United States Department of Agriculture

National Agricultural Statistics Service

Research and Applications Division

SRB Staff Report Number SRB-90-08

May 1990

MATHEMATICAL FORMULAE FOR THE 1989 SURVEY PROCESSING SYSTEM (SPS) SUMMARY

Phillip S. Kott

MATHEMATICAL FORMULAE FOR THE 1989 SURVEY PROCESSING SYSTEM (SPS) SUMMARY, by Phillip S. Kott, Research and Applications Division, National Agricultural Statistics Service, U.S. Department of Agriculture, Washington, DC 20250, May 1990. NASS Staff Report, Number SRB-90-08.

ABSTRACT

The 1989 Survey Processing System (SPS) Summary was designed to summarize the results of the six agricultural surveys from June 1989 to March 1990, the four labor surveys from July 1989 to April 1990, and the August 1990 rice stocks survey. This report is primarily an attempt to document statistical formulae incorporated into the 1989 SPS Summary. The focus is on providing equations for the expansions, ratios, standard errors, and coefficients of variation found in SPS Summary Tables 3 through 7. Although the emphasis is on stating the formulae not on justifying their use, informal arguments for particular formulations will be given - especially when there are no other readily available references.

ACKNOWLEDGMENTS

The report has had many drafts. In the course of making the many refinements that ultimately lead to a more accurate and (I hope) useful product, I have had the help of very many talented and knowledgable people. Among them were Huong Luong, Bill Iwig, Martha Roehrenbeck, Mike Clark, Ned Jones, Bob Hale, Jerry Thorson, Brad Pafford, Carol House, Dave Dillard, Tim Placke, Rick Kestle, and Barry Ford. I thank them all. In particular I would like to thank Jack Nealon for his early and continuous support of this endeavor. Jack reviewed about half of the 11 drafts and supplied the quote at the bottom of this page. I apologize to all those I have missed.

May 1990

Washington, DC

"I don't believe in mathematics."
- Albert Einstein

TABLE OF CONTENTS

AUTHOR'S FOREWORD

When the Survey Processing System Summary team was organized, I as the representative of the Survey Research Branch (SRB) was given the task of making sure the mathematical formulae incorporated into the new Summary were correct. It seemed to me that the best way to accomplish that task was to create a report spelling out exactly what those formulae were. Now over a year later the report in ready for publication. It is being issued as an SRB staff report bearing the year "1989" in its title. It is my sincere hope that this report will be updated every year, although whether that task should remain the responsibility of the SRB remains to be decided.

Phillip S. Kott
April 30, 1990
Washington, DC

I. INTRODUCTION

The 1989 Survey Processing System (SPS) Summary was designed (in principle) to summarize the results of the six agricultural surveys from June 1989 to March 1990, the four labor surveys from July 1989 to April 1990, and the August 1990 rice stocks survey. See [1] and [2] for more information on 1989 agricultural and labor surveys (note: "1989" refers to the year of the frame not necessarily the year of the survey).

The SPS Summary is by its very nature dynamic. It will change over the years as the National Agricultural Statistics Service (NASS) extends its use to reimbursable surveys and employs more sophisticated sampling and estimation strategies.

This report is primarily an attempt to document statistical formulae incorporated into the 1989 SPS Summary. The focus is on providing equations for the expansions, ratios, standard errors, and coefficients of variation found in SPS Summary Tables 3 through 7. Although the emphasis is on stating the formulae not on justifying their use, informal arguments for particular formulations will be given - especially when there are no readily available references.

The terminology used in this report does not always conform to standard NASS practice, which is constantly evolving and therefore the subject of some confusion. A glossary of many of the technical terms used here has been appended to the report.

One particular note on terminology is in order at this point. The text and glossary use tne word "estimator" to refer to an estimation formula. Following NASS practice, the word "indication" is used in this report in place of "estimate" to denote the result of applying an estimator to a set of survey values.

More information on the SPS Summary, including topics not discussed here (for example, Summary Tables 1 and 2 and exact Table terminology) and those covered only superficially (like unexpanded counts), can be found in [3; Section 8] and [4; Section 8].

II. OVERVIEW

Base and Follow-on Surveys

NASS uses multiple frame estimation strategies to calculate many of its indications. This is because most of NASS's lists of potential farms are incomplete, while estimates based purely on area samples are often statistically inefficient.

The June Agricultural Survey is called the base survey. The base survey is composed of a (stratified) random sample of area segments (see [5] or [6] for a description of the selection process) and a list sample of names associated with potential farms. Sometimes the area segments are used by themselves to make estimates for the states (for numbers of farms, land in farms, and crop acreages). More often they are combined with estimates calculated from list frames.

All other surveys covered by the 1989 SPS Summary employ multiple frame estimators exclusively. These follow-on surveys rely on subsamples of tracts from the June area sample to account for agriculture from farms not on NASS list frames.

Multiple Frame Estimators

As noted already, all follow-on surveys use multiple frame (MF) estimators. For the purpose of the 1989 SPS Summary, they are all dual-frame estimators (list frames coupled with area frames) with the exception of the agricultural services survey (a component of the labor survey) in California and Florida. The nature of that survey will be explained briefly in a later section of this chapter and with more detail in Chapter VIII.

There are five different types of MF estimators used with the base survey (more details on the terms overlap, nonoverlap, extreme operator, non-extreme operator, closed, open, and weighted can be found in later sections and the glossary):

1. The overlap MF with closed nonoverlap (the OL-MF/closed) estimator -- for indications of crop acreages.

2. The overlap MF with weighted nonoverlap (the OL-MF/weighted) estimator -- for indications of grain stocks, hogs, and cropland acreages in 10 states.

3. The extreme operator MF estimator with closed non-extreme operators (the EO-MF/closed) estimator -- for indications of hogs and (July) cattle.

4. The EO-MF with open non-extreme operators (the EO-MF/open) -- for indications of hogs in 10 states.

5. The EO-MF with weighted non-extreme operators (the EO-MF/weighted) estimator -- for indications of hogs in 10 states.

The EO-MF/closed estimator for the JULY CATTLE survey combines a July list sample with the June area sample. All other follow-on agricultural indications use the OL-MF/weighted estimator exclusively, while labor indications use the OL-MF/open estimator (heretofore unmentioned, but self-explanatory).

List Frame Estimators

Each of the lower 49 states has at least three list frames -- one for hogs, crops, stocks, and/or chickens (called the hog/crops frame), one for cattle/sheep/goats, and one for basic (or farm) labor. Many states also have trout and/or catfish list frames. For our purposes, the only other frame of interest is the agricultural services frame in California and Florida (there will be more on this frame in a later section).

Each frame is a list of names (associated with potential farms) organized into list strata. The names in a particular stratum are believed to be associated with farms having similar sizes with respect to a relevant agricultural characteristic (e.g., its peak number of cattle for the cattle/sheep/goats frame).

Some of the strata in each hogs/crops and cattle/sheep/goats frame are composed entirely of (presumably) large farms called extreme operators or EO's. Naturally enough, these strata are

called "EO strata."

NASS draws a without replacement simple random sample of names from each stratum in a list frame when that frame is relevant for a survey at hand; for example, a stratified simple random sample is drawn from a state's hogs/crops frame but not from its basic labor frame for the base survey.

After the sampling units are contacted, indications of list frame totals are calculated. More on the mechanics of this is reserved for a later section.

NASS is also interested in hog totals based on hogs/crops EO's sampled in the base survey and cattle totals based on cattle/sheep/goat EO's sampled in the July agricultural survey. This is for NASS's EO-MF estimators.

Area Frame Estimators

A random sample of area segments is selected as part of the base survey. A tract is the land area within a segment under one land operating arrangement. Indications of number of farms, land on farms, and crop acreages are computed using information from the operators of these tracts with the closed (land in farms, crop acreages), open (number of farms, land in farms) and weighted (number of farms, cropland acreages in 10 states) estimators. Precise definitions of these three estimators can be found in [5] or [6]. The glossary may also be of some help.

Tracts that are contained in farms that are on a particular list frame are called overlaps or OL's (with respect to that frame). Conversely, tracts not on farms on a list frame are called nonoverlaps or NOL's. Similarly, tracts can be EO's (extreme operators) or NEO's (non-extreme operators).

Multiple frame estimators combine estimators for a list frame (or its EO's) with area frame estimators. For the base survey, it is necessary to set data values to zero for OL tracts when employing an OL-MF estimator. Likewise, it is necessary to zero out EO's when employing an EO-MF

estimator. For follow-on surveys, only NOL tracts are subsampled, so nothing has to be zeroed out.

We will call the closed estimator that zeroes out OL tracts "the NOL-closed estimator" since using it can provide an indication for total values from farms not in the list frame. The NOL-open, NOL-weighted, NEO-closed, and NEO-weighted estimators are defined in analogous manner.

Indications

This short section tries to map the estimators discussed above with some of the indication terminology commonly used by NASS.

As noted in the introduction, an estimator is applied to survey data to get an indication. Since there are no closed or weighted estimators for hogs based purely on area frame data from the base survey, NASS often calls indications calculated from the EO-MF/closed estimator for hogs "closed indications." Likewise, the EO-MF/open and EO-MF/weighted estimator for hogs are said to produce open and weighted indications, respectively. In the same spirit, the combination of EO's from the July cattle survey and the NEO-closed cattle indication from the base survey is called the "mid-year closed cattle indication."

Another name often used for a closed indication is a "tract indication." An open indication (produced from using the open estimator with area frame data only) is also called a "farm indication."

Finally, an indication from an OL-MF estimator is often called simply an "MF indication," since indications from EO-MF estimators are named after their component NEO estimators; e.g., the closed indication of hogs.

The Labor Survey

The labor survey has two parts: a <u>basic</u> (or farm) labor survey and an <u>agricultural</u> (ag) <u>services</u> survey. The basic labor survey is conducted in all states, while the agricultural services survey is conducted exclusively in California and

Florida. The former is a survey of list names (of potential farms) and NOL tracts, the latter a survey of ag services - providers of labor-related services that need not themselves be farms.

In states other than California and Florida, the reporting units of the basic labor survey also provide information on labor provided to them by ag services. A survey item that involves both farm labor and ag services labor is called a total labor item. In states other than California and Florida, indications for ag services and total labor items are derived in the same manner as farm labor items.

In Chapters III through VII, references to the labor survey apply specifically to the **basic** labor survey. Chapter VIII discusses the ag services survey in California and Florida. It also deals with estimators of total labor items in those two states. (Note: in states other than California and Florida, total labor items are handled as normal survey items of the basic labor survey.)

Another feature of the SPS Summary is the computation of yearly aggregates for certain labor survey items. This is discussed in Chapter IX. It should also be noted that survey-to-survey and base-to-base ratios (dealt with in Chapter V) are not computed for any labor survey items.

