THE
QUARTERLY
AGRICULTURAL SURVEYS

R.R. BOSECKER

July 1987

National Agriculture Statistics Service
U.S. Department of Agriculture
Washington, D.C.

CONTENTS

THE QUARTERLY AGRICULTURAL SURVEYS

I. Introduction

The Quarterly Agricultural Surveys (QAS), conducted by the National Agricultural Statistics Service (NASS), provide inventory and production estimates for crops and livestock at state and national levels. The QAS utilizes two frames: a list frame for sampling efficiency and an area frame for coverage completeness. The sampling unit for the list frame is a name. The area unit for the area frame is an area of land (segment). The reporting unit in both cases is land operated by one or more persons under a single land operating arrangement. Each quarter a list sample of farm operators (75,000) are contacted by mail, telephone or personal interview for inventory information on the land they operate. Sample segments (16,000) selected from the area frame are also screened for farm operators (55,000). The multiple frame estimator utilized by NASS requires the matching of names between the two frames to identify those in the area frame who had no chance of selection for the list. These are referred to as nonoverlap (NOL) operators. The NOL estimate from their data compensates for the incompleteness of the list, thereby completing coverage of the target agricultural population.

II. List Frame Construction and Maintenance

The purpose of the NASS list frame for the QAS is to improve sampling efficiency. Names, addresses, phone numbers, and measures of size for farm operators permits stratification for more efficient sample selection and allows the use of less expensive survey methods for more efficient data collection. This list is not expected to be complete. Farming operations go in and out of business too quickly to expect to have a complete list. However, considerable gains in efficiency can be expected from utilizing a list frame containing a significant proportion of the larger operations.

Incompleteness of the list is not a coverage problem when the list is backed up by a complete area frame. The list incompleteness contributes to less sampling efficiency in the combined estimator, but there would be no theoretical coverage bias due to omission of population units. Duplication in the list also lessens sampling efficiency, but one may appropriately compensate for it by the detection of duplicates in the sample (Gurney and Gonzalez, 1972).

The NASS list frame for the QAS is the result of numerous input sources of farm operators. Many of the same operators appeared on several source lists. Therefore, two to three times as many names were brought together during creation of the list than eventually comprised the list frame. Record linkage procedures based on work by Fellegi and Sunter (1969) and described by Coulter and Mergerson (1977) were used to standardize and remove duplication in the construction of the final composite list.

Master lists were built for several states in 1979, and all states were using the list frame system by 1982. After the initial "build" phase, a continuous maintenance program has been in place to keep the frame current. New operations are added, those no longer

operating are deleted, and the "control" data associated with each active operation for stratification purposes are updated as new information becomes available.

Six subsystems are utilized by NASS to facilitate management and utilization of the list sampling frame. The Source List Editor Subsystem (SLEDSS) standardizes input records into a common format; reduces identically matched records to a single record; identifies all components within the name and address of each record; and codes all names as individual, partnership, or corporate. A Record Linkage Subsystem (RECLSS) employs different linkage procedures for each class of names (individual, partnership, or corporate) to group potential duplicates by class into definite links, possible links, and nonlinks. The Group Resolution Subsystem (GPRESS) codes a record to represent each linkage group, matches across the linkage groups and classes and outputs the computer identified possible matches for visual inspection. Duplication is kept to minimum by removal of the computer-determined definite links and the identified duplicates among match groups.

The fourth list sampling frame subsystem is Data Select. This program determines the "best" control data from among several input sources to attach to the list record. This might be the largest value, most current value or the one coming from the best source depending upon guidelines specified for each of the variables of interest. The Sample Select Subsystem then stratifies the list frame based on control variables for many different surveys and selects multiple samples simutaneously.

Finally, the Mail and Maintenance Subsystem is a frame and sample management system to create mailing lables and/or listings, amend the list frame with transactions from surveys, create special comments or changes for specific surveys, provide a history or "tracking" for all changes after sample selection, and combine or organize survey samples for special needs such as elimination of multiple contacts for different surveys in the same time period.

Utilization of all components of the list frame system provides the means to maximize list coverage for the agricultural variables of interest and minimize duplication within the list. Remaining undercoverage in the list is compensated for through the area frame sample and remaining list duplication is adjusted for based on information received for the sample.

The most serious problem that could befall the list frame in the QAS multiple frame context would be for names from the area frame sample to somehow be added to the list. This would compromise the ability of the area frame sample to estimate for the proportion of the population of farm operators who are not on the list. The necessity for independence between the list and area frames is emphasized in all NASS training manuals and classes. A thorough discussion of the potential for list contamination and the consequences are given in Vogel and Rockwell (1977).

