Survey Methodology, June 1989
Vol. 15, No. 1, pp. 13–27
Statistics Canada

Estimation of Livestock Inventories Using Several Area- and Multiple-Frame Estimators

GEORGE E. BATTESE, NANCY A. HASABELNABY, and WAYNE A. FULLER[1]

ABSTRACT

Estimation of total numbers of hogs and pigs, sows and gilts, and cattle and calves in a state is studied using data obtained in the June Enumerative Survey conducted by the National Agricultural Statistics Service of the U.S. Department of Agriculture. It is possible to construct six different estimators using the June Enumerative Survey data. Three estimators involve data from area samples and three estimators combine data from list-frame and area-frame surveys. A rotation sampling scheme is used for the area frame portion of the June Enumerative Survey. Using data from the five years, 1982 through 1986, covariances among the estimators for different years are estimated. A composite estimator is proposed for the livestock numbers. The composite estimator is obtained by a generalized least-squares regression of the vector of different yearly estimators on an appropriate set of dummy variables. The composite estimator is designed to yield estimates for livestock inventories that are "at the same level" as the official estimates made by the U.S. Department of Agriculture.

KEY WORDS: June Enumerative Surveys; Rotation sample; Composite estimator; Generalized least squares.

1. INTRODUCTION

The National Agricultural Statistics Service (NASS), formerly the Statistical Reporting Service, of the U.S. Department of Agriculture (USDA) conducts probability surveys in June each year (the June Enumerative Surveys) to obtain data on farming operations. The survey data are a critical input in the construction of the official estimates of livestock numbers, crop acreages, grain stocks, *etc.* for the different states and for the United States as a whole. The sampling units in the farm surveys are selected from area frames and from list frames.

The area frame for a given state is the geographic area of the state stratified according to land use. The strata are defined by the percentage of the area that is cultivated, and whether the area is mainly urban, woodland, lakes, or other nonagricultural land. The sampling units for the area samples are called "segments", which vary in size in different states and strata, but are approximately one square mile in rural areas.

For the estimation of livestock inventories, samples of farm operators are also drawn from lists of farmers who raise the particular livestock. These list frames are stratified by measures of size. The area-frame and list-frame survey data are combined to obtain multiple-frame estimators for livestock numbers at the state level.

Different estimators can be constructed from the area-frame and list-frame samples. Statisticians within a state office of NASS calculate several estimators and make a recommendation for the official estimate of the number of livestock in the state. These materials are sent to the Agricultural Statistics Board within NASS in Washington D.C. The Board considers the different sample estimators, the recommendations of the state office, industry data, regional-level summaries and balance sheets when constructing the official estimates. Charting techniques

[1] George E. Battese, Department of Econometrics, University of New England, Armidale, N.S.W. 2351 Australia. Nancy A. Hasabelnaby and Wayne A. Fuller, Department of Statistics, Iowa State University, Ames, Iowa 50011, United States.

to maintain historical relationships among the data sources are also used by the Board. The Agricultural Statistics Board sets official estimates so that the official state estimates sum to the official national estimate.

A major drawback of the present procedure of establishing the official estimate is that there is no statistical measure of precision available for the official estimate. In 1983, a long-range planning group within NASS recommended that an objective procedure be developed for combining the different probability-based estimators into a composite estimator for the official estimate [see Allen, *et al.* 1983]. In 1984 it was recommended that a composite estimator should be made available for the consideration of the Agricultural Statistics Board [see Bynum, *et al.* 1985, p. 2].

The pooling of data from different, but related, samples and the combining of two or more estimators has been a subject of statistical research for many years. Some of this research is cited by Kuo (1986). Kuo also considers a composite estimator for livestock inventories based on USDA survey data.

In this study we investigate a procedure for constructing a composite estimator for livestock numbers. The values of several estimators for livestock inventories for a number of years and the variances and covariances among estimators for the different years are used in the construction. Assuming that the relationships among these estimators are defined by a simple linear model, we obtain the generalized least-squares estimator for the livestock inventories in the last year of sample data. Because the time-series of estimates is important, the set of composite estimators is constrained such that the average of the estimates for all years prior to the current year is equal to the average of the corresponding official estimates. This preserves the level of the time-series relative to previous official estimates. Alternative level constraints could be imposed.

2. AREA- AND MULTIPLE-FRAME ESTIMATORS

In the area-frame June Enumerative Survey, sample segments are identified on maps and all farm operators who have farming activities within these segments are identified and interviewed. The interviewers determine whether or not the farm operators in a given sample segment have their residences located within that segment. An area (or a collection of areas) of land within a sample segment that is under one type of management arrangement is called a "tract". A tract may be part of a farm or an entire farm.

The interviewer obtains information on the farming operation for each tract within a sample segment, including the size of the tract. In addition, information is obtained on the total farming operation of each sample farm operator. This information can be used to construct three different estimators of totals. The three estimators are called the closed-, open- and weighted-segment area-frame estimators. They differ mainly in the way in which farm values are associated with the segment.