III. MATHEMATICAL PRELIMINARIES

Reporting Units, Survey Items, and List Adjustment Factors

NASS draws a distinction between the names or selected units on its list frames and the reporting units or sub-tracts that actually respond to its surveys. Each selected unit is composed of at least one sub-tract. (NASS often calls a reporting unit in the base survey from the list frame a "tract" rather than a "sub-tract." We will not follow that practice here.)

Area tracts are the reporting units for the area frame portion of the base survey. For follow-on surveys, however, it is possible for an area tract (the selected unit) to subdivide into separately reporting sub-tracts.

A survey item value is a quantitative answer to a particular survey question or a function of those answers. One such function is the sum of winter wheat, durum wheat, and spring wheat from the June Agriculture Survey. Two derived survey items (i.e., functions of survey answers) merit particular mention. A domain variable (used in domain estimation) is a function of another survey item value that equals that value only when the respondent is within a certain domain; it is zero otherwise (for example, a respondent's inventory of hogs if he has 1 to 99 hogs, zero if he has no hogs or more than 99 hogs). A count or count variable is a survey item value that can only be zero or one: 1 when a particular survey item value is positive, 0 otherwise.

Let q_{is} be a raw survey item value for sub-tract s of selected unit i, and let

$$y_{is} = L_{is}q_{is}, \qquad (1)$$

where L_{is} is the list adjustment factor (LAF) for sub-tract s (see below).

An LAF is a scalar adjustment to item values that accounts for more than one sub-tract being associated with the same farming arrangement. (Note: L_{is} is equal to 1 for an area tract in the base survey.)

IN THIS REPORT, WE WILL REFER TO y_{is} (not q_{is}) AS THE SURVEY ITEM VALUE of s.

The survey item value for the selected unit i is

$$y_i = \Sigma\, y_{is} \qquad (2)$$

where the summation is over every sub-tract s within selected unit i. (In Chapter V, selected units for follow-on surveys are denoted with the subscript m rather than i.)

Nonresponse Adjustments

NASS has three methods of adjusting to nonresponse caused by a reporting unit being either inaccessible, refusing to participate in the survey at all, failing to answer particular questions, or giving obviously incorrect (i.e., edit failing) answers to particular questions. These methods are hand imputation, machine imputation, and reweighting.

Hand imputation is conducted by NASS usually in a state office. This is the act of filling in an unreported survey value or changing an obviously incorrect value based usually on judgement but it can also happen during the edit program (which means it need not literally be done by hand). Hand imputation occurs before the SPS Summary gets the data. Survey item values based entirely on information provided either by a reporting unit or by hand imputation (this includes LAF information) are called usables. In contrast to previous practice, usability is defined here on an item by item basis. If a survey item is the sum of other items, it is usable only when its components are all usable.

Missing item values for CROPS and STOCKS (except COUNTS) are machine imputated by the imputation program described in detail in [7]. Missing LAF's for crops, stocks, and TOTAL LAND expansions are also imputed. SURVEY VALUES THAT ARE EVEN PARTIALLY MACHINE IMPUTED (including the L_{is} from equation (1)) ARE NOT CONSIDERED USABLES

Reweighting is a process of adjusting the expansion factor in an estimator formula to account for missing responses. The LABOR survey

uses reweighting both in its area frame and list frame estimators. One LIVESTOCK estimator (called the presence estimator) employs a more complicated reweighting scheme on the list side only (all missing livestock values on NOL tracts are hand imputed). There is also direct expansion livestock estimator that uses reweighting in a straightforward manner, again, on the list side only.

Reweighting is done as part of the SPS Summary. Moreover, since variances can not be correctly estimated in the conventional manner when there is machine imputed data (see [8]), variances for crops and stocks are conservatively estimated using the formula based on reweighting (described in the following chapters). Additional errors caused by hand imputation are ignored by the SPS Summary.

Variances, Standard Errors, and CV's

Suppose Z is an estimator used in the SPS Summary, and V is the estimator for the variance of Z. (When Z is an estimated total, it will be denoted by Y. When Z is a ratio, by R.)

The estimator for the standard error of Z is

$$SE = \sqrt{V}, \qquad (3)$$

while the estimator for the coefficient of variation (CV) of Z (in percent terms) is

$$CV = (SE/Z) \times 100. \qquad (4)$$

This report contains formulas for estimators of totals and ratios and for estimators of variances. Equations (3) and (4) show how the estimated standard errors and CV's displayed in SPS Summary tables are derived from estimated variances.

IV. LIST EXPANSIONS

In this chapter, we focus on estimators for the total of a survey item in a list stratum in a particular state. State, region, and U.S. level aggregates will be addressed in Chapter VI.

A SELECTED UNIT VALUE FOR A SURVEY ITEM IS SAID TO BE **USABLE** WHEN ALL ITS SUB-TRACT VALUES FOR THAT ITEM ARE USABLE, which (again) can mean that they have been hand imputed but does not allow for (any part of) them to be machine imputed.

Direct Expansion

There is machine imputation for all nonusable CROP and STOCK sub-tract values (except COUNTS). As a result, both usable and nonusable item values are employed in the estimator of a crop or stock list stratum total. We begin with such an estimator.

The estimator for a crop or stock item total in list stratum h is

$$Y_h = e_h \sum_{i \epsilon S_h} y_i = \sum_{i \epsilon S_h} y_i^e, \qquad (5)$$

where e_h is the expansion factor for all selected units in list stratum h,
S_h is the set of all selected units in h ($\sum_{i \epsilon S_h}$ denotes the summation of all selected units i in set S_h),
y_i is the item value for selected unit i (which may involve MACHINE IMPUTATION of sub-tract values), and
$y_i^e = e_h y_i$ is the item expanded value for selected unit i.

The expansion factor in this case is the inverse of the selection probability of each name in list stratum h; that is, $e_h = N_h/n_h$, where N_h is the total number of names in list stratum h (sampled or not), and n_h is the number of names sampled from h.

Reweighting

The estimator for a item total in list stratum h of a LABOR SURVEY (also a livestock item direct expansion) is

$$Y_h = e_h^* \sum_{i \epsilon U_h} y_i$$
$$= \sum_{i \epsilon U_h} y_i^r, \qquad (6)$$

where U_h is the set of selected units with usable item values in stratum h,
e_h^* is the reweighted expansion factor for all selected units in list stratum h,
$y_i^r = e_h^* y_i$ is the reweighted item expanded value for selected unit i,
and y_i is as before.

The reweighted expansion factor is equal to N_h/u_h, where N_h is as before, and u_h is the number of selected units with usable item values (if u_h is zero, then Y_h is undefined and remedial action must be taken).

The theory behind reweighting is discussed rigorously in [9]. Essentially, the set of randomly chosen selected units from h with usable item values is assumed to be equivalent to a simple random sample of size u_h.

The estimator in (6) is also applied to CROP and STOCK items for CV estimation purposes, but the resulting Y_h's are not displayed in SPS Summary tables.

EXPANDED COUNTS from agricultural surveys are also computed using (6). These are the estimators of the numbers of farms in particular domains (e.g., places having hogs). Expanded counts are item totals for count variables. Unexpanded counts, while included in SPS Summary tables, are not of concern here. See [3; Section 8] for more details about them.

Livestock Estimators

NASS uses two estimators for livestock items (including AQUACULTURE items, i.e., catfish and trout, but excluding all COUNTS): a direct expansion estimator which uses the formula in equation (6) and a presence estimator described below.

A distinction can be drawn between a nonusable selected unit recorded as having (within at least one of its sub-tracts) some quantity of a particular livestock type (such as hogs, cattle, sheep, goats, or chickens) and one for which the presence or absence of the livestock type is unknown for every one of its sub-tracts (note: there is often no more than one sub-tract per selected unit).

Consider a survey item (e.g., market hogs) related to a livestock type (hogs). IF A SELECTED UNIT HAS A USABLE ITEM VALUE (i.e, if all of its sub-tracts have usable item values) AND ANY SUB-TRACT IS RECORDED AS HAVING A POSITIVE QUANTITY FOR THE TOTAL NUMBER OF THE LIVESTOCK TYPE, THEN THE SELECTED UNIT IS A POSITIVE USABLE. Let u_h^+ be the number of positive usables in stratum h, and U_h^+ denote the set of those usables in h.

IF A SELECTED UNIT DOES NOT HAVE A USABLE ITEM VALUE BUT ONE OF ITS SUB-TRACTS RECORDED AS HAVING A POSITIVE QUANTITY FOR THE TOTAL NUMBER OF THE LIVESTOCK TYPE, IT IS A POSITIVE NONUSABLE. Let

$\bar{u}_h^+$ be the number of positive nonusables in h.

Note that it is possible for a selected unit to be a usable zero. This happens when all its subtracts have usable item values of zero for every item in the livestock type (or have LAF's of zero).

The PRESENCE estimator for an item LIVESTOCK total in list stratum h is (when u_h^+ is positive).

$$Y_h = N_h p_h x_{h(+)}, \qquad (7)$$

where $p_h = (u_h^+ + \bar{u}_h^+)/(u_h + \bar{u}_h^+)$,

$$x_{h(+)} = \sum_{i \epsilon U_h} y_i/u_h^+,$$

and N_h, U_h, and y_i are as before. (See [10] for a justification of this estimator.)

When u_h^+ is zero and p_h is zero, then Y_h is also zero. Otherwise, when u_h^+ is zero, Y_h is undefined (and some sort of remedial action must be taken).

Variances

The variance estimator for the direct expansion Y_h in either (5) or (6) is

$$V_h = [(N_h - u_h)/N_h][u_h/(u_h - 1)] \quad x \quad [\sum_{i \epsilon U_h} (y_i^r)^2 - (\sum_{i \epsilon U_h} y_i^r)^2/u_h] \qquad \text{for } u_h > 1 \tag{8}$$

$$= Y_h^2/2 \qquad \text{otherwise, when } Y_h \text{ exists. (It is undefined when } Y_h \text{ does not exist.)}$$

The first equality in (8) is unbiased for labor, count, and livestock items when the assumption underpinning reweighting is correct. Similarly, this equality is conservative (biased upward) for crop and stock items if NASS's machine imputation program is more statistically efficient than imputing missing item values with stratum means (which is mathematically equivalent to reweighting).

The second equality in (8) is based on the realization that $E[(Y_h)^2] = [E(Y_h)]^2 + Var(Y_h)$.

Since $Var(Y_h)$ is usually less than $[E(Y_h)]^2$,

$(Y_h)^2/2$ will most often be conservative as an

estimator of the variance of Y_h.

