# MULTIPLE FRAME ESTIMATION WITH STRATIFIED OVERLAP DOMAIN

SAMPLE SURVEY RESEARCH BRANCH
RESEARCH DIVISION
STATISTICAL REPORTING SERVICE
U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C.

July 1976
SF 76-01

# MULTIPLE FRAME ESTIMATION WITH STRATIFIED OVERLAP DOMAIN

By

RAYMOND R. BOSECKER

and

BARRY L. FORD

SAMPLE SURVEY RESEARCH BRANCH
RESEARCH DIVISION
STATISTICAL REPORTING SERVICE
U.S. DEPARTMENT OF AGRICULTURE
WASHINGTON, D.C.

July 1976

CONTENTS

## ACKNOWLEDGEMENTS

We would like to thank Norman Beller, Harold Huddleston, Fred Vogel, and Margaret Weidenhamer for reviewing this paper. Special thanks to Cindy Musselman for her special efforts in typing this report.

## SUMMARY AND CONCLUSIONS

The Statistical Reporting Service (SRS) currently uses a "screening" estimator for most multiple frame estimates. The estimate and variance for the area frame non-overlap domain are simply added to the estimate and variance from a stratified list frame. The overlap domain from the area frame is discarded in the screening multiple frame estimator.

The "Hartley" full multiple frame estimator which utilizes both frames to estimate the overlap domain before adding the non-overlap component has also been investigated. Overlap domain estimates were weighted together based on the relative size of the variances from the two frames. The variance of the list frame estimate was computed by stratum while the variance for the area frame overlap estimate was computed over the entire overlap domain. In most SRS surveys, the area frame overlap estimate received a very small weight because it had a large variance relative to the stratified list frame estimate. Only small gains were achieved over the screening estimator for SRS programs so the full estimator is not generally computed.

This paper presents a multiple frame estimator which takes advantage of stratification within the overlap domain for two-frame estimation. The overlap domain strata are defined by the stratified list frame. Matching the area frame units against the list frame units enables each stratum to be estimated by both frames. A vector of optimum weights combines the stratum estimates from the two frames.

The proposed multiple frame estimator therefore combines the overlap domain estimates from the two frames optimally at the stratum level before summing over strata. Proof is presented that the variance is less than or equal to the Hartley multiple frame estimator. This was demonstrated by applying the proposed estimator to livestock survey data for two states. These estimates had smaller sampling errors than either the screening or Hartley estimators in both states. Sampling errors were 14 percent below those of the screening estimator for both hog and cattle estimates in one State.

The stratum-by-stratum multiple frame procedure also provides a way to measure the effect on variance from combining list and area frame estimates in each stratum. Only a minimal reduction in variance resulted from the optimum combination of the list frame with the area frame in the smaller size group livestock strata.

No additional data from respondents or change in office edit routine is required to use the proposed estimator. Each area frame tract operator matched with the list frame would simply be coded according to the list stratum of the matching operator on the list.

# Multiple Frame Estimation With Stratified Overlap Domain

By

Raymond R. Bosecker and Barry L. Ford

## 1. Introduction

The multiple frame procedure is a common tool of the survey statistician. Random samples are drawn from two or more sampling frames to estimate a specified variable. Most multiple frame surveys involve only two frames: an area frame and a list frame. The area frame is a complete sampling frame comprised of small blocks of land called segments. The entire population of segments contains all the variable of interest. Within each segment, land operated under a specific operator is called a tract. If the variable being estimated were number of livestock, for example, then livestock counts on all tracts within the sample segments are expanded to provide an area frame estimate.

A list frame may be composed of names of farm operators stratified by size based on number of livestock in the operation. A stratified sample of operators who report all livestock on their land provides estimates with smaller sampling errors for a given cost than the area frame. However, the list frame is incomplete. It does not cover the entire livestock population as does the area frame. To compensate for this incompleteness, a livestock estimate for names found in the area frame but not on the list is added to the list frame estimate. The portion of the area frame not found on the list is called the "nonoverlap domain". The combination of list frame and nonoverlap domain in called a "screening" estimate.

## 2. Alternative Estimators and Their Variances

### 2.1 Hartley Multiple Frame Estimator

From work done by Hartley (1.) a single multiple frame estimator may be obtained from the two frames by combining estimates for the "overlap" domain they have in common as follows:

$$\hat{y}_H = \hat{y}_{no\ell} + q\,\hat{y}_{o\ell} + p\,\hat{y}_\ell$$

where:

$\hat{y}_{no\ell}$ = area frame estimate of total livestock located on land operated by someone <u>not</u> on the list. "*NOL*" therefore is "not on list" or "nonoverlap" domain.

$\hat{y}_{o\ell}$ = area frame estimate of total livestock located on land operated by someone on the list. "*OL*" therefore stands for "on list" or "overlap" domain.

$\hat{y}_\ell$ = list frame estimate obtained by stratified sampling from list of farm operators (overlap domain).

$p$ = weight attached to the list frame estimate of the overlap domain.

$q$ = weight attached to the area frame estimate of the overlap domain.

and $p+q$ = $1$ to combine the independent estimates from the list and area frames respectively for the overlap domain.