The closed-segment area-frame estimator uses values associated with the operation of each tract within a sample segment. The open-segment area-frame estimator uses the values for the entire farm operation for those farms whose operators have their residences within the sample segment. The weighted-segment area-frame estimator uses values for the entire farm operation for farms with tracts in the sample segment. The values are prorated to the tract level by multiplying the farm total by the proportion of the total farm area that is within the sample segment. The weighted-segment value for a segment is the sum of the prorated values summed over all tracts within the sample segment. The closed-, open- and weighted-segment area-frame estimators of totals are defined by multiplying the corresponding segment values by their segment weights (inverses of the probabilities of selection of the segments) and adding these values over all sample segments and strata within the state. The three estimators are compared and discussed by Houseman (1975) and Nealon (1984).

The closed-segment area-frame estimator is considered to have a smaller variance than the open-segment area-frame estimator for most variables that can be easily reported on a tract basis. Items such as farm expenditures and livestock deaths are not easily reported at the tract level. The closed-segment area-frame estimator is preferred for estimation of national crop acreages and is also calculated, along with other estimators, for livestock inventories in most states. When values of variables can easily be associated with tracts, the closed-segment area-frame estimator is generally preferred because it is believed that the data obtained are less subject to reporting error by farm operators than information for the whole farm.

The weighted-segment area-frame estimator generally has the smallest variance of the three area-frame estimators. The weighted-segment estimator can be used for estimation of the population total for any agricultural item. Nealon (1984, p. 19) cites several research studies which show that the weighted-segment area-frame estimator is biased because the total farm size is frequently underreported. It is generally believed that some areas in woodland, pastureland, idle land, and farmsteads are not reported as part of the farm. If so, the ratio of the tract area to the total farm area will be too large and the weighted-segment area-frame estimator will be positively biased.

Multiple-frame estimators for livestock inventories use sample data from two or more frames. In the case of livestock, there are usually two frames, the area frame and a list frame. The list frame is a list of operators that were known, at one time, to have the livestock of interest. The list frame is incomplete but generally contains many of the large operators. For estimation of hog inventories in the study state, multiple-frame estimators are obtained by summing the estimator for the total of the list frame constructed with the list sample and an estimator for the nonoverlap domain (those operators not found in the list frame) from the area sample. The list sample is considered to be independent of the area sample. Different multiple-frame estimators are obtained when the closed-, open- and weighted-segment area-frame estimators are used for the nonoverlap domain.

3. COMPOSITE ESTIMATOR

We propose a composite estimator for the livestock inventory constructed under the assumption that a linear model defines the relationship among the different estimators. Suppose that N estimators for a given livestock inventory are available in each of T years and that the Agricultural Statistics Board has made official estimates for the first T-1 years. It is assumed that a composite estimator for the livestock inventory in the T-th year is desired.

Let Y_{ti} represent the i-th estimator for the t-th year, where $t = 1, 2, \ldots, T$ and $i = 1, 2, \ldots, N$. We assume the linear model,

$$Y_{ti} = \alpha_t + \beta_i + e_{ti}, \tag{3.1}$$

where α_t is the livestock inventory for the t-th year;

β_i is the effect associated with the i-th estimator; and

e_{ti} is a random error which has mean zero.

The estimator effects, $\beta_1, \beta_2, \ldots, \beta_N$, are included to account for the fact that nonsampling errors may cause different estimators to have different expectations. Model (3.1) specifies the estimator effects to be additive and constant over years. The assumption of constant effects is a simple specification that is consonant with the data.

The model (3.1) is a classical two-way, analysis-of-variance model, whose parameters are not estimable without additional model assumptions. To identify the parameters of the model, we restrict the average of the true livestock inventories in the first (T-1) years to be equal to the average of the corresponding official estimates of the Agricultural Statistics Board. This restriction is

$$\sum_{t=1}^{T-1} \alpha_t = \sum_{t=1}^{T-1} a_t, \tag{3.2}$$

where a_t is the official estimate for the t-th year. This constraint forces the estimates of livestock inventories to be at the same level as the previous official estimates. This is judged a reasonable constraint because actual values for α_t cannot be obtained and the time series nature of the estimates is important.

Given the restriction (3.2), the linear model (3.1) can be expressed in terms of the parameters, $\alpha_2, \alpha_3, \ldots, \alpha_T$ and $\beta_1, \beta_2, \ldots, \beta_N$, as

$$\begin{aligned} Y_{1i}^* &= -\sum_{j=2}^{T-1} \alpha_j + \beta_i + e_{1i} \\ Y_{ti} &= \alpha_t + \beta_i + e_{ti} \end{aligned} \tag{3.3}$$

where $t = 2, 3, \ldots, T$; and $Y_{1i}^* \equiv Y_{1i} - \sum_{j=1}^{T-1} a_j$, $i = 1, 2, \ldots, N$.

The model in matrix notation is

$$Y* = X\underset{\sim}{\gamma} + e, \tag{3.4}$$

where $Y* \equiv (Y_{11}^*, \ldots, Y_{1N}^*, Y_{21}, \ldots, Y_{2N}, \ldots, Y_{T1}, \ldots, Y_{TN})'$;

X is the ($NT \times K$) matrix of dummy variables associated with the model (3.3), where $K = T - 1 + N$;

$\underset{\sim}{\gamma} \equiv (\alpha_2, \alpha_3, \ldots, \alpha_T, \beta_1, \beta_2, \ldots, \beta_N)'$; and

e is the NT-column vector of random errors, having covariance matrix, V.