The variance estimator for Y_h in (7) (the PRESENCE estimator for LIVESTOCK) is given in [10]:

$$V_h = N_h^2\{[(x_{h(+)})^2 - s_h^{+2}/u_h^+] \times (1 - n_h'/N_h)p_h(1-p_h)/(n_h'-1) + p_h^2(1 - u_h^+/[p_hN_h])s_h^{+2}/u_h^+\}, \quad \text{when } u_h^+ > 0 \text{ and } n_h' > 1 \tag{9}$$

$$= Y_h^2/2 \quad \text{otherwise}$$

where

$$n_h' = u_h + \bar{u}_h^+, \text{ and}$$

$$s_h^{+2} = [\sum_{i\epsilon U_h^+} y_i^2 - (\sum_{i\epsilon U_h^+} y_i)^2/u_h^+]/(u_h^+ - 1) \quad \text{when } u_h^+ > 1$$

$$= (x_{h(+)})^2/2 \quad \text{when } u_h^+ = 1.$$

V. AREA EXPANSIONS

In this chapter, we focus on estimators for the total of a survey item in the area frame of a particular state. July cattle is somewhat special and is treated in a separate section. It should also be noted that, unlike its labor surveys, there are no area expansions for Hawaii's two agricultural surveys (December hogs and January cattle). Moreover, there are no area expansions for TROUT and CATFISH.

Direct Expansion - the Base Survey

As noted in Chapter II, a random sample of area segments is selected as part of the base survey. This sampling can be thought of here as a three step process. First, the area segments of a state are divided into land-use strata, such an intensively cultivated, extensively cultivated, range land, bodies of water, and non-agricultural. Different states can have different land-use strata definitions.

Second, the land-use strata are subdivided into substrata based on geography and other considerations (see [11] for more details). Finally, a (nearly) simple random sample of segments are selected from each substratum ([11] explains the need for the modifier "nearly").

Once a segment is sampled, the tracts within it are enumerated so that they can be expanded in as many as eight different ways (as discussed in chapter II, these are the closed, open, weighted, NOL-closed, NOL-weighted, NEO-closed, NEO-open, or NEO-weighted estimators). The three CLOSED direct expansion estimators for a land-use stratum total of a survey item have the following form:

$$Y_h = \sum_{j \epsilon B_h} \sum_{k \epsilon G_{hj}} e_{hjk} \sum_{m \epsilon T_{hjk}} w_{hjkm} t_{hjkm}, \qquad (10)$$

where h is the land-use stratum in question,
B_h is the set of all substrata in h,
G_{hj} is the set of all segments in substratum j of land-use stratum h,
T_{hjk} is the set of all tracts in segment k of substratum j of land-use stratum h,

e_{hjk} is the expansion factor for all tracts in segment T_{hjk},
t_{hjkm} is the survey value for tract m in segment T_{hjk}, which may (in principle) be machine imputed, and
w_{hjkm} is an indicator variable, i.e., it takes on either the value 0 or 1 depending on the estimator.

For the the closed estimator w_{hjkm} is always 1. For the NOL-closed estimator, w_{hjkm} is 1 if tract m is an NOL tract and zero if it is an OL tract. For the NEO-closed estimator, w_{hjkm} is 1 if tract m is an NEO tract and zero if it is an EO tract.

The two OPEN estimators (there is no NOL-open estimator in the 1989 base survey) have a form similar to (10):

$$Y_h = \sum_{j \epsilon B_h} \sum_{k \epsilon G_{hj}} e_{hjk} \sum_{m \epsilon T_{hjk}} w_{hjkm} b_{hjkm} y_{hjkm}, \qquad (11)$$

where b_{hjkm} is 1 if the farm headquarters for tract m of segment T_{hjk} is within the segment and zero otherwise,
y_{hjkm} is the entire farm value associated with tract m, which may be machine imputed, and

all the other notation is as before.

The three WEIGHTED estimators have the form:

$$Y_h = \sum_{j \epsilon B_h} \sum_{k \epsilon G_{hj}} e_{hjk} \sum_{m \epsilon T_{hjk}} w_{hjkm} a_{hjkm} y_{hjkm}, \qquad (12)$$

where a_{hjkm} is the weight used to prorate y_{hjkm} (presently, the tract acres of m divided by the farm acres of m; other weights are under study), and
all the other notation is as before.

No matter which of the eight estimators is employed to determine the land-use stratum totals, the state total is

$$Y_{ST} = \sum_{h \epsilon A_L} Y_h, \qquad (13)$$

where A_L is the set of all land-use strata in the

state.

Variances - the Base Survey

Let the expanded item value for a tract be

$$y_{hjkm}^{e} = e_{hjk} w_{hjkm} t_{hjkm} \quad \text{for closed estimators}$$

$$= e_{hjk} w_{hjkm} b_{hjkm} y_{hjkm} \quad \text{for open estimators}$$

$$= e_{hjk} w_{hjkm} a_{hjkm} y_{hjkm} \quad \text{for weighted estimators,}$$

and n_{hj} be the number of segments in G_{hj}. No matter how y_{hjkm}^{e} is defined, let the expanded item value for a segment be

$$y_{hjk\cdot}^{e} = \sum_{m \epsilon T_{hjk}} y_{hjkm}^{e}, \tag{14}$$

A slightly conservative variance estimator of Y_h in (10), (11), or (12) is given in [12] (assuming, incorrectly, that there is no additional variance from using machine imputed data for stocks):

$$V_h = \sum_{j \epsilon B_h} [(n_{hj}/[n_{hj} - 1]) \times \{ \sum_{k \epsilon G_{hj}} (y_{hjk\cdot}^{e})^2 - (\sum_{k \epsilon G_{hj}} y_{hjk\cdot}^{e})^2 / n_{hj} \}]. \tag{15}$$

The variance estimator for Y_{ST} in (13) is

$$V_{ST} = \sum_{h \epsilon A_L} V_h. \tag{16}$$

Direct Expansion - a Follow-on Survey

For an area expansion of a follow-on survey, a distinction must be drawn between selected units (tracts) and sub-tracts (reporting units), because a tract in the base survey may subdivide into different reporting units before the follow-on survey is conducted.

Tracts are selected for the area frame (or NOL) component of follow-on surveys in the following

manner. Those NOL tracts contacted for the base survey are restratified based on their responses to that survey and the equivalent of simple random samples are chosen from what are called select strata. For variance calculation purposes, however, several select strata are often collapsed into a single so-called summary stratum (see [13] for a more detailed description of this process).

Let y_{ms} denote an item value of interest for sub-tract s of selected unit m (which may be hand or **machine** imputed). For LABOR surveys this is the full (open) farm value of the survey item; for AGRICULTURAL surveys, it is the weighted farm value. (In both cases, the raw survey values are multiplied by LAF's; see equation (1).)

Let e_m be the expansion factor associated with selected unit m. This is the product of the first phase expansion factor from the base survey and the second phase expansion factor from the follow-on subsampling process (which itself has two phases in both March and August; but that need not concern us here; the second phase expansion factor for the HAWAII LABOR survey is unity).

The direct expansion for summary stratum r of an item from an AGRICULTURAL survey (except COUNTS) is

$$Y_r = \sum_{ms \epsilon R_r} e_m y_{ms}, \tag{17}$$

where R_r is the set of all REPORTING UNITS in the follow-on survey subsample that are within summary stratum r (land-use stratum in Hawaii).

The direct expansion for summary stratum r of an item from a LABOR survey (or a COUNT from an agricultural survey) is

$$Y_r^* = \sum_{ms \epsilon U_r} e_m^* y_{ms}, \tag{18}$$

where U_r is the set of all reporting units in R_r with USABLE item responses,
$e_m^* = e_m d$ is the reweighted expansion factor for item t and sub-tract s, and
d is the number of NOL reporting units in the state divided by the number of NOL

reporting units in the state with usable item responses.

What has been implicitly assumed is that each NOL reporting unit is equally likely to provide usable item values. If that assumption is true, Y_r^* is an unbiased estimator.

Equation (18) is also applied to crop and stock items for use in variance estimation. The set U_r (and thus the denominator of d) includes only those reporting units with non-machine imputed values.

No matter whether (17) or (18) is used, the area direct expansion for the state is

$$Y_{ST} = \sum_{r \epsilon A_S} Y_r^{(*)}, \qquad (19)$$

where A_S is the set of all summary strata in the state, and $Y_r^{(*)}$ is either equal to Y_r^* (for labor items and counts) or Y_r (otherwise) depending on which is appropriate.

Variances - a Follow-on Survey

The theory behind variance estimation for the NOL components of follow-on surveys is given in [13]. In every state save Hawaii, there are (at least) two phases of sampling for NOL tracts. As a result, variance estimators for the NOL component of a follow-on surveys have two components. Only the sum of the two is displayed in SPS Summary tables and that only for Y_{ST} in (19). This is because there is no meaningful way to estimate the portion of the first phase variance that is attributable to a particular summary stratum.

Second phase variance estimates for particular summary strata may exist, but estimated second phase standard errors and CV's are <u>not</u> displayed in SPS Summary tables.

In this section, we will be dealing with item values for selected units (tracts). The item value of a selected unit is the sum of its corresponding sub-tract values (the y_{ms}). Missing or machine imputed sub-tract values <u>are</u> <u>treated</u>

here as zeroes when determining selected unit item values. (This is because missing sub-tract values are considered to be the result of another phase of random sampling.)

Let m be a selected unit from summary stratum r, S_r be the set of all such m in r, and y_m be the item value m:

$$y_m = \Sigma\, y_{ms}, \tag{20}$$

where the summation is over all the sub-tracts within selected unit m (treating missing and machine sub-tract values as zero). Further, let f_m be the first phase expansion factor for the area segment in which selected unit m is located.

The estimator for the second phase variance of Y_r in (17) is

$$V_{2r} = [\{\sum_{m\epsilon S_r} (e_m^*/f_m)^2\} - N_r]/(v_r - 1) \quad x \quad \{\sum_{m\epsilon S_r} z_m^2 - (\sum_{m\epsilon S_r} z_m)^2/v_r\} \tag{21}$$

where N_r is the number of NOL tracts in summary stratum r <u>before</u> subsampling,
v_r is the number of selected units in S_r (after subsampling), and
$z_m = f_m y_m$.

Note that for livestock items e_m^* is by definition equal to e_m. Furthermore, V_{2r} is undefined when $v_r < 2$. Since there is no subsampling there, the HAWAII LABOR SURVEY HAS NO SECOND PHASE VARIANCE.

The expression V_{2r} can be biased, but the bias will be small when d is close to 1. Moreover, the bias is composed of two terms with opposite signs. This (we hope) will have a cancelling effect. See [14] for a fuller development.