The area and screening estimators are special cases of this multiple frame formulation. The area estimator may be expressed as:

$$\hat{y}_A = \hat{y}_{no\ell} + \hat{y}_{o\ell}$$

where $q = 1$ and $p = 0$. The variance is:

$$Var(\hat{y}_A) = Var(\hat{y}_{no\ell}) + Var(\hat{y}_{o\ell}) + 2Cov(\hat{y}_{no\ell}, \hat{y}_{o\ell})$$

The screening estimator may be written:

$$\hat{y}_L = \hat{y}_{no\ell} + \hat{y}_{\ell}$$

where $q = 0$ and $p = 1$. The variance is:

$$Var(\hat{y}_L) = Var(\hat{y}_{no\ell}) + Var(\hat{y}_{\ell})$$

With $q$ and $p$ each different from zero, both the area and list frame estimates of the overlap domain influence the estimate. The Hartley multiple frame estimator may then be rewritten in terms of only $p$, list frame weight, since $q = 1 - p$. This alternative expression is derived in Section A of the Appendix as:

$$\hat{y}_H = \hat{y}_A + p(\hat{y}_{\ell} - \hat{y}_{o\ell})$$

The variance is shown in Section B of the Appendix to be:

$$Var(\hat{y}_H) = Var(\hat{y}_A) - 2pCov(\hat{y}_A, \hat{y}_{o\ell}) + p^2[Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell})]$$

The advantage of this multiple frame estimator should be a smaller sampling error than either the screening ($\hat{y}_L$) or area ($\hat{y}_A$) estimators. To achieve this, an optimum $p$ value, $p_{opt}$, must be used. The derivation of $p_{opt}$ is presented in Section C of the Appendix.

$$p_{opt} = \frac{Cov(\hat{y}_A, \hat{y}_{o\ell})}{Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell})} = \frac{Var(\hat{y}_{o\ell}) + Cov(\hat{y}_{no\ell}, \hat{y}_{o\ell})}{Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell})}$$

Substituting $p_{opt}$, the variance for the Hartley estimator becomes:

$$Var(\hat{Y}_H) = Var(\hat{Y}_A) - p_{opt} \, [Cov(\hat{Y}_A, \hat{Y}_{o\ell})]$$

This is shown in Section D of the Appendix.

It would not be difficult to use the full Hartley multiple frame estimator since each tract operator appearing in the area frame must be checked against the list to determine his overlap-nonoverlap status. Although $Var(\hat{Y}_H) < Var(\hat{Y}_L)$ because optimum $p$ is used in the Hartley estimator, previous experience with the full estimator versus the screening estimator has shown only small reductions in sampling error because the weight attached to the list frame has been so large relative to the area frame weight that little gain was realized by including the area frame estimate. In other words, for the entire overlap domain the $Var(\hat{Y}_{o\ell})$ was large relative to $Var(\hat{Y}_\ell)$ so that $p_{opt}$ approached $1$.

When the list frame has been stratified:

$$\hat{Y}_\ell = \sum_{h=1}^{k} \hat{Y}_{\ell h} \quad \text{where } k \text{ is the number of independent strata.}$$

The Hartley estimator may then be written as $\hat{Y}_H = \hat{Y}_{no\ell} + q \, \hat{Y}_{o\ell} + p \sum_{h=1}^{k} \hat{Y}_{\ell h}$.

## 2.2 Proposed Strata Multiple Frame Estimator

This paper examines the extension of the Hartley estimator to a stratum-by-stratum combination of estimates from the two frames. Tract operators from the area frame who are found on the list are sampling units in the same livestock stratum as the matching list name. Area frame estimates are then also possible for each list stratum. An estimator, $\hat{y}_S$, is proposed using different $p$ values for each livestock stratum,

$$\hat{y}_S = \hat{y}_{no\ell} + \sum_{h=1}^{k} q_h \hat{y}_{o\ell h} + \sum_{h=1}^{k} p_h \hat{y}_{\ell h}.$$

The stratum-by-stratum multiple frame estimator written in terms of $p$ only is:

$$\hat{y}_S = \hat{y}_A + p_1 (\hat{y}_{\ell 1} - \hat{y}_{o\ell 1}) + p_2 (\hat{y}_{\ell 2} - \hat{y}_{o\ell 2}) + \ldots + p_k (\hat{y}_{\ell k} - \hat{y}_{o\ell k})$$

$$= \hat{y}_A + \sum_{h=1}^{k} p_h (\hat{y}_{\ell h} - \hat{y}_{o\ell h})$$

where $1$ to $k$ are the list stratum numbers from small to large livestock operations. This could be more easily expressed in matrix notation as:

$$\hat{y}_S = \hat{y}_A + \underline{p}' (\hat{\underline{y}}_{\ell} - \hat{\underline{y}}_{o\ell})$$

where $\underline{p}$, $\hat{\underline{y}}_{\ell}$ and $\hat{\underline{y}}_{o\ell}$ are $(kx1)$ vectors of values associated with the respective livestock strata (see Section E of Appendix). Individual tracts within a segment belong to one of the following domains: *NOL*, *OL* list stratum *1*, *OL* list stratum *2*, . . ., *OL* list stratum *k*. Domain estimates are not independent in the area frame since the tracts are all part of the same primary sampling unit, the segment. Therefore, covariances between the domains must be considered in variance computations involving the combination of these domains within the area frame. The variance of $\hat{y}_S$ is shown in Appendix Section F. The optimum $p_{(kx1)}$ vector, derived in Section G of the Appendix, is expressed as:

$$\underline{p}_{opt} = (\Sigma_{y_\ell} + \Sigma_{y_{o\ell}})^{-1} \quad Cov\ (\hat{y}_A, \underline{\hat{y}}_{o\ell})$$