The covariance matrix, V, is the covariance matrix of the sampling errors, e_{ti}, associated with the different estimation procedures. The estimators, Y_{ti}, $t = 1, 2, \ldots, T$; $i = 1, 2, \ldots, N$, are correlated within any given year because they are based on the same area segments and the same list sample. The estimators are also correlated among years because sample segments in the area sample are included in the surveys for several years, according to a rotation sampling scheme. The list sample is selected independently each year. The variances and covariances of the estimators for any given year can be estimated by standard survey-sampling methods. Because the same list-sample estimator is used in defining the three multiple-frame estimators in a given year, the covariance between any two of the multiple-frame estimators in the same year will have a component due to the variance of the estimator obtained from the list sample. The covariances between estimators in different years, $\mathrm{Cov}(Y_{ti}, Y_{t'j})$, where $t \neq t'$, can be estimated by standard methods, using the sample segments that are common to the two years. If it is assumed that the variances and covariances in V satisfy particular relationships, then these conditions can be imposed as part of the estimation procedure.

Given an estimator of the covariance matrix, denoted by $\hat{V}^*$, the estimated generalized least-squares estimator of the parameter vector, $\underset{\sim}{\gamma}$, is

$$\hat{\underset{\sim}{\gamma}} = (X'\hat{V}^{*-1}X)^{-1}(X'\hat{V}^{*-1}Y*). \tag{3.5}$$

The covariance matrix of $\hat{\underset{\sim}{\gamma}}$ is estimated by

$$\mathrm{C\hat{o}v}(\hat{\underset{\sim}{\gamma}}) = (X'\hat{V}^{*-1}X)^{-1}. \tag{3.6}$$

The estimated generalized least-squares estimator, $\hat{\alpha}_T$, which is the $(T\text{-}1)$-th element of $\hat{\underset{\sim}{\gamma}}$, is a possible composite estimator for the livestock inventory for the T-th year. Its variance is estimated by the corresponding element of the estimated covariance matrix (3.6). Furthermore, the estimated generalized least-squares estimators, $\hat{\alpha}_T + \hat{\beta}_i$, $i = 1, 2, \ldots, N$, are adjusted area-frame and multiple-frame estimators for livestock inventories in the T-th year which are based on the model (3.4). The variances of these adjusted estimators are estimated by obtaining the appropriate linear functions of the estimated covariance matrix (3.6).

If the model (3.4) is true and the random errors have a normal distribution, then the weighted sum of squares,

$$x^2 = (Y* - X\hat{\underset{\sim}{\gamma}})'\hat{V}^{*-1}(Y* - X\hat{\underset{\sim}{\gamma}}), \tag{3.7}$$

has a chi-square distribution with parameter, $NT - K$. Thus the weighted residual sum of squares obtained by using the estimated covariance matrix yields an approximate test of the adequacy of the model (3.1).

4. EMPIRICAL RESULTS

4.1 Introduction

In the USDA June Enumerative Surveys between 1982 and 1986, a total of 298 area segments were sampled in the study state. These segments were included in the June Enumerative Surveys according to a rotation sampling scheme in which approximately twenty percent of the segments are replaced each year. The actual replacement rate varies, but we construct estimators as if the rate was exactly twenty percent.

The area frame for the state consists of eleven strata: nine strata are agricultural land, with varying percentages cultivated; one stratum is agri-urban land; and one stratum consists of residential or commercial areas.

The list frame for hog producers in the study state consists of eleven strata, which are defined by the total number of hogs raised by the farm operators at a particular time. The strata are defined by operations with: no livestock, livestock but no hogs, 1-99 hogs, 100-199 hogs, ..., more than 6,000 hogs. The list frame for cattle that is sampled in June contains very large operators. It was a small list of less than 500 operators in each of the study years. The cattle list is divided into four strata. Three strata are defined by the total number of cattle and calves, where the strata are between 1,000 and 2,999, between 3,000 and 9,999, and more than 10,000. The fourth stratum is composed of farm operators with at least 200 dairy cattle.

The total number of farm operators in the area sample of the June Enumerative Surveys averaged about 2,350 during the years studied with a range of 120. The list sample for hogs averaged about 2,400 farm operators with a range of 100, whereas the list sample for cattle averaged about 70 farm operators with a range of 71. Using these data, the values of the closed-, open-, and weighted-segment area-frame estimators and the three corresponding multiple-frame estimators for the total number of hogs and pigs, sows and gilts, and cattle and calves were computed for each of the five years. The estimates where obtained by use of PC CARP, which is a computer program for performing survey-sampling estimation on personal computers [see Fuller, *et al.* 1986 and Schnell, *et al.* 1988]. The variance estimators are the usual estimators for an estimated total constructed from a stratified cluster sample. See, for example, Cochran (1977).

The data used for variance computations were treated as complete data although some data were imputed for nonresponse. The imputation, especially since the imputation methods draw heavily upon prior year data in the rotation scheme, may lead to an overestimate of the correlation between years.