Let T_{jk} be the set of subsampled tracts in segment k of substratum j from the base survey (note: land-use strata play no active role in follow-on survey estimators; also j denotes a "<u>variance stratum</u>" in Hawaii). Further, let

$$y_{jk\cdot} = \sum_{m \epsilon T_{jk}} e_m^* y_m \quad \text{when } T_{jk} \text{ has at least one member}$$
$$= 0 \quad \text{otherwise.} \tag{22}$$

Similarly, let

$$y_{j\cdot\cdot} = \sum_{k \epsilon G_j} y_{jk\cdot} \, , \tag{23}$$

where G_j is the set of all n_j sampled segments in substratum j.

Suppose there are L substrata in the state (note: substrata from different land-use strata are, by our definition, different substrata). We can define the following two variables:

$$V^* = \sum_{j=1}^{L} (n_j/[n_j - 1])\left(\sum_{k \epsilon G_j} y_{jk\cdot}^2 - y_{j\cdot\cdot}^2/n_j \right) \tag{24}$$

and

$$D^* = \sum_{r \epsilon A_s} \left(\left(\left[\sum_{m \epsilon R_r} (e_m/f_m)^2 \right] - N_r \right)/[v_r(v_r - 1)] \right) \times \tag{25}$$
$$\left[\sum_{j=1}^{L} (n_j/[n_j - 1]) \left(\left[\sum_{k \epsilon G_j} z_{jkr}^2 \right] - z_{j\cdot r}^2/n_j \right) - z_{\cdot\cdot r}^2 \right]$$

where z_{jkr} is equal to Σz_m, the summation being over the intersection of the sets T_{jk} and S_r (z_{jkr} is zero when the intersection has no members),

$z_{j\cdot r} = \sum_{k \epsilon G_j} z_{jkr}$, the summation of all the z_{jkr} in substratum G_j,

and $z_{\cdot\cdot r} = \sum_{j=1}^{L} z_{j\cdot r}$, the summation of the $z_{j\cdot r}$ over all L substrata for a given summary stratum r.

The variance estimator of Y_{ST} in (19) suggested by [13] is

$$V_{ST} = V^* + D^* \text{ when } V^* + D^* \geq \sum_{r \epsilon A_S} V_{2r}$$

$$= V^* \text{ when } V^* \geq \sum_{r \epsilon A_S} V_{2r} > V^* + D^* \quad (26)$$

$$= \sum_{r \epsilon A_S} V_{2r} \text{ when } \sum_{r \epsilon A_S} V_{2r} > V^*.$$

In the 1989 version of the SPS Summary, the more conservative V^* from equation (24) is used in variance estimation. For the HAWAII LABOR survey, which has no subsampling, V^* would be an unbiased variance estimator if it were not necessary to collapse substrata into variance strata due to sample size considerations.

The area expansions for both the MARCH and AUGUST Agricultural Surveys (the latter is exclusively a survey of rice stocks in six states) are based on subsamples of NOL SELECTED units from the December Agricultural Survey. (Note: the list and area frames for the August 1990 rice stocks survey were determined by the June 1989 base survey.) That is to say, there is a three phase NOL sampling design in those survey months. Nevertheless, V^* in equation (24) remains a conservative estimator of the NOL variance.

Formally, there is a stage of sampling in all follow-on surveys between the first and second phases as described above. Approximately 60% of the sampled June area segments are allocated for agricultural survey subsamples and the remaining segments for labor subsamples. This allocation process can be thought of as part of the first phase of sampling.

July Cattle and December Chicken

As mentioned previously, the July (or mid-year) NEO-closed estimator for a cattle item is calculated from the base (June) survey. There is also a NOL-weighted estimator based strictly on a July area subsample. This latter estimator is no different from the area expansion for any follow-on survey; it uses equations (17) through (24).

The NEO-closed estimator for cattle is a more complicated matter. Each cattle survey item in

the June area sample has two components. One component covers cattle on Public Industrial Grazing Association (PIGA) land and the other cattle on non-PIGA land. For the NEO-closed estimator, each t_{hjkm} in equation (10) is the sum of the on-tract non-PIGA cattle value of interest and a_{hjkm} (from equation (12)) times the <u>total farm</u> PIGA cattle value of interest. Variance formulae for July cattle NEO-closed area expansions follow equations (15) and (16).

The CHICKEN component of the December Agricultural Survey is even more complicated than the PIGA-effected parts of the NEO-closed estimator for cattle. As with the cattle estimator, the data file inputed into the SPS Summary is such that the appropriate equations for livestock surveys can be applied in a straightforward manner.

VI. MULTIPLE FRAME EXPANSIONS AND AGGREGATION

Direct Expansion

Equations (5), (6) and (7) provide estimators for an item total in list stratum h, as described in Chapter IV. (In this chapter, the LIVESTOCK PRESENCE estimator is included in discussions of direct expansion estimators.) If there are H list strata in a state (h=1, 2, ..., H), then the list direct expansion for the state is:

$$Y_L = \sum_{h=1}^{H} Y_h, \qquad (27)$$

regardless of whether (5), (6), or (7) is used in determining the Y_h. (Note: if necessary, each undefined Y_h is changed to zero).

For the base survey, the state list direct expansion in (27) can be added to the state area direct expansion in (13) to form either the state OL-MF/closed or OL-MF/weighted estimator, presuming that the Y_h in (13) are computed in the appropriate manner (equation (10) is used for the OL-MF/closed and (12) for the OL-MF/weighted estimator with w_{hjkm} =1 when m is an NOL tract and 0 otherwise).

For a follow-on survey, Y_L in (27) is simply added to Y_{ST} in (19) to form the state direct expansion estimator.

Returning to the list side, let h=1, ..., E be EO strata, and let

$$Y_E = \sum_{h=1}^{E} Y_h, \qquad (28)$$

For the base survey, Y_E in (28) can be added to the state area direct expansion in (13) to form either the state EO-MF/closed, EO-MF open, or EO-MF/weighted estimator, presuming that the Y_h in (13) are computed in the appropriate manner (equation (10) is used for the EO-MF/closed, (11) for the EO-MF/open, and (12) for the EO-MF/weighted estimator with w_{hjkm} = 1 when m is an NEO tract and 0 otherwise).

Regional or U.S. level direct expansions are determined by simply adding together all the component state direct expansions.

Variances

The variance estimator for either Y_L or Y_E is simply the sum of the variances of its component Y_h; i.e,

$$V_L = \sum_{h=1}^{L} V_h, \tag{29}$$

while

$$V_E = \sum_{h=1}^{E} V_h, \tag{30}$$

where V_h is defined by either (8) or (9), as described in Chapter IV. (Note: if necessary, each undefined V_h is changed to zero.)

Likewise, the variance estimator for a state direct expansion (list + area) is also the sum of its component variance estimators. Finally, the variance estimator of a regional or U.S. level direct expansions is also equal to the sum of its component variance estimators.

VII. RATIO ESTIMATORS

Introduction

The SPS Summary computes two types of ratios. One type is the ratio of different survey items on the same survey. For example, the ratio of wages paid to labor hours on a labor survey or the ratio of corn production to harvested acres on an agricultural survey. We call this an item-to-item ratio.

The second type of ratio is of the same (or similar) survey items on different surveys. When both surveys are base surveys, we call this a base-to-base ratio; otherwise, it is called a survey-to-survey ratio.

The principle of ratio estimation is that all respondents providing data (whether direct or imputed) for the numerator of the ratio also provide data for the denominator, and vice versa. Ratio estimators take the following form in the SPS Summary:

$$R = Y'/X', \tag{31}$$

where Y' and X' are estimators of the appropriate totals based, in principle, only on data from matched respondents.

The variance estimator for R generally has the form (there is one exception which will be addressed later):

$$V_R = R^2\left(\frac{V_{yy}}{(Y')^2} - 2\frac{V_{yx}}{Y'X'} + \frac{V_{xx}}{(X')^2}\right), \tag{32}$$

where V_{yy}, V_{xx}, and V_{yx} are, respectively, the variance estimator for Y', the variance estimator for X', and the covariance estimator for Y' and X'.

Now Y' and X' are always based on the same sample, so V_{yy}, V_{xx}, and V_{yx} have the same form. As a result, only the estimator V_{yx} will be presented in the text. It is a simple matter to derive V_{yy} and V_{xx} from V_{yx} (for example, one gets V_{yy} by changing all x-values in V_{xy} to y-values).

The SPS Summary also calculates ratio expansions from base-to-base and survey-to-survey ratio estimators (and sometimes, but not often, from item-to-item ratio estimators in the base survey; see the section on "digitized land expansions"). A ratio expansion has the form:

$$Y_{RE} = XR, \tag{33}$$

where X is an estimator for the same total as the denominator of R, but is (usually) based on more information than X'. We call X the root or root estimator of the ratio expansion.

The variance of Y_{RE} is estimated in the SPS Summary by

$$V_{RE} = (R^2 - V_R)\text{var}(X) + X^2 V_R, \tag{34}$$

where var(X) is the variance estimator for X and V_R is defined in (32) (the derivation of V_{RE} comes from (A2) and (A3) of [14]). Except in degenerate cases, equation (34) will provide a nearly unbiased estimator for the variance (more correctly, the mean squared error) of Y_{RE} when X and R are independent (or at least uncorrelated). Unfortunately, this may not be the situation in reality.

Item-to-Item - the Base Survey

The SPS Summary provides item-to-item ratios (such as harvested to planted wheat acres) for every list and land-use stratum in a state. FOR A LAND-USE STRATUM, THESE ARE SIMPLY THE RATIOS OF THE APPROPRIATE DIRECT EXPANSIONS (which may involve machine imputed data) described in Chapter V.

The situation is a bit more complicated for a list stratum, say h, where an analogue of (6) is used (when the set M_h is not empty):

$$Y_h' = e_h' \sum_{i \epsilon M_h} y_i$$

$$= \sum_{i \epsilon M_h} y_i^r, \tag{35}$$

and

$$X_h' = e_h' \sum_{i \epsilon M_h} x_i$$

$$= \sum_{i \epsilon M_h} x_i^r,$$

where M_h is the set of SELECTED UNITS with both usable y_i and x_i item values in stratum h, and

$e_h' = N_h/m_h$, where N_h is the total number of names on the list frame in stratum h and m_h is the number of selected units in M_h.

When M_h is empty, Y_h', X_h', and $R_h = Y_h'/X_h'$ are undefined. (From now on, we will not provide a definition like X_h' in (35) when there is an exactly parallel expression in y-values, in this case Y_h').