III. Area Frame Construction and Maintenance

The primary purpose of the area frame is to provide a probability mechanism for estimating the entire population of crops and livestock in the U.S., i.e., completeness. Since all crop acreage and all livestock are physically located on land, complete

representation is assured if the total land area is divided into sampling units. The description of how this is accomplished through land-use stratification by state and county throughout the United States is given by Nealon and Cotter, "Area Frame Design for Agricultural Surveys" (1987).

In the QAS the area sample supplements the list sample, accounting for list incompleteness, to provide full coverage of the agricultural population of interest. Typically the area frame nonoverlap (NOL) domain estimates for 10 to 20 percent of the total crop and livestock inventories.

Since completeness is a primary function of the area frame there are a number of control practices in place to insure all land is represented without duplication or omission. First, a premium is placed on the use of good, identifiable, permanent boundaries which can be marked on maps and photographs and recognized by interviewers at the site. Land use stratification boundaries and sample unit boundaries are drawn to provide a clear demarcation even at the expense of some sampling efficiency if necessary.

Second, the areas defined by strata and clusters of sample units are electronically digitized so the total for the frame can be computed and compared with the known land area for the county and state. The accumulated state area is allowed to vary $\pm$ 0.5 percent from the published area.

Both the original frame materials containing the boundaries and a graphic representation of the digitized boundaries are reviewed for completeness.

Finally, the selected sample segments are also digitized to determine land area. Interviewers accumulate reported acres in each segment and compare the reported total against the digitized total. This control insures complete coverage of each sample segment.

IV. Rules of Association

A given area of land may be represented in the QAS in several ways through the list frame as well as appearing also in the area frame. The operation may have a name unto itself as well as having the name(s) of one or more operators associated with the land. Any of several partners may be sampled and provide the information requested for the same parcel of land.

To control this potential for duplication there are several rules of association set forth in interviewer instructions and in supervising and editing (S&E) manuals. A list dominant rule provides for the list frame to account for any land which may reached through the list frame, that is, an area of land may belong to the area nonoverlap domain only if none of the names associated with the land are represented on the list.

Within the list frame, potential for duplication is controlled through priority rules governing which names associated with a given parcel of land will be considered the dominant sample unit. All data for an operation will be associated with the list name assigned greatest priority. An operation name, if any, is given top priority because the

name tends to be attached to land operated under that title for a longer peirod of time than the names of individual operators. This is particularly true in the case of managed land where the operation must have its own name appear on the list to be considered overlap. The name of the hired manager is not used to determine the overlap status of the operation.

In the absence of an operation name or if the operation name is not on the list, the land area may be represented through the list frame by a combination of the individual's names who comprise a partnership (second priority) or, finally by the name of any individual actively considered an operator, alone or in partnership, if he participates in making the "day-to-day" decisions affecting the farming of land.

Partnership operations are a particularly difficult situation. It must first be determined that a true partnership in operating the land exists, i.e., more than one person jointly operate the land. Thus, since each person can report for that operation, a rule to account for the data only once is needed. To do this the QAS utilizes a "largest stratum" rule. If more than one partner is on the list, the data will be accepted only from the partner in the largest list frame stratum. If more than one partner belongs to the same largest stratum the data will be divided equally between them. The procedure minimizes the division of data among sampling units and attaches the data with operators having the largest stratifying control data and smallest expansion factors.

By far the largest portion of the list frame is comprised of individual names who operate their own farms. However, an individual may be involved in more than one operating arrangment. According to the above rules of association, an individual should report for each of the different land operations in which he is an active operator. For example, if he has an individual operation and is a partner in another operation, he should provide a report for each of them. Each operation will then be considered separately according to the priority rules governing its representation on the list frame.

All of these rules presuppose the collection by the interviewer of <u>all</u> associated names for a farm operation. Emphasis is given during training, in the instruction manuals, and on the questionnaires to the importance of providing the operation name, if any, and the name or names of all operators. If the operation was found in an area frame sample segment, all of the names will be checked against the list. When any of the names are found, the list frame identification number is attached to the corresponding name and the operation belongs to the overlap domain. If the operation was sampled through the list frame all other associated names are obtained and checked against the list. Investigations into the operational application of these rules were reported by Bosecker and Kelly (1975), Hill and Rockwell (1977) and Nealon (1984).