Table 1

Estimates for livestock inventories in 1986

	Hogs and pigs	Sows and gilts	Cattle and calves
Area-Frame Estimators			
Closed-Segment	18.42	15.78	15.27
	(1.97)	(2.17)	(1.53)
Open-Segment	21.11	18.24	18.74
	(2.82)	(2.69)	(2.35)
Weighted-Segment	21.69	18.85	15.48
	(1.67)	(1.62)	(1.15)
Multiple-Frame Estimators			
Closed-Segment	18.11	15.59	16.12
	(1.11)	(1.28)	(1.38)
Open-Segment	18.06	15.29	19.97
	(1.26)	(1.39)	(2.08)
Weighted-Segment	18.50	15.82	16.22
	(1.00)	(1.00)	(1.00)

Estimates for the livestock inventories in 1986 and the estimated standard deviations of the estimators for 1986 are given in Table 1. Each standard deviation in Table 1 is the square root of the average of the five estimated variances for the five years. The units in the table are determined by coding the standard deviation of the weighted-segment multiple-frame estimators to be 1.00 for all livestock inventories. This makes comparison easy and also complies with confidentiality rules.

As expected from previous studies [*e.g.*, see Nealon (1984)], the open-segment area-frame estimator is the least precise estimator for livestock inventories. The most precise estimator is the multiple-frame estimator which uses the weighted-segment estimator for the nonoverlap domain. Coefficients of variation for the weighted-segment area-frame estimators are about 7% to 9%, whereas the weighted-segment multiple-frame estimators have coefficients of variation of about 5.5% to 6.5%. Because the list sample for hog inventories is larger than that for cattle and calves, the precision of the multiple-frame estimators relative to the area-frame estimators is much greater for hog iventories than for cattle and calves.

4.2 Estimation of Covariance Matrices

The estimation of the covariance matrix for the six estimators for the five years of data proceeded in several steps. The covariance matrix for the error vector, $\boldsymbol{e}$, in (3.4) can be written in the form

$$V = \begin{pmatrix} V_{11} & V_{12} & V_{13} & V_{14} & V_{15} \\ V_{21} & V_{22} & V_{23} & V_{24} & V_{25} \\ V_{31} & V_{32} & V_{33} & V_{34} & V_{35} \\ V_{41} & V_{42} & V_{43} & V_{44} & V_{45} \\ V_{51} & V_{52} & V_{53} & V_{54} & V_{55} \end{pmatrix} \tag{4.1}$$

where, for a particular inventory type, V_{tj} is the 6 x 6 matrix of covariances between the six estimators for year t and the six estimators for year j. With the rotation scheme used, the covariance of estimators in two years is a function of the number of rotation groups common to the two years. Let $k = |t - j|$ for $k = 0, 1, \ldots, 4$. Then the covariance matrix, V_{tj}, can be estimated from the area segments of the $5 - k$ rotation groups which are common to the two years t and j.

We estimate the elements of the covariance matrix (4.1) imposing some additional assumptions about its structure. Our primary interest is to compare the precision of the alternative estimators and this comparison is facilitated by the assumptions which follow.

We assume that the covariance matrices for years that are the same distance apart are the same and are symmetric. This is, we assume

$$V_{tj} = V_{t+r,j+r}$$

and (4.2)

$$V_{tj} = V_{tj}'$$

where V_{tj} are the submatrices of (4.1); $r = 0, 1, \ldots, \max(5\text{-}t, 5\text{-}j)$; and $t,j = 1, 2, \ldots, 5$. For $t = j$ and $r = 0$, the assumptions of (4.2) imply

$$V_{11} = V_{22} = V_{33} = V_{44} = V_{55} \equiv V_0.$$

For $t \neq j$, the assumptions of (4.2) imply the following:

$$V_{12} = V_{23} = V_{34} = V_{45} \equiv V_1,$$
$$V_{13} = V_{24} = V_{35} \equiv V_2,$$
$$V_{14} = V_{25} \equiv V_3,$$

and

$$V_{15} \equiv V_4.$$

These assumptions are in reasonable agreement with the data. Good agreement was anticipated because the sample size is very stable over the five years and there were no large shifts in livestock inventories.

We estimate the distinct submatrices of (4.1) by averaging the corresponding estimated covariance matrices obtained from common segments. The averaging process was based on the correlation matrices. Let the covariance matrix of the estimated totals defined in (4.1) be expressed as

$$V = S\,C\,S\,,$$

where S is the 30 x 30 diagonal matrix of the estimated standard deviations of the six estimators for the five years and C is the 30 x 30 correlation matrix, partitioned in the same manner as V of (4.1).