where

$$\Sigma_{y_\ell} = \begin{pmatrix} Var(\hat{y}_{\ell 1}) & 0 & 0 \ \cdots\cdots & 0 \\ 0 & Var(\hat{y}_{\ell 2}) & 0 \ \cdots\cdots & 0 \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ \cdot & \cdot & \cdot & \cdot \\ 0 & 0 & 0 & Var(\hat{y}_{\ell k}) \end{pmatrix}$$

$$\Sigma_{y_{o\ell}} = \begin{pmatrix} Var(\hat{y}_{o\ell 1}) & Cov(\hat{y}_{o\ell 1}, \hat{y}_{o\ell 2}) & \cdots\cdots & Cov(\hat{y}_{o\ell 1}, \hat{y}_{o\ell k}) \\ Cov(\hat{y}_{o\ell 1}, \hat{y}_{o\ell 2}) & Var(\hat{y}_{o\ell 2}) & \cdots\cdots & Cov(\hat{y}_{o\ell 2}, \hat{y}_{o\ell k}) \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ \cdot & \cdot & & \cdot \\ Cov(\hat{y}_{o\ell 1}, \hat{y}_{o\ell k}) & Cov(\hat{y}_{o\ell 2}, \hat{y}_{o\ell k}) & \cdots\cdots & Var(\hat{y}_{o\ell k}) \end{pmatrix}$$

and

$$Cov(\hat{y}_A, \underline{\hat{y}}_{o\ell}) = \begin{pmatrix} Cov(\hat{y}_A, \hat{y}_{o\ell 1}) \\ Cov(\hat{y}_A, \hat{y}_{o\ell 2}) \\ \cdot \\ \cdot \\ \cdot \\ Cov(\hat{y}_A, \hat{y}_{o\ell k}) \end{pmatrix}$$

The variance of this stratum-by-stratum estimator with $\underline{p}_{opt}$ is:

$$Var(\hat{Y}_S) = Var(\hat{Y}_S) - \underline{p}'_{opt} (\Sigma_{Y_\ell} + \Sigma_{Y_{o\ell}}) \underline{p}_{opt}$$

The derivation of this variance expression is shown in Section H of the Appendix. It is also possible to prove $Var(\hat{Y}_S) \leq Var(\hat{Y}_H)$ so that the strata multiple frame estimator is always better or at least as good as the Hartley estimator (see Section I of the Appendix).

## 3. Empirical Study

Formulas for computing the components of the estimators are presented in Section J of the Appendix. To illustrate how the four estimators $\hat{Y}_A$, $\hat{Y}_L$, $\hat{Y}_H$, and $\hat{Y}_S$ compare, data for cattle and hogs from June 1974 area and list frame surveys conducted by the Statistical Reporting Service in two states were used to obtain totals and variances by state for each estimator. The estimates and standard errors for each frame estimating the overlap domain are presented by livestock stratum in Tables 1A and 1B for each state. Note that estimates and standard errors vary considerably between frames for the same stratum. In general, standard errors are smaller for the list frame estimates and are much smaller in the higher strata.

TABLE 1A--List and Area Hog and Pig Inventory Estimates by Livestock Stratum, Two States, June, 1974

| Hogs and Pigs | Overlap Domain | | | | | |
|---|---|---|---|---|---|---|
| | List ($\hat{y}_{\ell}$) | | | Area ($\hat{y}_{o\ell}$) | | |
| Multiple Frame Strata | Stratum Estimates | Standard Error | Coefficient of Variation | Stratum Estimates | Standard Error | Coefficient of Variation |
| | (000) | (000) | (%) | (000) | (000) | (%) |
| <u>State A</u> | | | | | | |
| 1 (No Livestock) | 67.0 | 24.0 | 35.9 | 101.7 | 42.2 | 41.5 |
| 2 (Zero Hogs Pos. Cattle) | 52.1 | 18.1 | 34.7 | 139.5 | 55.3 | 39.7 |
| 3 (1-124 Hogs) | 506.5 | 38.6 | 7.6 | 621.9 | 130.5 | 21.0 |
| 4 (125-199 Hogs) | 559.4 | 37.4 | 6.7 | 450.4 | 103.9 | 23.1 |
| 5 (200-299 Hogs) | 425.5 | 31.4 | 7.4 | 325.0 | 103.2 | 31.8 |
| 6 (300+ Hogs) | 727.0 | 40.6 | 5.6 | 830.6 | 228.0 | 27.5 |
| <u>State B</u> | | | | | | |
| 1 (Unknown) | 44.0 | 8.5 | 19.3 | 44.7 | 22.8 | 51.1 |
| 2 (0-9 Hogs) | 305.4 | 70.0 | 22.9 | 200.3 | 50.5 | 25.2 |
| 3 (10-49 Hogs) | 140.1 | 26.9 | 19.2 | 195.1 | 76.6 | 40.8 |
| 4 (50-499 Hogs) | 314.4 | 24.3 | 7.7 | 328.9 | 110.8 | 33.7 |