The covariance estimator for Y_h and X_h for a LAND-USE STRATUM is analogous to equation (15) (remember that $Y_h = Y_h'$ in this case):

$$V_{h \cdot xy} = \sum_{j \epsilon B_h} (n_{hj}/[n_{hj} - 1]) \times$$

$$\left(\sum_{k \epsilon G_{hj}} (y_{hjk \cdot}^{e})(x_{hjk \cdot}^{e}) - \left(\sum_{k \epsilon G_{hj}} y_{hjk \cdot}^{e} \right)\left(\sum_{k \epsilon G_{hj}} x_{hjk \cdot}^{e} \right)/n_{hj} \right), \quad (36)$$

where the notation is the same as in (15).

The covariance estimator of Y_h' and X_h' for a LIST STRATUM is analogous to (8) for count variables:

$$V_{h \cdot xy} = [(N_h - m_h)/N_h][m_h/(m_h - 1)] \times \left[\sum_{i \epsilon M_h} y_i^r x_i^r - \left(\sum_{i \epsilon M_h} y_i^r \right)\left(\sum_{i \epsilon M_h} x_i^r \right)/m_h \right] \quad \text{for } m_h > 1$$

$$= Y_h' X_h'/2 \quad \text{for } m_h = 1 \quad (37)$$

is undefined otherwise,

where y_i^r and x_i^r are defined in (35).

WHEN m_h=1, EQUATION (32) IS OVERRIDDEN, AND V_R (the variance estimator for R_h) IS UNDEFINED. The

second line of (37) is used for ratios that involve aggregation across strata.

The covariance of a sum (across strata, states, or regions) is the sum of the component covariances. When aggregating across strata, all undefined Y_h', X_h', $V_{h\cdot yy}$, $V_{h\cdot xx}$, and $V_{h\cdot xy}$ are changed to zero.

As an example of aggregation, the ratio estimator for the list frame of a state is:

$$R_{ST} = \frac{\sum_{h=1}^{H} Y_h'}{\sum_{h=1}^{H} X_h'} . \qquad (38)$$

The variance estimator for R_{ST} is

$$V_{R\cdot ST} = R_{ST}^2 \left\{ \frac{\sum V_{h\cdot yy}}{(\sum Y_h')^2} - 2 \frac{\sum V_{h\cdot yx}}{(\sum Y_h')(\sum X_h')} + \frac{\sum V_{h\cdot xx}}{(\sum X_h')^2} \right\}, \qquad (39)$$

where all the summations are from 1 to H.

Equations (38) and (39) are simply the applications of (31) and (32) at the state list frame level. Applications to other levels of aggregation are straightforward.

Item-to-Item - a Follow-on Survey

Item-to-item ratios for list strata are handled the same way on follow-on surveys as on the base survey. Item-to-item ratios for NOL's are calculated at the same levels of aggregation as NOL direct expansions. For any item, we use an analogue of (18):

$$Y' = \sum_{ms \in M_R} e_m d' y_{ms}, \qquad (40)$$

where M_R is the set of all REPORTING UNITS from the NOL subsample for the state that

has both usable y_{ms} and x_{ms} values,

e_m is the expansion factor for SELECTED UNIT m,

and d' is the number of NOL reporting units in the state divided by the number of reporting units in M_R.

In order to estimate the covariance of Y' in (40) and its companion X', it is necessary to first construct an analogue of $y_{jk\cdot}$ in equation (22). This is simple enough to do. Let

$$y_{jk\cdot}' = \Sigma\, e_m d' y_{ms}, \qquad (41)$$

where the summation is over all reporting units in set M_R and segment k of substratum (or variance stratum) j. If there are no such reporting units, set $y_{jk\cdot}' = x_{jk\cdot}' = 0$. Similarly, let

$$y_{j\cdot\cdot}' = \Sigma\, y_{jk\cdot}', \qquad (42)$$

where the summation is over all segments k in substratum j.

The covariance estimator of Y' and X' in (40) is then

$$V_{xy} = \sum_{j=1}^{L} (n_j/[n_j - 1]) \times \qquad (43)$$

$$\left(\sum_{k \epsilon G_j} y_{jk\cdot}' x_{jk\cdot}' - y_{j\cdot\cdot}' x_{j\cdot\cdot}'/n_j \right),$$

where n_j is the number of segments sampled from substratum j in the base survey, L the number of substrata in the state, and G_j the the set of all sampled segments in substratum j (as in (22)).

(We have dropped the analogue to D^* in equation (25); it is likely small and, if anything, has a tendency to be negative.)

Base-to-Base

Base-to-Base ratio estimators involve only area base survey samples. They compare June 1989 survey values with June 1988 survey values for the area segments found in both base surveys. For crop acreages, the closed approach is used (the

weighted approach is also used for cropland acreages in 10 states); for land in farms, both the closed and open approaches; for the number of farms, the open and weighted approaches; for hogs, the NEO-closed, NEO-open (for 10 states), and NEO-weighted (for 10 states) approaches; and for cattle, the NEO-closed approach.

Borrowing notation from Chapter V, let $y_{hjk\cdot}^{e}$ be the appropriate 1989 expanded value for segment k in substratum j of land-use stratum h (see equation (14), and let $x_{hjk\cdot}^{e}$ be the matching segment expanded value in 1988 (if one exists). The ratio estimator for land use stratum h is:

$$R_h = \frac{Y_h'}{X_h'} = \frac{\sum_{j\epsilon B_h} (n_{hj}/n_{hj}') \sum_{k\epsilon G_{hj}'} y_{hjk\cdot}^{e}}{\sum_{j\epsilon B_h} (n_{hj}/n_{hj}') \sum_{k\epsilon G_{hj}'} x_{hjk\cdot}^{e}} = \frac{\sum_{j\epsilon B_h} \sum_{k\epsilon G_{hj}'} y_{hjk\cdot}'}{\sum_{j\epsilon B_h} \sum_{k\epsilon G_{hj}'} x_{hjk\cdot}'} \qquad (44)$$

where G_{hj}' is the set of all segments in substratum j of land-use stratum h in both the 1988 and 1989 area samples (note: such segments have a "comparability code" equal to 1 on the edit master file),

n_{hj}' is the number of segments in G_{hj}',

$y_{hjk\cdot}' = (n_{hj}/n_{hj}')y_{hjk\cdot}^{e}$,

and all the other notation follows that of Chapter V (n_{hj} is as defined for the 1989 base survey). For those state with new frames in 1989 (Iowa, the New England states, and West Virginia), Y_h', X_h', and R_h are undefined.

The covariance estimator for Y_h' and X_h' is

$$V_{h\cdot xy} = \sum_{j\epsilon B_h} (n_{hj}'/[n_{hj}' - 1]) \quad x$$

$$\{ \sum_{k\epsilon G_{hj}'} (y_{hjk\cdot}')(x_{hjk\cdot}') \quad - \qquad (45)$$

$$(\sum_{k\epsilon G_{hj}'} y_{hjk\cdot}')(\sum_{k\epsilon G_{hj}'} x_{hjk\cdot}')/n_{hj}'),$$

The real purpose behind base-to-base ratios is ratio expansion (see equation (33)). Ratio expansions can be determined for each land-use stratum h by multiplying the 1988 land-use stratum direct expansion, call it X_h, by R_h in (44).

To determine the variance estimator of $Y_{h\cdot RE} = X_h R_h$, the SPS Summary employs equation (34), where the variance estimator for X_h is exactly parallel to equation (15) but based on the 1988 sampled segments and their item values. (The SPS Summary actually recalculates X_h and var(X_h) internally.)

The land-use stratum ratio expansions are <u>not</u> summed to get the area frame ratio expansion for a state. Rather,

$$Y_{ST\cdot RE} = (\Sigma\ X_h) \frac{\Sigma\ Y_h'}{\Sigma\ X_h'}, \qquad (46)$$

where all the summations are over the land-use strata in the state, is used. This ratio expansion is in required $Y_{RE} = XR$ form, and so equation (34) is invoked to determine the variance estimator for $Y_{ST\cdot RE}$.

The right hand side of equation (46) is also used to calculate regional and U.S. level <u>area frame</u> ratio expansions. The summations are redefined appropriately.

Livestock ratio expansions based on (46) only cover NEO tracts. To estimate the livestock totals for a state, $Y_{ST\cdot RE}$ in (46) is added to Y_E

from equation (28) (the July Y_E for cattle items). The variance of this sum is the sum of the two variances. The same thing is done for regional and U.S. level livestock ratio expansions, an EO total is added to the appropriate version of the right hand side of (46). Variances are estimated

analogously to state variances (NEO variance plus EO variance).

Survey-to-Survey

We are concerned in this section with the ratio of the same item (or similar items) on two agricultural surveys less than one full year apart. These ratios are always in the form of current to previous survey values.

Survey-to-survey ratio estimators for a LIST STRATUM follow exactly the formula in equation (35) (note: the y_i are determined using current LAF's, while the x_i use the previous LAF's). Likewise, equation (37) provides the needed information for variance estimation.

There is one complication that needs to be mentioned. Sometimes a selected unit contains either more or less reporting units (sub-tracts) in the current survey than in the previous survey. Since equation (35) is a summation of selected unit item values, this should not be a problem: the current (previous) item value for a selected unit is the sum of all its current (previous) sub-tract values.

Unfortunately, when there are more sub-tracts for a particular selected unit in June 1989 than in December 1989, it is possible that not all of the June sub-tracts appear in the December edited data file (changes in the makeup of edited data files prevent such phenonema from ocurring in March 1990 and afterwards). This can cause December-to-June ratio estimators to be biased slightly upward (remember, most selected units contain only a single sub-tract in all survey months).

The SPS Summary also calculates ratio expansions from survey-to-survey ratios. It uses equation (5) as the root estimator for CROP and STOCK items and (7) as the root for livestock items. For the variance of the root, it uses (8) and (9), respectively. (As with base-to-base ratios, the roots and their variance estimates are recalculated internally.) Equation (34) is again used as the variance estimator for the stratum ratio expansion.

NOL survey-to-survey estimators are much more difficult since they depend on what two agricultural surveys are being compared. The potential pairs of surveys fall into two groups which are treated separately:

Group 1: September to June, December to June, March to December, August to March, August to December.

Group 2: December to September, March to September, January to July.

In the first group, the selected units of the current survey are a subsample of the selected units of the previous survey. In the second group, the selected units of both surveys are independently drawn from the sampled tracts of the base survey.

For an NOL survey-to-survey ratio from Group 1 or 2, let y_m be the total item value for a selected tract in the current survey (see equation (20)), and x_m be the total item value for the tract in the previous survey.

Item values on the tract or sub-tract level are almost always weighted values (in the sense of the weighted estimator). The one exception to this is that CROP VALUES FOR THE BASE SURVEY ARE CLOSED.