The estimator of the correlation matrix C is constructed by averaging estimates of the submatrices of C. Using the segments common to two years, the covariance matrix of the two vectors of estimated totals constructed with those segments was estimated by the usual stratified cluster formulae. The estimated covariance matrices were converted to correlation matrices and these estimates were called the direct estimates. Let

$$\hat{C}_0 = (\tfrac{5}{5})\tfrac{1}{5}(\hat{C}_{11} + \hat{C}_{22} + \hat{C}_{33} + \hat{C}_{44} + \hat{C}_{55})$$
$$\hat{C}_1 = (\tfrac{4}{5})\tfrac{1}{4}(\hat{C}_{12} + \hat{C}_{23} + \hat{C}_{34} + \hat{C}_{45})$$
$$\hat{C}_2 = (\tfrac{3}{5})\tfrac{1}{3}(\hat{C}_{13} + \hat{C}_{24} + \hat{C}_{35})$$
$$\hat{C}_3 = (\tfrac{2}{5})\tfrac{1}{2}(\hat{C}_{14} + \hat{C}_{25})$$
$$\hat{C}_4 = (\tfrac{1}{5})\hat{C}_{15},$$

where the $\hat{C}_{tj}$ are the directly estimated correlation matrices based on common segments. The factors in parentheses represent the fraction of segments that are common to the estimates. This fraction arises from the rotation-sampling scheme in which twenty percent of the segments in the area sample are dropped from the sample each year and twenty percent new segments are added. By the independence assumption, the correlation between the segments rotated out and those rotated in is zero.

Since the estimated correlation matrices, $\hat{C}_{tj}$, are not symmetric when $t \neq j$, the symmetric assumption, $V_{tj} = V'_{tj}$, in (4.2) is imposed on the estimated covariance matrix by defining

$$\overset{*}{\hat{C}}_r = ½(\hat{C}_r + \hat{C}'_r)\,, \; r = 1, 2, 3, 4.$$

Let $\overset{*}{S}$ be the 6 x 6 diagonal matrix of the square roots of the average estimated variances of the six estimators, where the average is over the five years. Again, for confidentiality requirements, the estimated variances are standardized such that the estimated variance of the weighted-segment multiple-frame estimator is equal to 1.00. Then the estimated covariance matrix for the six estimators for the five years is

$$\overset{*}{V} = \begin{pmatrix} \overset{*}{S} & 0 & 0 & 0 & 0 \\ 0 & \overset{*}{S} & 0 & 0 & 0 \\ 0 & 0 & \overset{*}{S} & 0 & 0 \\ 0 & 0 & 0 & \overset{*}{S} & 0 \\ 0 & 0 & 0 & 0 & \overset{*}{S} \end{pmatrix} \begin{pmatrix} \hat{C}_0 & \overset{*}{C}_1 & \overset{*}{C}_2 & \overset{*}{C}_3 & \overset{*}{C}_4 \\ \overset{*}{C}_1 & \hat{C}_0 & \overset{*}{C}_1 & \overset{*}{C}_2 & \overset{*}{C}_3 \\ \overset{*}{C}_2 & \overset{*}{C}_1 & \hat{C}_0 & \overset{*}{C}_1 & \overset{*}{C}_2 \\ \overset{*}{C}_3 & \overset{*}{C}_2 & \overset{*}{C}_1 & \hat{C}_0 & \overset{*}{C}_1 \\ \overset{*}{C}_4 & \overset{*}{C}_3 & \overset{*}{C}_2 & \overset{*}{C}_1 & \hat{C}_0 \end{pmatrix} \begin{pmatrix} \overset{*}{S} & 0 & 0 & 0 & 0 \\ 0 & \overset{*}{S} & 0 & 0 & 0 \\ 0 & 0 & \overset{*}{S} & 0 & 0 \\ 0 & 0 & 0 & \overset{*}{S} & 0 \\ 0 & 0 & 0 & 0 & \overset{*}{S} \end{pmatrix} \tag{4.3}$$

The estimated covariance matrices, $\overset{*}{V}_0 \equiv \overset{*}{S}\,\hat{C}_0\overset{*}{S}$, for the livestock inventories are given in Table 2. The estimates of the four unique off-diagonal submatrices, $\overset{*}{V}_r \equiv \overset{*}{S}\,\overset{*}{C}_r\overset{*}{S}$, $r = 1, 2, 3, 4$, are available from the authors on request.

Table 2

Estimated covariance matrices for the six estimators of livestock inventories within a year

	Area-Frame Estimators			Multiple-Frame Estimators		
	Closed	Open	Weighted	Closed	Open	Weighted
A. Hogs and pigs						
	3.886	4.077	2.366	0.654	0.688	0.405
	4.077	7.959	2.394	0.698	1.150	0.430
	2.366	2.394	2.784	0.373	0.409	0.481
	0.654	0.698	0.373	1.242	1.239	0.936
	0.688	1.150	0.409	1.239	1.590	0.937
	0.405	0.430	0.481	0.936	0.937	1.000
B. Sows and gilts						
	4.720	4.274	2.455	1.102	1.112	0.572
	4.274	7.260	2.322	1.119	1.427	0.548
	2.455	2.322	2.621	0.481	0.487	0.499
	1.102	1.119	0.481	1.638	1.658	1.033
	1.112	1.427	0.487	1.658	1.934	1.033
	0.572	0.548	0.499	1.033	1.033	1.000
C. Cattle and calves						
	2.355	1.951	1.141	1.853	1.655	0.907
	1.951	5.527	1.014	1.652	4.418	0.912
	1.141	1.014	1.321	0.913	0.891	0.925
	1.853	1.652	0.913	1.910	1.756	0.992
	1.655	4.418	0.891	1.756	4.310	1.017
	0.907	0.912	0.925	0.992	1.017	1.000