TABLE 1B--List and Area Cattle and Calf Inventory Estimates by Livestock Stratum, Two States, June, 1974

| Cattle and Calves | Overlap Domain | | | | | |
|---|---|---|---|---|---|---|
| | List ($\hat{Y}_{\ell}$) | | | Area ($\hat{Y}_{o\ell}$) | | |
| Multiple Frame Strata | Stratum Estimates | Standard Error | Coefficient of Variation | Stratum Estimates | Standard Error | Coefficient of Variation |
| | (000) | (000) | (%) | (000) | (000) | (%) |
| State A | | | | | | |
| 1 (No Livestock) | 327.5 | 80.6 | 24.6 | 558.4 | 111.7 | 20.0 |
| 2 (0-24 Cattle) | 907.2 | 73.1 | 8.1 | 1135.6 | 141.9 | 12.5 |
| 3 (25-49 Cattle) | 1091.5 | 105.2 | 9.6 | 1279.7 | 206.1 | 16.1 |
| 4 (50-99 Cattle) | 1274.4 | 64.5 | 5.1 | 1589.4 | 268.6 | 16.9 |
| 5 (100-199 Cattle) | 1091.6 | 52.5 | 4.8 | 1024.8 | 209.1 | 20.4 |
| 6 (200+ Cattle) | 1239.8 | 55.5 | 4.5 | 1280.0 | 216.3 | 16.9 |
| State B | | | | | | |
| 1 (Unknown) | 363.6 | 56.5 | 15.5 | 175.8 | 33.2 | 18.9 |
| 2 (0-9 Cattle) | 436.3 | 69.4 | 15.9 | 327.3 | 60.2 | 18.4 |
| 3 (10-49 Cattle) | 1179.8 | 53.1 | 4.5 | 975.2 | 111.1 | 11.4 |
| 4 (50-499 Cattle) | 1334.8 | 56.7 | 4.2 | 1263.7 | 180.9 | 14.3 |

The variances of $\hat{y}_H$ and $\hat{y}_S$ are determined by $Var(\hat{y}_A)-p_{opt}[Cov(\hat{y}_A,\hat{y}_{o\ell})]$ and $Var(\hat{y}_A)-\underline{p}'_{opt}[Cov(\hat{y}_A,\underline{\hat{y}}_{o\ell})]$ respectively. The $p_{opt}$ values, correlations between segment totals and overlap domain, and the amounts by which $Var(\hat{y}_A)$ is reduced to equal $Var(\hat{y}_H)$ and $Var(\hat{y}_S)$ are presented for the estimators $\hat{y}_S$ and $\hat{y}_H$ in Tables 2A and 2B. Values of $p_{opt}$ differ considerably between strata for the $\hat{y}_S$ estimator and from the $p_{opt}$ obtained for $\hat{y}_H$. The value of $Cov(\hat{y}_A,\hat{y}_{o\ell})$ summed over the strata is the same as for $\hat{y}_H$. Therefore, the $p_{opt}$ values optimized on a stratum-by-stratum basis are weighted by the $Cov(\hat{y}_A,\hat{y}_{o\ell})$ values for each stratum. The larger the weighted value of $p_{opt}$ for $\hat{y}_S$ compared to the unweighted $p_{opt}$ of $\hat{y}_H$ the smaller the variance of $\hat{y}_S$ relative to $Var(\hat{y}_H)$.

Some of the individual stratum $p_{opt}$ values are greater than 1 which in effect gives a negative weight to the area frame. The size of $p_{opt}$ depends upon $Var(\hat{y}_{o\ell})$ + $Cov(\hat{y}_{no\ell},\hat{y}_{o\ell})$ relative to $Var(\hat{y}_{\ell})$ + $Var(\hat{y}_{o\ell})$. If the covariance between the nonoverlap domain and overlap domain is larger than $Var(y_{\ell})$ then $p_{opt}$ can exceed 1.

It is also noteworthy that for both hogs and cattle the contribution of the small livestock strata in reducing the total variance was minimal. This was due to smaller overlap domain standard errors for these strata relative to the higher strata and poor correlation between $\hat{y}_A$ and $y_{o\ell}$ in the lower strata.

TABLE 2A--Multiple Frame Parameters for $\hat{y}_H$ and $\hat{y}_S$, State A, June 1974

| Estimator | Stratum 1/ | $p_{opt}$ 2/ | $Corr(\hat{y}_A, \hat{y}_{o\ell})$ 3/ | $p_{opt}[Cov(\hat{y}_A, \hat{y}_{o\ell})]$ 4/ | Percent of $Var(\hat{y}_A)$ 5/ |
|---|---|---|---|---|---|
| | | | | (000,000) | (%) |
| Hogs | 1 | .635 | .045 | 459 | - |
| | 2 | .875 | .190 | 3,468 | 2 |
| | 3 | .971 | .341 | 16,311 | 12 |
| | 4 | .849 | .354 | 11,782 | 8 |
| | 5 | .832 | .253 | 8,178 | 6 |
| | 6 | 1.068 | .673 | 61,746 | 43 |
| $\hat{y}_S$ | Total | .990 6/ | - | 101,944 | 71 |
| $\hat{y}_H$ | Overall | .979 7/ | .869 | 100,838 | 70 |
| Cattle | 1 | .554 | .044 | 1,175 | 1 |
| | 2 | .746 | .257 | 11,828 | 6 |
| | 3 | .760 | .366 | 24,888 | 13 |
| | 4 | .893 | .480 | 49,927 | 26 |
| | 5 | .856 | .384 | 29,835 | 16 |
| | 6 | .886 | .349 | 29,049 | 15 |
| $\hat{y}_S$ | Total | .842 6/ | - | 146,702 | 77 |
| $\hat{y}_H$ | Overall | .827 7/ | .952 | 144,057 | 76 |

1/ List frame strata based on number of head in the operation.
2/ Weight attached to the list frame estimate.
3/ Correlation between area frame total and area frame overlap domain.
4/ Amount by which $Var\hat{y}_A$ is reduced to equal $Var\hat{y}_S$ and $Var\hat{y}_H$.
5/ $p_{opt}[Cov(\hat{y}_A, \hat{y}_{o\ell})]$ as percent of $Var(\hat{y}_A)$ for percentage reduction from $Var(\hat{y}_A)$
6/ Weighted value of $p_{opt}$ for $\hat{y}_S$ from weighting individual stratum $p_{opt}$ by stratum $Cov(\hat{y}_A, \hat{y}_{o\ell})$.
7/ Value of $p_{opt}$ for $\hat{y}_H$ ignoring strata.