The following is the estimator of Y' for a state NOL survey-to-survey ratio from Group 1:

$$Y' = \sum_{m \epsilon M_S} e_m d'' y_m, \qquad (47)$$

where M_S is the set of all tracts from the current NOL subsample for the state that has <u>both</u> usable y_m and x_m values (a y_m value is said here to be usable if all if its component y_{ms} values are usable; they can be neither missing nor machine imputed),

e_m is the expansion factor for tract m in the CURRENT survey,

and d'' is the number of current NOL subsampled tracts in the state divided by the number of tracts in M_S.

(Note: as on the list side, the y_m are determined using current LAF's, while the x_m use previous LAF's.)

Paralleling (41), let

$$y_{jk\cdot}' = \Sigma\, e_m d'' y_m, \qquad (48)$$

where the summation is over all sub-tracts in both M_S and segment k of substratum (or variance stratum) j. If there are no such reporting units, set $y_{jk\cdot}' = x_{jk\cdot}' = 0$. Similarly, let

$$y_{j\cdot\cdot}' = \Sigma\, y_{jk\cdot}', \qquad (49)$$

where the summation is over all segments k in substratum j.

The covariance of Y' and X' is then expressed by (43). Ratio expansions for the state NOL component use (13) or (19) as their root, depending on whether the previous survey was conducted in June or December. Variance estimators again are based on (34).

For area frame DECEMBER-TO-SEPTEMBER ratios, equation (47) is again used but the numerator of d'' is now the number of tracts in both the December and September samples, while $e_m = e_m^D e_m^S / f_m$ is the product of the base period expansion factor (f_m in (21)), the December subsampling expansion factor (e_m^D/f_m, where e_m^D is the DECEMBER expansion factor), and the September subsampling expansion factor (e_m^S/f_m, where e_m^S is the SEPTEMBER factor). Equations (48) and (49) follow. Area frame MARCH-TO-SEPTEMBER ratios are analogous (e.g., $e_m = e_m^M e_m^S / f_m$, where e_m^M is the March expansion factor) as are area-frame JANUARY-TO-JULY ratios.

To determine the survey-to-survey ratio for a state, the SPS Summary follows the principle of equation (38), adding the Y_h' values from the list strata ((35)) to Y' from the NOL component ((47)) in the numerator of the ratio and doing the same with the X_h' and X' in the denominator. Ratio expansions and variance estimation follow as always.

Digitized Land Expansions

The SPS Summary produces another type of item-to-item ratio expansion for the base survey. The root is total land area (for a state, region, or the whole U.S) provided by Statistical Abstracts put out by the Bureau of Census. The numerator values are CROP acreages by segment calculated using equations (14) and (10) (the closed estimator). The denominator values are digitized acres for the segments. These are derived by the Area Frame Section using digitizing tablets to electronically (and very accurately) measure the area in each segment.

The ratio estimator for a land-use stratum item-to-item ratio in this case is equation (44) with $n_{hj} = n_{hj}'$. As a result, $y_{hjk\cdot}'$ in (44) is replaced by $y_{hjk\cdot}^{e}$ in (14) (while the replacement for $x_{hjk\cdot}'$ can be found in the input data set.) Variance estimation (based on (45) with $y_{hjk\cdot}'$ set equal to $y_{hjk\cdot}^{e}$) and aggregation to higher levels are staightforward.

The variance of the root of one of the item-to-item ratio expansion under discussion is zero, because the census land area data is considered truth. As a result, the variance estimator of a ratio expansion has the form

$$V_{RE} = X^2 V_{R\cdot} \qquad (50)$$

Equation (50) can be applied at any level of aggregation in a straightforward manner.

VIII. AGRICULTURAL SERVICES

As noted in Chapter II, the agricultural services survey is conducted in two states - California and Florida. A survey item from this survey is often paired with an item from the basic (or farm) labor survey (in fact, they share the same "item code" on the questionaire; for example, total workers is item code 360 on both surveys). The SPS Summary produces three types of expansions (or ratios) from such paired survey items: farm labor, ag services, and TOTAL LABOR (farm labor plus ag services). This section treats the latter two types of expansions for the California and Florida labor surveys.

A sample is drawn from a supplemental list frame of ag services (contractors) in California and Florida. In addition to this, farms enumerated for the basic labor survey report the ag services firms they use. When these firms are not on the list frame of ag services, NASS adds them to the ag services survey. Their quantitative responses are for all the labor services they provide. Consequently, these responses must be divided by the number of farms provided services by the firm. This is handled (before the Summary) in the LAF.

The direct expansion estimator for a STRATUM OF AG SERVICES uses the same formula as equation (6). Its variance estimator is the same as equation (8). Likewise, the direct expansion of the numerator and denominator of an item-to-item ratio is expressed by equation (35) with the variance of the ratio following the same pattern as that of any list stratum.

Let us now turn our attention to labor from ag services firms not on the ag services list frame. Suppose one of the sub-tracts of a selected unit from either the basic labor list frame or the NOL area subsample uses an ag services firm not on the ag services list frame. That firm is added as a sub-tract of the selected unit. ITS RAW SURVEY ITEM VALUES ARE MULTIPLIED BY TWO LAF's -- one for the ag services firm itself and one for the basic list or NOL area reporting unit that originated it (deemed to be the first sub-tract of the tract).

Since total labor survey items include both basic labor and ag services items, it would be difficult

to handle nonresponse simply by adjusting the expansion factors. Instead, we employ these variations to our usual procedures:

1) The value for an ag services survey item from an added ag services sub-tract is modified to be

$$y_{is} = (g/g')y_{is}' \quad \text{when } g' > 0$$
$$\text{is undefined} \quad \text{otherwise (and remedial action should be taken),} \tag{51}$$

where y_{is}' is the original survey item value (including the two LAF's) when such a valid value exists and ZERO OTHERWISE,

g is the number of added ag services in the state: contractors not on the ag services list frame that were added to the sample, and

g' is the number of those added ag services with valid item responses (for an item-to-item ratio, it is, in principle, the number of added ag services with valid item values for <u>both</u> the numerator and the denominator).

(Note: The subscript "i" in (51) is replaced by "m" for NOL area sub-tracts.)

2) All ag services survey item values for BASIC LIST sub-tracts are (by definition) zero (they are not defined for basic NOL sub-tracts).

3) The survey value (ag services or total labor) for a selected unit (tract) from a BASIC LIST stratum is calculated using equation (2). The resultant sum is deemed <u>nonusable</u> when and only when one of the basic labor sub-tract fails to provide a usable response for the paired farm (basic) labor item. (IMPORTANT NOTE: <u>Even when part of a nonusable tract, an ag services sub-tract providing a valid response counts in computing g'. Conversely, even when part of a usable tract, an ag services sub-tract not providing a valid response does not count in computing g'</u>.)

4) The NOL AREA direct expansion for an AG SERVICES survey item uses equation (18) with d

defined as the number of basic NOL reporting units in the state divided by the number of those units reporting the ag services they use (which can be zero). For an item-to-item ratio, d' in (40) equals d as described above. (This is because in 1989 surveys when any item response in a section is not valid all item responses are deemed not valid.)

5) The NOL AREA direct expansion for an TOTAL LABOR survey item uses equation (18) with d defined separately for ag services and basic labor sub-tracts (the former as in point (4) above, the latter as in the basic labor survey). The item-to-item ratio in (40) is handled analogously.

The variance estimator for a BASIC LIST stratum direct expansion in (6) is equation (8) (this estimator is biased when g/g' is greater than one; see the discussion of V_{2r} in equation (21)). The variance estimator for an item-to-item ratio follows equation (32) (as before) with equation (37) serving as the formula for the stratum covariance estimator.

The (conservative) variance estimator for the NOL area direct expansion is equation (24). For an item-to-item ratio, the key covariance estimator is expressed by equation (43). NASS has no plans to estimate second phase variances for ag services and total labor items.

When aggregating, we once more change any undefined expansion (and its variance estimator) to zero. Expansions from all three frames must be added together for state totals, separately in the numerator and the denominator of item-to-item ratios.

IX. YEARLY AGGREGATES

The SPS Summary has been equipped to provide yearly estimates of some labor survey items after the July 1989 survey and again after the October 1989 survey.

Consider a particular labor survey item. Let t=1 denote October 1988, t=2 denote January 1989, t=3 denote April 1989, t=4 denote July 1989, and t=5 denote October 1989. Further, let Y_t denote the item direct expansion (at some level of aggregation) for survey t calculated using the formulae in Chapters III through VIII (Y_t may also be the numerator or denominator of an item-to-item ratio).

Each Y_t is measured on a per-week basis. The yearly aggregate of interest is also measured on a per-year basis. Thus for an item requiring a yearly direct expansion after the JULY survey, the formula used is

$$Y = 13\{\sum_{t=1}^{4} Y_t\}. \qquad (52)$$

(The SPS Summary actually resummarizes Y_1, Y_2, and Y_3 values to determine Y). For an item requiring a yearly direct expansion after the OCTOBER survey, the formula used is

$$Y = 13\{\sum_{t=2}^{5} Y_t\}. \qquad (53)$$

(Again, the SPS Summary actually resummarizes Y_2, Y_3, and Y_4 values to determine Y).

The variance estimators for the Y in (52) and (53) are not so straightforward. Their derivation takes up the remainder of the section. To this end, the Y are be broken up into list and area components. For ag services and total labor items, the list component includes both the basic and ag services list expansions.

The NOL Component of Variance for July

Let Y_{At} be that part of Y_t in (52) that comes from the NOL subsample. The variance estimator for $Y_A = 13(Y_{A1} + Y_{A2} + Y_{A3} + Y_{A4})$ has the form:

$$V_A = 169(V_{A(1,2,3)} + V_{A4}), \qquad (54)$$

where $V_{A(1,2,3)}$ is the variance estimator for $Y_{A(1,2,3)} = Y_{A1} + Y_{A2} + Y_{A3}$. This is because the July 1989 NOL subsample comes from the 1989 area sample. As a result, it is assumed independent of the three subsamples that come from the 1988 area sample. (This assumption neglects the fact that, in some states, only 20% of sampled area segments were rotated out of the sample in June 1989. As a result, the variance estimation described below can be biased in those states.)

Paralleling equation (48), let

$$y_{jk\cdot}' = \sum_{t=1}^{3} \sum_{ms \epsilon M_{jkt}} e_m d_t Y_{mst}, \qquad (55)$$

where M_{jkt} is the set of all NOL sub-tracts in segment k of substratum (or variance stratum) j with usable item responses in the survey for month t,

e_m is the expansion factor for NOL tract m,

d_t is the number of NOL sub-tracts in the state divided by those providing usable item responses in the survey for month t (with TOTAL LABOR items in California and Florida, d_t is defined separately for ag services and basic labor sub-tracts),

Y_{mst} is the survey item value for sub-tract ms and survey month t.