Consider a sample composed of a common set of rotation groups observed in each of the five years, rather than the existing sample in which twenty percent of the sample segments are dropped each year. For the sample with no rotation, the covariance matrix of the six estimators for the five years, expressed in terms of the submatrices of (4.1), is

$$\begin{pmatrix} V_{11} & \frac{5}{4}V_{12} & \frac{5}{3}V_{13} & \frac{5}{2}V_{14} & 5V_{15} \\ \frac{5}{4}V_{21} & V_{22} & \frac{5}{4}V_{23} & \frac{5}{3}V_{24} & \frac{5}{2}V_{25} \\ \frac{5}{3}V_{31} & \frac{5}{4}V_{32} & V_{33} & \frac{5}{4}V_{34} & \frac{5}{3}V_{35} \\ \frac{5}{2}V_{41} & \frac{5}{3}V_{42} & \frac{5}{4}V_{43} & V_{44} & \frac{5}{4}V_{45} \\ 5V_{51} & \frac{5}{2}V_{52} & \frac{5}{3}V_{53} & \frac{5}{4}V_{54} & V_{55} \end{pmatrix} \tag{4.4}$$

Direct sample estimates of the submatrices, V_{tj}, obtained from segments common to years t and j sometimes gave a covariance matrix (4.4) that was not positive definite. For example, this can happen if operators with very large holdings are among those operators in the one rotation common to all five years. When the assumptions of (4.2) are imposed in the estimation process, the estimates of the covariance matrix (4.4) were positive definite for all three livestock inventories.

Table 3

Composite estimates for the livestock inventories in 1986 and the effects for different estimators

	Hogs and pigs[1]	Sows and gilts[1]	Cattle and calves[1]
Composite Estimator	18.84	18.06	16.43
	(1.01)	(1.02)	(1.03)
Effects of Area-Frame Estimators			
Closed-Segment	−1.13	−2.26	−0.21
	(1.30)	(1.36)	(0.99)
Open-Segment	0.26	−1.09	1.03
	(1.86)	(1.78)	(1.45)
Weighted-Segment	1.24	−0.94	−0.26
	(1.14)	(1.10)	(0.80)
Effects of Multiple-Frame Estimators			
Closed-Segment	−0.33	−1.86	0.04
	(0.66)	(0.78)	(0.92)
Open-Segment	−0.11	−1.82	1.40
	(0.75)	(0.84)	(1.32)
Weighted-Segment	0.19	−1.74	−0.31
	(0.59)	(0.59)	(0.69)

[1] Standard errors are given in parentheses.

4.3 Model Estimation

Given the estimated covariance matrix, $\hat{V}^*$, we estimate the parameters of model (3.4) by using the estimated generalized-least-squares estimator (3.5). The values of the composite estimator, $\hat{\alpha}_T$, for 1986 livestock inventories are given in the first line of Table 3. The six estimator effects, denoted by β_i in model (3.3), are also given in the table. The estimated standard deviation of the composite estimator is slightly larger than that of the weighted-segment multiple-frame estimator. The increase in variance comes from the fact that the level of the estimator is estimated using past sample estimates and past official estimates.

The residual sums of squares defined in (3.7) were 18.22, 15.38, and 24.59, for hogs and pigs, sows and gilts, and cattle and calves, respectively. The degrees of freedom is 20 because, for each livestock inventory, there are thirty observations in the $Y*$-vector and ten parameters are estimated in $\underset{\sim}{\gamma}$. In no case does the residual sum of squares exceed 31.41, which is the 95-th percentile for the chi-square distribution with 20 degrees of freedom.

The composite estimator in Table 3 has nearly the same standard deviation as the weighted-segment multiple-frame estimator, the estimator with the smallest standard deviation (Table 1). Thus, one would expect the optimal linear combination of the six estimators for a single year to assign the majority of the weight to the weighted-segment multiple-frame estimator, and this is the case. The minimum variance weights for the data of a single year are calculated as

$$(\mathbf{1}'\hat{V}^*_0{}^{-1}\mathbf{1})^{-1}\mathbf{1}'\hat{V}^{*\,-1}_0,$$

where $\mathbf{1}' = (1, 1, 1, 1, 1, 1)$ and $\hat{V}^*_0$ is the covariance matrix of the six estimators given in Table 2 [see the diagonal elements of (4.3)]. The optimal weights and the estimated standard deviation of the optimal combination of the six estimators are presented in Table 4. Note that the sum of the weights is one for each livestock inventory. The difference between these standard errors and those of the first line of Table 3 is due to the estimation of level in the construction of the estimates of Table 3.

Table 4

Optimal weights for six estimators in a single year.