TABLE 2B--Multiple Frame Parameters for Estimates $\hat{y}_H$ and $\hat{y}_S$, State B, June 1974

| Estimator | Stratum 1/ | $p_{opt}$ 2/ | $Corr(\hat{y}_A,\hat{y}_{o\ell})$ 3/ | $p_{opt}[Cov(\hat{y}_A,\hat{y}_{o\ell})]$ 4/ | Percent of $Var(\hat{y}_A)$ 5/ |
|---|---|---|---|---|---|
| | | | | (000,000) | (%) |
| Hogs | 1 | 1.216 | .187 | 918 | 3 |
| | 2 | .324 | .300 | 862 | 3 |
| | 3 | .989 | .583 | 7,806 | 25 |
| | 4 | .995 | .689 | 13,415 | 42 |
| $\hat{y}_S$ | Total | .927 6/ | - | 23,001 | 73 |
| $\hat{y}_H$ | Overall | .829 7/ | .913 | 20,556 | 65 |
| Cattle | 1 | .288 | .152 | 337 | 1 |
| | 2 | .406 | .201 | 1,139 | 2 |
| | 3 | .806 | .388 | 8,034 | 15 |
| | 4 | .915 | .719 | 27,517 | 51 |
| $\hat{y}_S$ | Total | .841 6/ | - | 37,027 | 69 |
| $\hat{y}_H$ | Overall | .774 7/ | .920 | 34,077 | 64 |

1/ List frame strata based on number of head in the operation.
2/ Weight attached to the list frame estimate.
3/ Correlation between area frame total and area frame overlap domain.
4/ Amount by which $Var(\hat{y}_A)$ is reduced to equal $Var(\hat{y}_S)$ and $Var(\hat{y}_H)$.
5/ $p_{opt}$ $[Cov(\hat{y}_A,\hat{y}_{o\ell})]$ as percent of $Var(\hat{y}_A)$.
6/ Weighted value of $p_{opt}$ for $\hat{y}_S$ from weighting individual stratum $p_{opt}$ by stratum $Cov(\hat{y}_A,\hat{y}_{o\ell})$.
7/ Value of $p_{opt}$ for $\hat{y}_H$ ignoring strata.

Except for Stratum 1, State B Hogs, the weight for the list frame, $p_{opt}$, was lower for the smaller strata. Even when the list frame weight was greater than one in Stratum 1, the reduction in total variance was small. The resulting minimal reduction in variance realized by sampling both frames makes the practice of gathering data in both frames questionable for these strata. The stratum-by-stratum multiple frame procedure provides a means of measuring the contribution of the frames for each stratum.

The combinations of estimates from each of the two frames into the various multiple frame estimates are presented in Table 3. There was little difference between results for the screening estimator $(\hat{Y}_L)$ and the Hartley estimator $(\hat{Y}_H)$. The weight attached to the list frame was so dominant that very little reduction in sampling error was realized from the contribution of the area frame overlap domain.

TABLE 3--Multiple Frame Livestock Estimates Using Alternative Estimators - June, 1974

| Multiple Frame Estimator | Hogs | | | Cattle | | |
|---|---|---|---|---|---|---|
| | Estimate | Standard Error | Coefficient of Variation | Estimate | Standard Error | Coefficient of Variation |
| | (000) | (000) | (%) | (000) | (000) | (%) |
| State A | | | | | | |
| $\hat{y}_A$ (*Area*) | 3540.6 | 378.0 | 10.7 | 8597.0 | 436.3 | 5.1 |
| $\hat{y}_L$ (*Screening*) | 3409.1 | 205.2 | 6.0 | 7660.2 | 228.7 | 3.0 |
| $\hat{y}_H$ (*Hartley*) | 3411.8 | 204.9 | 6.0 | 7822.7 | 215.0 | 2.8 |
| $\hat{y}_S$ (*Strata*) | 3395.6 | 202.3 | 5.9 | 7895.7 | 209.0 | 2.6 |
| State B | | | | | | |
| $\hat{y}_A$ (*Area*) | 1301.1 | 177.4 | 13.6 | 3615.4 | 231.4 | 6.4 |
| $\hat{y}_L$ (*Screening*) | 1299.9 | 106.6 | 8.4 | 4188.9 | 149.6 | 3.6 |
| $\hat{y}_H$ (*Hartley*) | 1300.1 | 104.4 | 8.0 | 4067.1 | 139.5 | 3.4 |
| $\hat{y}_S$ (*Strata*) | 1265.4 | 92.0 | 7.3 | 3952.2 | 128.6 | 3.3 |

The $\hat{Y}_S$ estimates for hogs in both states are slightly below the other multiple frame estimates while the $\hat{Y}_S$ cattle estimates are between $\hat{Y}_A$ and $\hat{Y}_L$, and near $\hat{Y}_H$. This reflects the relative size of the area frame estimate and its weight compared to the list frame estimate and weight by stratum. The estimates for each frame are shown in Tables 1A and 1B, while the weights, $p_{opt}$, are shown in Tables 2A and 2B.