(Note: Variance strata are used in place of substrata in Hawaii and in those states whose 1988 labor subsamples were not selected from the June 1988 area sample.)

Equations (49) and (43) follow immediately. The latter (in V_{yy} form) providing a variance estimator for $Y_{A(1,2,3)}$ which can serve as

$V_{A(1,2,3)}$ in (54).

The NOL Component of Variance for October

Let Y_{At} be that part of Y_t in (53) that comes from the NOL subsample. The variance estimator for $Y_A = 13(Y_{A2} + Y_{A3} + Y_{A4} + Y_{A5})$ has the form:

$$V_A = 169\{V_{A(2,3)} + V_{A(4,5)}\}, \qquad (56)$$

where $V_{A(2,3)}$ is the variance estimator for $Y_{A(2,3)} = Y_{A2} + Y_{A3}$, and $V_{A(4,5)}$ is the variance estimator for $Y_{A(4,5)} = Y_{A4} + Y_{A5}$.

To calculate $V_{A(2,3)}$ and $V_{A(4,5)}$, the principles used in calculating $V_{A(1,2,3)}$ are again employed. Now, however, the first summation in equation (55) runs from 2 to 3 for $V_{A(2,3)}$ and from 4 to 5 for $V_{A(4,5)}$.

Again, equations (49) and (43) follow immediately. The latter (in V_{yy} form) providing a variance estimators $V_{A(2,3)}$ and $V_{A(4,5)}$ which are plugged into equation (56).

The List Component of Variance for July

Let Y_{ht} be that part of Y_t in (52) (or (53)) that comes from stratum h of the list sample (h can be a basic labor list stratum or an ag services list stratum). The variance estimator for $Y_h = 13(Y_{h1} + Y_{h2} + Y_{h3} + Y_{h4})$ has the form:

$$V_h = 169\{V_{h(1,2,3)} + V_{h4}\}, \qquad (57)$$

where $V_{h(1,2,3)}$ is a variance estimator for $Y_{h(1,2,3)} = Y_{h1} + Y_{h2} + Y_{h3}$. This is because the July 1989 list sample is drawn independently of the samples used for the other three surveys.

Unlike the NOL subsample, the list samples are not the same in October 1988, January 1989, and April 1989. They are, however, highly correlated (and all have the same size). Half of the selected units from list stratum h in October and January are the same. The other half of the January sample corresponds to half of the April sample. That part of the October sample that does not correspond to the January sample is called Replicate 2. Replicate 3 is the overlapping portions of the samples in October and January, while Replicate 4 is the overlap in January and April. Replicate 1 is that part of the April sample that is not surveyed in January. (Replicates 1 and 2 were also enumerated for the July 1988 survey.)

To estimate the variance of $Y_{h(1,2,3)}$, a selected unit in Replicate 2 of the October survey is matched to a selected unit in Replicate 4 (note: all replicate mathching is done before the SPS Summary). Likewise, a selected unit in Replicate 1 of the April survey is matched to one in Replicate 3. The result is a "matched" sample of selected units for the three surveys with the same sample size as each of the surveys.

Paralleling equation (6), let

$$Y_{h(1,2,3)}' = \sum_{i \epsilon U_h} \sum_{t=1}^{3} e_h^* y_{it}, \qquad (58)$$

$$= \sum_{i \epsilon U_h} y_i^r \qquad \text{if } u_h > 0$$

$$\text{is undefined} \qquad \text{otherwise,}$$

where U_h is the set of all matched selected units in list stratum h with usable item values in all three surveys,
$e_h^* = N_h/u_h$,
N_h is the total number of names in list stratum h,
u_h is the number of selected units in U_h,
y_{it} is the month t survey item values for selected unit i,
and $y_i^r = \Sigma\, e_h^* y_{it}$, where the summation is over t = 1, 2, and 3.

A conservative estimator for the variance of $Y_{h(1,2,3)}$ is

$$V_{h(1,2,3)} = [u_h/(u_h - 1)] \times \left[\sum_{i \epsilon U_h} (y_i^r)^2 - \left(\sum_{i \epsilon U_h} y_i^r \right)^2 / u_h \right] \quad \text{when } u_h > 1 \qquad (59)$$

$$= (Y_{h(1,2,3)})^2/2 \quad \text{when } u_h < 2 \text{ and } Y_{h(1,2,3)} \text{ exists}$$

is undefined otherwise.

The first line of this estimator would be unbiased if $Y_{h(1,2,3)}$ were equal to $Y_{h(1,2,3)}'$ and the selected units for each replicate were drawn <u>with</u> replacement. The replicates, however, were drawn <u>without</u> replacement, which means that the with replacement variance formula is conservative.

Moreover, $Y_{h(1,2,3)}$ (unlike $Y_{h(1,2,3)}'$) can use information from selected units that provide usable data in one survey but not in another, so it will often have less variance than $Y_{h(1,2,3)}'$.

Equation (59) can be plugged into (57) which then serves as a variance estimator for Y_h.

Item-to-item ratio estimation follows along the lines laid out in Chapter VII. Of course, M_{jk} in equation (55) and U_h in (58) need be modified to include only tracts or selected units with usable item values in all three surveys for both the numerator and the denominator. The covariance analogue to (59) is

$$V_{h(1,2,3)\cdot xy} = [u_h/(u_h - 1)] \times \left[\sum_{i \epsilon U_h} y_i^r x_i^r - \left(\sum_{i \epsilon U_h} y_i^r \right) \left(\sum_{i \epsilon U_h} x_i^r \right) / u_h \right] \quad \text{when } u_h > 1 \qquad (60)$$

$$= Y_{h(1,2,3)} X_{h(1,2,3)}/2$$

when $u_h < 2$ and both $Y_{h(1,2,3)}$ and $X_{h(1,2,3)}$ exist

is undefined otherwise.

Aggregation is fairly straightforward also. For direct expansions, equation (52) holds at every level of aggregation. The same applies to the numerator and denominator of an item-to-item ratio. For variance estimators, the rule that the variance (or covariance) of a sum equals the sum of the variances still applies. As long as the component parts consist of state NOL variance estimators (equation (54)), list stratum variance estimators (equation (57)), or their sums.

The List Component of Variance for October

The variance estimator for the list stratum total, $Y_h = 13(Y_{h2} + Y_{h3} + Y_{h4} + Y_{h5})$, has the form:

$$V_h = 169\{V_{h(2,3)} + V_{h(4,5)}\}, \qquad (61)$$

where $V_{h(2,3)}$ is a variance estimator for $Y_{h(2,3)}$, and $V_{h(4,5)}$ is a variance estimator for $Y_{h(4,5)}$.

The 1989 list sampling design follows the same replicate pattern as the 1988 survey. Consequently, to estimate the variance of $Y_{h(2,3)}$, a selected unit in Replicate 3 of the January 1989 survey is matched (before the SPS Summary) to a selected unit in Replicate 1 of the April survey (both surveys enumerate Replicate 4). Likewise, to estimate the variance of $Y_{h(4,5)}$, a selected unit in Replicate 1 of the July 1989 survey is matched to a selected unit in Replicate 3 of the October survey.

Paralleling equation (58), let

$$Y_{h(2,3)}' = \sum_{i\epsilon U_h} \sum_{t=2}^{3} e_h^* y_{it},$$

$$= \sum_{i\epsilon U_h} y_i^r \qquad \text{if } u_h > 0 \tag{62}$$

$$\text{is undefined} \qquad \text{otherwise,}$$

where U_h is the set of all matched selected units in list stratum h with usable item responses in all both surveys (January and April),
$e_h^* = N_h/u_h$,
N_h is the total number of names in list stratum h,
u_h is the number of selected units in U_h,
y_{it} is the month t survey item values for selected unit i,
and $y_i^r = \sum e_h^* y_{it}$, where the summation is over t = 2 and 3.

A conservative estimator for the variance of $Y_{h(2,3)}$ is

$$V_{h(2,3)} = [u_h/(u_h - 1)] \times \left[\sum_{i\epsilon U_h} (y_i^r)^2 - \left(\sum_{i\epsilon U_h} y_i^r\right)^2/u_h\right] \qquad \text{when } u_h > 1 \tag{63}$$

$$= (Y_{h(2,3)})^2/2 \qquad \text{when } u_h < 2 \text{ and } Y_{h(2,3)} \text{ exists}$$

$$\text{is undefined} \qquad \text{otherwise.}$$

It is a straigtforward matter to modify equations (62) and (63) to provide a variance estimator for $Y_{h(4,5)}$ (the subscript (2,3) is everywhere

replaced by (4,5), U_h is concerned with usable item values in July and October, and the summation for y_i^r is over 4 and 5). Results from equation (63) can then be plugged into (61) which then serves as a variance estimator for Y_h.

Extensions to item-to-item ratio estimation and aggregation follow along the lines laid out in the last section (with the subscript (1,2,3) in equation (60) replaced by (2,3) and (4,5)).

X. GLOSSARY OF TERMS

This glossary is not intended as a lexicon of mathematical terms used in the text, rather it is a listing of terminology that is somewhat distinct to the surveys covered by the 1989 SPS Summary.

Terms

All terms in bold face that are not underlined are contained in the alphabetical listing below. Terms both bold faced and underlined are not in the listing. These are either obvious or begin with an acronym. There is an alphabetical listing of acronyms in the following section.

Agricultural (Ag) Services Survey - That component of the **Labor Survey** in California and Florida that involves labor and wages from labor provided by contractors that need not themselves be farms.

Ag Services Item - A survey **item** related to labor or wages from labor provided by contractors (note: in states other than California and Florida, such items come from the **basic labor survey**).

Agricultural Survey - National **multiple frame survey** conducted in June (the **base survey**), July, August, September, January, and March of crops (acreages, production, and planting intentions), livestock, land in farms, number of farms, and grain and rice stocks (not every question is asked each month). Sometimes (incorrectly) called "the Quarterly Agricultural Survey " or "QAS."

Area Frame - The complete partition of a state into nonoverlapping area **segments**. A stratified sample of **segments** is drawn for use in the **base survey**. A subsample of **tracts** contained within those **segments** is drawn for each **follow-on survey**.

Area Frame Estimator - An estimator (usually a **direct expansion**) either for the entire state or for that part of the state not made up of **overlap** farms or **extreme operators**. There are three types of area frame estimators: **closed estimators**, **open estimators**, and **weighted estimators**.

Area Segment - See **segment.**

Base Survey - The June **Agricultural Survey.**

Basic Labor Survey - A survey of potential farms and area **tracts** concerning labor and wages from labor. The basic labor survey is the only component of the labor survey in states other than California and Florida. Sometimes called the **farm labor survey**.