Estimators	Inventory type: Hogs and pigs	Sows and gilts	Cattle and calves
Area-Frame			
Closed-Segment	0.0541	−0.0152	0.0525
Open-Segment	−0.0084	0.0152	0.0656
Weighted-Segment	0.1463	0.1909	0.0909
Multiple-Frame			
Closed-Segment	0.1640	−0.0218	−0.0353
Open-Segment	−0.0116	−0.0191	−0.0772
Weighted-Segment	0.6556	0.8500	0.9035
Estimated standard error of optimal combination	0.94	0.95	0.99

Table 5
Estimated correlation coefficients between the weighted-segment area-frame estimators based on a common rotation group

h	Hogs & pigs	Sows & gilts	Cattle & calves
0	1.000	1.000	1.000
1	0.606	0.590	0.592
2	0.478	0.456	0.433
3	0.365	0.336	0.258
4	0.304	0.217	0.097

4.4 Estimation Using Rotation Group Means

In obtaining the estimates of Section 4.3, we did not use all of the available information. We used the estimators for each year, but did not decompose the estimators into the parts associated with each rotation group. In this section we construct an estimator using the individual rotation group means of the weighted-segment area-frame estimator. We retain the assumption that the variance of the estimator is the same across years. Under that assumption, the correlation coefficients are assumed to depend only on the number of years between the estimators involved. Let ρ_h represent the correlation coefficient between the weighted-segment area-frame estimators for a common rotation group which is observed h years apart, $h = 0, 1, \ldots, 4$. For the three inventory types, the estimated correlation coefficients are given in Table 5. The estimated correlation coefficients between estimators for h years apart are the averages of the correlation coefficients estimated from the $5 - h$ rotation groups involved. There are a total of nine rotation groups for the five years.

Let Z_{tj} represent the weighted-segment area-frame estimator in rotation group j for year t, where $j = t, t+1, \ldots, t+4$ and $t = 1, 2, \ldots, 5$. Then, for a given year, t, we assume that Z_{tj} is an unbiased estimator of the unknown total inventory, α_t. It is known that a rotation group bias may exist and need to be estimated, but we ignore that effect in this illustration. The model is

$$\begin{aligned} Z_{tj} = \alpha_t + \epsilon_{tj}\,,\ t &= 1, 2, \ldots, 5; \\ j &= t, t+1, \ldots, t+4\,, \end{aligned} \tag{4.5}$$

where the errors, ϵ_{tj}, have zero mean. The model (4.5), in matrix notation, is

$$Z = D\underset{\sim}{\alpha} + \underset{\sim}{\epsilon}\,,$$

where

$$\begin{aligned} Z' &= (Z_{11}, Z_{12}, \ldots, Z_{15}; Z_{22}, Z_{23}, \ldots, Z_{26}; \ldots; \\ &\quad Z_{55}, Z_{56}, \ldots, Z_{59})\,, \\ \underset{\sim}{\alpha} &= (\alpha_1, \alpha_2, \ldots, \alpha_5)\,, \end{aligned}$$

$D = I_5 \otimes \mathbf{1}_5$, I_5 is the identity matrix of order 5 and $\mathbf{1}_5$ is the (5×1) vector with all elements equal to one.

Let the correlation matrix for the rotation-group estimators, $\boldsymbol{Z}$, be

$$\boldsymbol{W} = \begin{pmatrix} \boldsymbol{W}_0 & \boldsymbol{W}_1' & \boldsymbol{W}_2' & \boldsymbol{W}_3' & \boldsymbol{W}_4' \\ \boldsymbol{W}_1 & \boldsymbol{W}_0 & \boldsymbol{W}_1' & \boldsymbol{W}_2' & \boldsymbol{W}_3' \\ \boldsymbol{W}_2 & \boldsymbol{W}_1 & \boldsymbol{W}_0 & \boldsymbol{W}_1' & \boldsymbol{W}_2' \\ \boldsymbol{W}_3 & \boldsymbol{W}_2 & \boldsymbol{W}_1 & \boldsymbol{W}_0 & \boldsymbol{W}_1' \\ \boldsymbol{W}_4 & \boldsymbol{W}_3 & \boldsymbol{W}_2 & \boldsymbol{W}_1 & \boldsymbol{W}_0 \end{pmatrix}$$

where $\boldsymbol{W}_0 = \boldsymbol{I}_5$,

$$\boldsymbol{W}_1 = \begin{bmatrix} 0 & \rho_1 & 0 & 0 & 0 \\ 0 & 0 & \rho_1 & 0 & 0 \\ 0 & 0 & 0 & \rho_1 & 0 \\ 0 & 0 & 0 & 0 & \rho_1 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{W}_2 = \begin{bmatrix} 0 & 0 & \rho_2 & 0 & 0 \\ 0 & 0 & 0 & \rho_2 & 0 \\ 0 & 0 & 0 & 0 & \rho_2 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix},$$

$$\boldsymbol{W}_3 = \begin{bmatrix} 0 & 0 & 0 & \rho_3 & 0 \\ 0 & 0 & 0 & 0 & \rho_3 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}, \qquad \boldsymbol{W}_4 = \begin{bmatrix} 0 & 0 & 0 & 0 & \rho_4 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 \end{bmatrix}.$$