The stratum-by-stratum combination of area and list frame estimates results in smaller sampling errors for $\hat{Y}_S$ than for either $\hat{Y}_L$ or $\hat{Y}_H$. The decrease in sampling error for the strata estimator, $\hat{Y}_S$, is especially evident in State B where both hog and cattle standard errors were about 14% lower than for the screening estimator, $\hat{Y}_L$, and approximately 12% and 8% respectively below $\hat{Y}_H$. State A reductions were 1.5% and 8.6% for hogs and cattle respectively compared to the screening estimator and about 1% and 3% below $\hat{Y}_H$.

In summary, reductions in sampling errors are more pronounced using a multiple frame estimator with different $p$ values for each stratum in the overlap domain than for the screening on Hartley estimators which have one value for $p$. Independent stratum estimates from each domain are combined in an optimum, objective manner. Weights attached to the list frame, $p$ values, differ considerably between strata as do the covariances between segment totals and overlap domain totals, both of which are important to the reduction of variance. Most of the reduction in variance resulting from the combination of area and list frames by strata occurs in the larger size group livestock strata.

# APPENDIX

## A. Alternative Expressions of Hartley Estimator

$$\hat{y}_H = \hat{y}_{no\ell} + q\hat{y}_{o\ell} + p\hat{y}_{\ell}$$

$$= \hat{y}_{no\ell} + (1-p)\ \hat{y}_{o\ell} + p\hat{y}_{\ell}$$

$$= \hat{y}_{no\ell} + \hat{y}_{o\ell} + p(\hat{y}_{\ell} - \hat{y}_{o\ell})$$

$$= \hat{y}_A + p(\hat{y}_{\ell} - \hat{y}_{o\ell})$$

## B. Variance of Hartley Estimator

$$Var(\hat{y}_H) = Var(\hat{y}_{no\ell}) + q^2 Var(\hat{y}_{o\ell}) + p^2 Var(\hat{y}_{\ell}) + 2q\ Cov(\hat{y}_{no\ell}, \hat{y}_{o\ell})$$

$$= [Var(\hat{y}_{no\ell}) + Var(\hat{y}_{o\ell}) + 2Cov(\hat{y}_{no\ell}, \hat{y}_{o\ell})] - 2p[Var(\hat{y}_{o\ell}) + Cov(\hat{y}_{no\ell}, \hat{y}_{o\ell})]$$

$$+ p^2[Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell})]$$

$$= Var(\hat{y}_A) - 2pCov(\hat{y}_A, \hat{y}_{o\ell}) + p^2[Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell})]$$

## C. Derivation of Optimum $p$

$$\frac{\partial Var(\hat{y}_H)}{\partial p} = -2\ Cov(\hat{y}_A, \hat{y}_{o\ell}) + 2p(Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell}))\ \underline{\underline{Set}}\ 0$$

$$p_{opt} = \frac{Cov(\hat{y}_A, \hat{y}_{o\ell})}{Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell})} = \frac{Var(\hat{y}_{o\ell}) + Cov(\hat{y}_{no\ell}, \hat{y}_{o\ell})}{Var(\hat{y}_{\ell}) + Var(\hat{y}_{o\ell})}$$

D. Alternative Expressions of Hartley Variance Using $p_{opt}$

$$Var(\hat{y}_H)=Var(\hat{y}_A)-2p_{opt}[Cov(\hat{y}_A,\hat{y}_{o\ell})]+p^2_{opt}[Var(\hat{y}_\ell)+Var(\hat{y}_{o\ell})]$$

$$=Var(\hat{y}_A)-2p_{opt}[Cov(\hat{y}_A,\hat{y}_{o\ell})]+p_{opt}[Cov(\hat{y}_A,\hat{y}_{o\ell})]$$

$$=Var(\hat{y}_A)-p_{opt}[Cov(\hat{y}_A,\hat{y}_{o\ell})]$$

or

$$=Var(\hat{y}_A)-2p^2_{opt}[Var(\hat{y}_\ell)+Var(\hat{y}_{o\ell})]+p^2_{opt}[Var(\hat{y}_\ell)+Var(\hat{y}_{o\ell})]$$

$$=Var(\hat{y}_A)-p^2_{opt}[Var(\hat{y}_\ell)+Var(\hat{y}_{o\ell})]$$

E. Extension of Hartley Estimator to Stratified Case

$$\hat{y}_S = \hat{y}_A+p_1(\hat{y}_{\ell 1}-\hat{y}_{o\ell 1})+p_2(\hat{y}_{\ell 2}-\hat{y}_{o\ell 2})+ \ . \ . \ . \ +p_k(\hat{y}_{\ell k}-\hat{y}_{o\ell k})$$

$$= \hat{y}_A+(p_1\ p_2\ .\ .\ .\ p_k) \begin{pmatrix} \hat{y}_{\ell 1}-\hat{y}_{o\ell 1} \\ \hat{y}_{\ell 2}-\hat{y}_{o\ell 2} \\ \vdots \\ \hat{y}_{\ell k}-\hat{y}_{o\ell k} \end{pmatrix}$$

$$= \hat{y}_A + \underline{p}'(\underline{\hat{y}}_\ell-\underline{\hat{y}}_{o\ell})$$

F. Variance of Estimator Using Stratification

$$Var(\hat{y}_S)=Var(\hat{y}_A)+\underline{p}'(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})\underline{p}+2Cov[\hat{y}_A,\ \underline{p}'(\underline{\hat{y}}_\ell-\underline{\hat{y}}_{o\ell})]$$

$$= Var(\hat{y}_A)+\underline{p}'(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})\underline{p}-2\underline{p}'Cov(\hat{y}_A,\hat{\underline{y}}_{o\ell})$$

where $\Sigma_{\hat{y}_\ell}$ and $\Sigma_{\hat{y}_{o\ell}}$ are $kxk$ covariance matrices.