Closed Estimator (also known as the **closed-segment estimator**) - An **area frame estimator** used only for the **base survey** that enumerates **item** values contained on **tracts** identified to be within sampled **segments.** The **NOL-closed estimator** enters zeroes for **item** values on **tracts** that are part of **overlap** farms. The **NEO-closed estimator** enters zeroes for **item** values on **tracts** that are part of **extreme operators.**

Count (Variable) - A binary variable determined from survey responses that indicates when a reporting unit is within a particular category; e.g., whether it has hogs.

Direct Expansion - An estimate of a total calculated by first multiplying survey values by an **expansion factor** and then summing the results.

Expansion Factor - A number (which may be a fraction) that in some sense refects how many potential farm names (in a **list frame estimator**) or **tracts** (in an **area frame estimator**) are being represented by a survey value.

Expanded Count - The **direct expansion** of a **count variable.**

Expanded Value - A survey value (sometimes called "an unexpanded value") times its **expansion factor** and its **list adjustment factor.**

Extreme Operator (EO) - A name on a **list frame** associated with a presumably large farm in terms of crops, livestock, or storage capacity. All EO's are **overlaps** and are found within **list strata** containing only EO's.

Farm Labor Item - A survey **item** from the **basic labor survey** relating labor or wages from labor

employed directly by a farm. Sometimes called a **basic labor item**.

Follow-on Survey - A **labor survey**, a **reimbursable survey** or an **agricultural survey** other than the **base survey**. This multiple frame survey enumerates a subsample of **tracts** contained in those area **segments** sampled for the **base survey** as well as a sample of names associated with potential farms from the relevant **list frame**.

Hand Imputed - The act of filling in an unreported survey value or changing an obviously incorrect one (as determined by an edit program) based on the judgement of some NASS employee or employees (usually at a State office) or by the edit program.

Imputed Value - A value determined by NASS that replaces an unreported or obviously incorrect survey value (see **hand imputed** and **machine imputed**).

Item - A survey question calling for a quantitative response or a mathematical function of such a response or responses (e.g., a **count variable**). Also called a **survey item**.

Labor Survey - A quarterly (July, October, January and April) **follow-on survey** of labor and wages from labor on farms whether the labor is directly employed or provided by **agricultural services**.

Land-Use Stratum - A collection of segments in the area frame of a state that are thought to have similar land-use characteristics (intensive cultivation, range land, body of water, etc.).

List Adjustment Factor (LAF)- A scalar adjustment to **item** values to account for more than one **sub-tract** being associated with the same farming arrangement.

List Frame - A list of names associated with potential farms within a state that is eligible for the list sample of a particular survey. Thus, there are different list frames for different surveys. Also known as a **list universe**. (NASS sometimes refers to the collection of all the list frames or universes in a state as the list frame for that state. In this report, we always mean

the frame for a particular survey.)

List Frame Estimator - An estimator (usually a **direct expansion**) for all farms associated with names on a **list frame.** This estimator is based on a stratified simple random sample of names drawn from the **list frame.**

List Stratum - A collection of names within a **list frame** that are believed to be associated with farms with similar characteristics.

Machine Imputed - The act of filling in an unreported survey value or replacing an obviously incorrect one (as determined by an edit program) by means of an imputation progam.

Multiple Frame Estimator - An estimator combining estimates for two (or more) frames. The component frames need not be mutually exclusive but their combination should exhaust the target population. For the base survey, the **OL-MF estimator** combines an estimator for the **list frame** with an **NOL** (**closed** or **weighted**) estimator for the **area frame.** Also for the base survey, the **EO-MF estimator** combines an estimator for the extreme operators in the **list frame** with an **NEO** (**closed**, **open**, or **weighted**) estimator for the **area frame.**

Nonoverlap - Farms, or tracts on farms, that are not associated with names on a particular **list frame. Follow-on surveys** have **nonoverlap strata** (see **select** and **summary strata**) made up of nonoverlap tracts enumerated in the base survey.

Open Estimator (also known as the **open-segment estimator**) - An **area frame estimator** used for the base survey that enumerates whole farm values for **items** when the farm operator lives on a **tract** identified to be within a sampled **segment** - with the following exception - the **NEO-open estimator** records zeroes for **item** values on **extreme operators.** A subsample of all **non-overlap** (NOL) farms where the farmer lives within a sampled segment is used for the **labor survey.**

Overlap - A farm, or a tract on a farm, that is associated with a name on a particular **list frame.**

Positive Nonusable - A survey **item** response that is not valid but for which we have the following information: the **sampling unit** is recorded as having a positive quantity of the **item** or a related **item** and a nonzero **list adjustment factor** (e.g., a farmer may indicate that he has hogs but refuse to give information on particular hog questions like his number of market hogs). For the purposes of this report, only a livestock item response can be a positive nonusable.

Presence Estimator - A livestock estimator for a list stratum composed of an estimator for the fraction of sampled units with livestock of the type in question times the estimated mean among such units.

Ratio Expansion - An estimator for a current total that is the product of, (a), a ratio estimator containing direct expansions for the current total in the numerator and the previous total in the denominator, and, (b), an estimator for the previous total (called the **root**).

Root (estimator) - That part of a ratio expansion that is multiplied by the ratio estimator. The root is an estimator for the denominator of the ratio estimator that is not restricted to matched respondents.

Reimbursable Survey - A **follow-on** survey conducted by NASS for another agency or Congress. Such a survey is often a one time event for which NASS will be reimbursed.

Segment - A usually contiguous piece of land. Each state is divided into segments (see **area frame** and **area frame estimator**). Also known as an **area segment**.

Select Stratum - A collection of **NOL tracts** enumerated by the **base survey** with similar characteristics. A simple random sample of tracts from a select stratum is enumerated for a **follow-on survey**. Also known as a **NOL select stratum**.

Selected Unit - A name from a **list frame** or an NOL tract from a **select stratum** that has been sampled for a survey.

Substratum - A subdivision of a **land-use stratum** used in selecting the sample or area **segments** for the base survey. Every segment is in one and only one substratum.

Sub-tract - A reporting unit for a survey. Each **selected unit** has at least one sub-tract. (NASS often calls a reporting unit from the **list frame** in the base survey a "tract" rather than a "sub-tract." We do not follow that practice here. There is no distinction between area tracts and sub-tracts in the base survey).

Summary Stratum - A collection of **NOL tracts** enumerated by the **base survey** with similar characteristics, usually made up of one or more select strata. The sample for a **follow-on survey** may only contain a single tract from a particular **select stratum,** but there must be at least two sampled tracts from each summary stratum. Also known as a **NOL summary stratum**.

Total Labor Item - The combination of a **farm labor** and **ag services** survey **item**.

Tract - That part of an area **segment** that is under one land operating arrangement.

Usable (response) - A valid response for a survey **item**. May be given by farmer (or agent), hand-imputed by NASS (usually a state office), or calculated from other usable items (as in the case of a **count variable**). Note: On occasion a missing or invalid response is treated as a usable zero for estimation purposes.

Variance Stratum - The union of several similar area substrata. It is treated as a single substratum for variance estimation purposes.

Weighted Estimator (also known as the **weighted-segment estimator**) - **An area frame estimator** that prorates whole farm values for **tracts** identified to be within sampled **segments** (the area in the tract divided by the area in the farm as reported in the base survey is used for prorating). The **NOL-weighted estimator** records zeroes for **item** values on **tracts** that are part of **overlap** farms. The **NEO-weighted estimator** records zeroes for **item** values on **tracts** that are part of **extreme operators**.

Yearly Aggregate - A statistic produced for the **labor survey** (**basic**, **ag services** or **total**) that combines survey values for the present and previous three quarters.

Acronyms

The text uses many acronyms. The original form of each survey related acronym is given below. When that form can not be found in the previous section of this glossary, direction to a related term will be given.

EO - Extreme Operator; can be used along with a MF (multiple frame) estimator, as in "EO-MF estimator" (see multiple frame estimator)

MF - Multiple Frame

NEO - Not an extreme operator (see extreme operator); can be used along with an area frame estimator, as in NEO-closed estimator (see area frame estimator, closed estimator, open estimator, weighted estimator)

NOL - Nonoverlap (or "not-on-the-list"); can be used along with an area frame estimator, as in NOL-closed estimator (see area frame estimator, closed estimator, open estimator, weighted estimator); can also be used along with terms select and summary stratum, as in NOL summary stratum.

OL - Overlap (or "on-the-list"); can be used along with a MF (multiple frame) estimator, as in OL-MF estimator (see multiple frame estimator)

XI. REFERENCES

[1] SPECIFICATIONS -- Agricultural Surveys: June 1989 - March 1990, Estimates Division, October 1988.

[2] Quarterly Labor Survey Specifications, Estimates Division, August 1988.

[3] Agricultural Surveys Supervising and Editing Manual, December 1989 - March 1990, Estimates Division, 1989.

[4] Agricultural Labor Surveys Supervising and Editing Manual, October 1989 - April 1990, Estimates Division, 1990.

[5] Houseman, Earl E. Area Frame Sampling in Agriculture. SRS-20, 1975.

[6] Cotter, Jim & Nealon, Jack. Area Frame Design for Agricultural Surveys, Research and Applications Division, August 1987.

[7] Atkinson, Dale. The Scope and Effect of Imputation in Quarterly Agricultural Surveys, RAD Staff Report No. SSB8804, February 1988.

[8] Herzog, Thomas & Rubin Donald B. "Using Multiple Imputations to Handle Nonresponse in Sample Surveys" in W. G. Madow, I. Olkin, and D. B. Rubin, Eds., Incomplete Data in Sample Surveys, Volume 2: Theory and Biliographies. Academic Press, New York, 210-245, 1983.

[9] Oh, H. & Scheuren F. "Weighting Adjustment for Unit Nonresponse" in W. G. Madow, I. Olkin, and D. B. Rubin, Eds., Incomplete Data in Sample Surveys, Volume 2: Theory and Biliographies. Academic Press, New York, 143-83, 1983.

[10] Kott, P. S. & Thorson, J. Improving Variance Estimates for Livestock Surveys, RAD Staff Report No. SRB-NERS-8901, February 1989.

[11] Geuder, Jeff. Paper Stratification in SRS Area Sampling Frames, SF & SRB Staff Report No. 79, February 1984.

[12] Kott, P. S. Estimating Variances for the June Enumerative Survey, RAD Staff Report No. SRB-NERS-8806, May 1988.

[13] Kott, P. S. & Johnston R. Estimating the Non-Overlap Variance Component for Multiple Frame Agricultural Surveys, RAD Staff Report No. SRB-NERS-8805, May 1988.

[14] Kott, P. S. "Estimating Variances when the First Phase Sample is Restratified," Survey Methodology, forthcoming.