Then, the generalized least-squares weighted-segment area-frame estimator is

$$\hat{\underset{\sim}{\alpha}} = (\boldsymbol{D}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{Z},$$

where $\hat{\boldsymbol{W}}$ is the estimator for the correlation matrix, $\boldsymbol{W}$. The covariance matrix of $\hat{\underset{\sim}{\alpha}}$ is estimated by

$$\hat{Cov}(\hat{\underset{\sim}{\alpha}}) \equiv (\boldsymbol{D}'\hat{\underset{\sim}{\Sigma}}^{-1}\boldsymbol{D})^{-1},$$

where $\hat{\underset{\sim}{\Sigma}} \equiv 5\hat{\boldsymbol{W}}$ is the covariance matrix of $\boldsymbol{Z}$, whose units are such that the estimated variance of the weighted-segment area-frame estimator is one. The estimated covariance matrices, $\hat{Cov}(\hat{\underset{\sim}{\alpha}})$, for the three livestock types are given in Table 6. We see that the estimators obtained using the individual rotation group estimates are about 10% more efficient than the weighted-segment area-frame estimators for 1986.

The optimal weights for the vector of individual rotation estimates are

$$(\boldsymbol{D}'\hat{\boldsymbol{W}}^{-1}\boldsymbol{D})^{-1}\boldsymbol{D}'\hat{\boldsymbol{W}}^{-1}.$$

The weights are available from the authors.

The generalized least squares procedure can be applied to other combinations of rotation group and year estimators, but the results suggest that additional gains would be modest.

Table 6

Estimated covariance matrices for weighted-segment area-frame estimators using information in the rotation scheme

	1982	1983	1984	1985	1986
Hogs and pigs					
1982	0.899	0.436	0.283	0.180	0.124
1983	0.436	0.857	0.412	0.273	0.180
1984	0.283	0.412	0.844	0.412	0.283
1985	0.180	0.273	0.412	0.857	0.436
1986	0.124	0.180	0.283	0.436	0.899
Sows and gilts					
1982	0.908	0.429	0.272	0.167	0.099
1983	0.429	0.866	0.405	0.262	0.167
1984	0.272	0.405	0.853	0.405	0.272
1985	0.167	0.262	0.405	0.866	0.429
1986	0.099	0.167	0.272	0.429	0.908
Cattle and calves					
1982	0.914	0.438	0.264	0.135	0.061
1983	0.438	0.870	0.412	0.253	0.135
1984	0.264	0.412	0.856	0.412	0.264
1985	0.135	0.253	0.412	0.870	0.438
1986	0.061	0.135	0.264	0.438	0.914

5. CONCLUSION

The composite estimator suggested in this paper provides a method for combining the values of several estimators for livestock inventories. The composite estimator uses the values of the different area-frame and multiple- frame estimators in several preceding years, as well as the values in the year for which the official estimate is sought. The optimal linear combination of the six estimators within a particular year has a variance that is two to twelve percent less than that of the weighted-segment multiple-frame estimator. Including the estimators from the other four years produces an additional reduction of one to two percent in the variance of the composite estimator for the current year. The data required to calculate the weighted-segment multiple-frame estimator are those required for the other five area- and multiple-frame estimators. The greatest effort required in constructing the composite estimator is the estimation of the covariance matrix for the estimators over the years in which sample data are available. Because the variances are relatively constant over years, the weight vector can be calculated in advance and applied to the estimates of the current year. Then the marginal effort required for the composite estimator during the estimation year is very small.

ACKNOWLEDGMENTS

This research was partially supported by Research Agreement No. 53-319T-6-00073 with the National Agricultural Statistics Service of the U.S. Department of Agriculture. The authors thank Ron Fecso and Vic Tolomeo for their assistance. Ron Fecso's comments on an earlier draft are gratefully acknowledged. The work was conducted when the first author was on study leave at Iowa State University.

REFERENCES

ALLEN, R., CLAMPET, G., DUNKERLEY, C., TORTORA, R., and VOGEL, F. (1983). Framework for the Future. Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C.

BYNUM, H., DOWDY, W., HANUSCHAK, G., HUDSON, C., MURPHY, R., STEINBERG, J., and VOGEL, F.A. (1985). Crop Reporting Board Standards. Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C.

COCHRAN, W.G (1977). *Sampling Techniques*. New York: Wiley.

FULLER, W.A., KENNEDY, W., SCHNELL, D., SULLIVAN, G., and PARK, H.J. (1986). PC CARP. Statistical Laboratory, Iowa State University, Ames, Iowa.

HOUSEMAN, E.E. (1975). Area Frame Sampling in Agriculture. SRS Report No. 20. Statistical Reporting Service, U.S. Department of Agriculture, Washinton, D.C.

KUO, L. (1986). Composite Estimation of Totals for Livestock Surveys. SF & SRB Staff Report No. 92. Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C.

NEALON, J.P. (1984). Review of the Multiple and Area Frame Estimators. SF & SRB Staff Report No. 80. Statistical Reporting Service, U.S. Department of Agriculture, Washington, D.C.

SCHNELL, D., KENNEDY, W.J., SULLIVAN, G., PARK, H.J., and FULLER, W.A. (1988). Personal computer variance software for complex surveys. *Survey Methodology* 14, 59-69.