G. Optimum $\underline{p}$, $\underline{p}_{opt}$

$$\underline{p}_{opt} = \frac{\partial Var(\hat{y}_S)}{\partial \underline{p}} = 2(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})\underline{p} - 2Cov(\hat{y}_A,\hat{\underline{y}}_{o\ell}) \;\underline{\underline{\underline{Set}}}\; 0$$

$$\underline{p}_{opt} = (\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})^{-1}\, Cov(\hat{y}_A,\hat{\underline{y}}_{o\ell})$$

H. Alternative Expression for $Var(\hat{y}_S)$ with $\underline{p}_{opt}$

$$Var(\hat{y}_S) = Var(\hat{y}_A)+\underline{p}_{opt}'(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})\;(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})^{-1}\, Cov(\hat{y}_a,\hat{\underline{y}}_{o\ell})-2\underline{p}_{opt}'\, Cov(\hat{y}_a,\hat{\underline{y}}_{o\ell})$$

$$= Var(\hat{y}_A) - \underline{p}_{opt}'\, Cov(\hat{y}_a,\hat{\underline{y}}_{o\ell})$$

or

$$= Var(\hat{y}_A)+\underline{p}_{opt}'(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})\underline{p}_{opt}-2\underline{p}_{opt}'(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})\underline{p}_{opt}'$$

$$= Var(\hat{y}_A)-\underline{p}_{opt}'(\Sigma_{\hat{y}_\ell}+\Sigma_{\hat{y}_{o\ell}})\underline{p}_{opt}$$

I. Proof that $Var(\hat{y}_S) \leq Var(\hat{y}_H)$

A theorem from Rao (2.) states that if $A$ is a positive definite $mxm$ matrix and $\underline{\mu}$ and $\underline{x}$ are m-vectors, then:

$$\frac{(\underline{\mu}'\ \underline{x})^2}{\underline{x}'A\underline{x}} \leq \underline{\mu}'\ A^{-1}\ \underline{\mu}$$

based on the Couchy-Schwartz inequality.

Rewriting the variance of Hartley's estimator from Appendix Section D as:

$$Var(\hat{Y}_H) = Var(\hat{Y}_A) - \frac{[Cov(\hat{Y}_A,\hat{Y}_{o\ell})]^2}{[Var(\hat{Y}_\ell) + Var(\hat{Y}_{o\ell})]}$$

And rewriting the variance of the stratum estimator from Appendix Section H as:

$$Var(\hat{Y}_S) = Var(\hat{Y}_A) - Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})'[\Sigma_{\hat{Y}_\ell}+\Sigma_{\hat{Y}_{o\ell}}]^{-1}\ Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})$$

the proof is dependent upon:

$$\frac{[Cov(\hat{Y}_A,\hat{Y}_{o\ell})]^2}{[Var(\hat{Y}_\ell)+Var(\hat{Y}_{o\ell})]} \leq Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})'\ [\Sigma_{\hat{Y}_\ell}+\Sigma_{\hat{Y}_{o\ell}}]^{-1}\ Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell}).$$

Since $\Sigma_{\hat{Y}_\ell}$ and $\Sigma_{\hat{Y}_{o\ell}}$ are positive definite so $[\Sigma_{\hat{Y}_\ell}+\Sigma_{\hat{Y}_{o\ell}}]^{-1}$ is positive definite, it only remains to be shown that the inequality is consistent with the theorem. Letting $\underline{\mu} = Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})$, $A = [\Sigma_{\hat{Y}_\ell}+\Sigma_{\hat{Y}_{o\ell}}]$ and $\underline{x} = \underline{1}$, and expressing $[Var(\hat{Y}_\ell) + Var(\hat{Y}_{o\ell})]$ as $\underline{1}'[\Sigma_{\hat{Y}_\ell}+\Sigma_{\hat{Y}_{o\ell}}]\underline{1}$ and $Cov(\hat{Y}_A,\hat{Y}_{o\ell}) = Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})'\underline{1}$ gives the inequality:

$$\frac{[Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})'\underline{1}]^2}{\underline{1}'[\Sigma_{\hat{Y}_\ell}+\Sigma_{\hat{Y}_{o\ell}}]\underline{1}} \leq Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})'\ [\Sigma_{\hat{Y}_\ell}+\Sigma_{\hat{Y}_{o\ell}}]^{-1}\ Cov(\hat{Y}_A,\underline{\hat{Y}}_{o\ell})$$

which is the same expression as the theorem above and this completes the proof that $Var(\hat{Y}_S) \leq Var(\hat{Y}_H)$.

J. Computation Formulas

Area

$y_{ij}$ = Number of head on tract $j$ of segment $i$.

$$y_i = \sum_{j=1}^{m_i} y_{ij}$$ where $m_i$ = number of tracts in $i^{th}$ segment.

= total number of head in segment $i$

$$\hat{y}_A = \frac{N}{n} \sum_{i=1}^{n} y_i,$$ where $n$ = number of sample segments and $N$ = population of segments in the state.

= estimated total in the state.

$$Var(\hat{y}_A) = \frac{N^2}{n} \left[ \frac{\sum_{i=1}^{n} (y_i - \bar{y})^2}{n-1} \right] \frac{N-n}{n} .$$

$$_{no\ell}y_i = \sum_{j=1}^{m_i} y_{ij}$$ where $y_{ij} = \begin{cases} \text{Number of head on tract } j \text{ if operator is not on the list.} \\ 0 \text{ if operator is on the list.} \end{cases}$

= segment total livestock on nonoverlap tracts

$$\hat{y}_{no\ell} = \frac{N}{n} \sum_{i=1}^{n} {}_{no\ell}y_i$$ = Estimated total livestock on land operated by person(s) not on the list.

$$Var(\hat{y}_{no\ell}) = \frac{N^2}{n} \left[ \frac{\sum_{i=1}^{n} ({}_{no\ell}y_i - {}_{no\ell}\bar{y})^2}{n-1} \right] \frac{N-n}{N}$$

$$_{o\ell}y_i = \sum_{j=1}^{m_i} y_{ij},$$ where $y_{ij} = \begin{cases} \text{Number of head on tract } j \text{ if operator is found on the list.} \\ 0 \text{ if operator is not on the list.} \end{cases}$

= segment total livestock on overlap tracts

$$\hat{y}_{o\ell} = \frac{N}{n} \sum_{i=1}^{n} {}_{o\ell}y_i$$ = Estimated total livestock on land operated by person(s) on the list.

$$Var(\hat{y}_{o\ell}) = \frac{N^2}{n} \left[ \frac{\sum_{i=1}^{n} ({}_{o\ell}y_i - {}_{o\ell}\overline{y})^2}{n-1} \right] \frac{N-n}{N}$$

$$Cov(\hat{y}_A, \hat{y}_{o\ell}) = \frac{N^2}{n} \left[ \frac{\sum_{i=1}^{n} (y_i - \overline{y}) ({}_{o\ell}y_i - {}_{o\ell}\overline{y})}{n-1} \right] \frac{N-n}{N}$$

$${}_{o\ell h}y_i = \sum_{j=1}^{m_i} y_{ij}, \text{ where } y_{ij} = \begin{cases} \text{Number of head on tract } j \text{ if name of operator is found on the list in stratum } h, \\ 0 \text{ if operator is not in list stratum } h. \end{cases}$$

$$\hat{y}_{o\ell h} = \frac{N}{n} \sum_{i=1}^{n} {}_{o\ell h}y_i$$ = Estimated total livestock on land operated by person(s) on the list in stratum $h$.

$$Var(\hat{y}_{o\ell h}) = \frac{N^2}{n} \left[ \frac{\sum_{i=1}^{n} ({}_{o\ell h}y_i - {}_{o\ell h}\overline{y})^2}{n-1} \right] \frac{N-n}{N}$$

$$Cov(\hat{y}_{o\ell h}, \hat{y}_{o\ell h+1}) = \frac{N^2}{n} \left[ \frac{\sum_{i=1}^{n} ({}_{o\ell h}y_i - {}_{o\ell h}\overline{y}) ({}_{o\ell h+1}y_i - {}_{o\ell h+1}\overline{y})}{n-1} \right] \frac{N-n}{N}$$

List

$y_{\ell hx}$ = Total number of head on land operated by list name $x$.

$$\hat{Y}_{\ell h} = \frac{N_h}{n_h} \sum_{x=1}^{n_h} y_{\ell h x}$$ where $n_h$ = sample number of names selected from list stratum $h$ and $N_h$ is population size of stratum $h$.

= Estimated total number of livestock in list stratum $h$.

$$Var(\hat{Y}_{\ell h}) = \frac{N_h^2}{n_h} \left[ \frac{\sum_{x=1}^{n_h} (y_{\ell h x} - \overline{y}_{\ell h})^2}{n_h - 1} \right] \frac{N_h - n_h}{N_h}$$

$$\hat{Y}_{\ell} = \sum_{h=1}^{k} \hat{Y}_{\ell h}$$ = Estimated total livestock in the list domain.

$$Var\hat{Y}_{\ell} = \sum_{h=1}^{k} Var(\hat{Y}_{\ell h})$$

## REFERENCES


1. Hartley, H. O., (1962) "Multiple Frame Surveys." Proceedings of the Social Science Section of the American Statistical Association Meeting, Minneapolis, Minnesota.

2. Rao, C. R., (1973) Linear Statistical Inference and Its Applications, (p.60), John Wiley & Sons, New York, New York

3. Fuller, Wayne A. and Burmeister, Leon F., (1973) "Estimates for Samples Selected From Two Overlapping Frames." A report for the U.S. Bureau of the Census, Iowa State University, Ames, Iowa.

4. Cochran, Robert S., (1964) "Multiple Frame Sample Surveys." Proceedings of the Social Science Section of the American Statistical Association Meeting, Chicago, Illinois.

5. Vogel, F. and Bosecker, R., (1974) "Multiple Frame Livestock Surveys: A Comparison of Area and List Sampling," Statistical Reporting Service, USDA

6. Vogel, F. and Thorson, J., (1975) "Optimizing the Use of List and Area Samples in Livestock Multiple Frame Surveys," Statistical Reporting Service, USDA