V. Error Avoidance

Quality control measures during construction of the frames and unbiased survey design are only the first steps in ensuring proper coverage. The rules governing the representation of each population unit must be adhered to during data collection. The QAS makes use of written instructions, formal training, active supervision, questionnaire prompting, performance evaluations and reinterview samples to aid and monitor interviewer

activities. Completed questionnaires are reviewed by office personnel and submitted to computer edit and analysis both within and between questionnaires.

Problem areas requiring a great deal of attention to minimize coverage errors in the QAS include:

1) obtaining all names actively associated with the sampling unit,
2) determination of the nonoverlap domain,
3) obtaining an accurate report of the total acres being operated,
4) reporting all data, regardless of ownership, on land operated,
5) nonresponse.

Beller (1979) documented these areas of concern in "Error Profile for Multiple Frame Surveys."

If interviewers do not obtain all the names appropriate for the sampling unit, the rules of association described earlier cannot be applied properly. Errors could lead to either omission or duplication depending upon the frame from which the unit was sampled and the status of the missing names on the list frame. The importance of obtaining all legitimate names for the unit is impressed upon interviewers. Names from each survey are retained for verification in following surveys.

Even if all names are available, it is not always an easy task to determine whether a name from the area frame is the same as one on the list. More than one individual may have the same name near the same location. In these cases middle initials, telephone numbers, social security numbers and other identifiers help determine true matches. Spellings may differ slightly or nicknames may have been used. Great care is exercised to investigate possible matches. Even after the pairings have been made and the list identification number has been attached to the name from the area sample another verification on a computer listing of matched names is required.

An accurate report of the total acres operated is important to the area frame estimates of list incompleteness, i.e., the nonoverlap domain component. This stems from the weighted-type estimator employed. Whole farm data is pro-rated for the area nonoverlap respondent proportional to the amount of land operated inside the sample segment versus the total land operated. Complete coverage is achieved and duplication is avoided when the sum of a farm's land parcels across all possible area frame sample units equals the total farm size, i.e., the sum of proportional weights equals one.

The potential for reporting error exists because the respondent may not include all types of land when he provides the total farm acres. The portion of the farm inside the area sample segments is outlined on an aerial photograph facilitating a full accounting of the acreage. The remaining acreage in the farm is more dependent upon the respondent's concept of acreage to report. The questionnaire is designed to remind the respondent of all the acres he operates, whether owned, rented or managed, and all types of land, including woods, waste and roads, so the land outside the sample segment is reported comparably to the land inside.

Once the total land operated has been established, all requested data is to be reported regardless of who owns the crop or livestock commodity. This applies to both area and list sample respondents. Emphasis on this concept is required because a natural inclination of some respondents is to report only what they own. Since coverage for the QAS is dependent upon accounting for the variables of interest through the acres where they are located, a lot of effort is expended to ensure compliance with the concept and accuracy in the reported data.

Nonresponse in the QAS varies by state but typically ranges from 10 to 20 percent. Two types of adjustments are made to extend data from only the respondents to all on the sampling frames which in turn represent the total agricultural population. The first procedure adjusts sample sizes downward to the number of respondents by list stratum. Therefore, the assumption is made that whithin each list stratum, nonrespondents share the same agricultural characteristics as respondents.

Evidence that respondents tend to be smaller than nonrespondents is provided by Gleason and Bosecker (1978) and Crank (1979). Therefore a second approach is also used. Information is provided by the interviewer through observation or secondary sources on the presence or absence of individual commodities for nonrespondents. Through imputation or summary adjustment this information is used to more appropriately utilize respondent data for nonrespondents with similar operations.

A coverage problem posed by nonrespondents which is sometimes overlooked concerns the status or classification of the sampling unit as a viable operating entity. Simple adjustments to sample sizes for nonresponse assumes the same proportions of nonrespondents are out of business as for respondents. Imputation may assume the unit is in business. However, there are two main sources of nonresponse--refusals and inaccessibles. Refusals most often have the items of interest (which they do not want to report so they refuse) while inaccessibles may be in business but unreachable or may not be found because they are out of business.

Nonrespondents in the QAS are coded for their in or out of business status based on available information in the same way individual commodities are coded for their likelihood of existing on an operating unit. Units having no evidence of operating currently may therefore be more properly handled as a zero contribution sampling unit.

Identification of error sources in multiple frame surveys has been the subject of numerous special studies and survey quality assurance programs in the National Agricultural Statistics Service. Expenditures for these efforts are unlikely to decline since full coverage of the agricultural sector is the desired product to be achieved.

VI. Comparative Analysis

Many of the commodity totals estimated through the QAS move through the agricultural marketing channels and are therefore amenable to comparison with administrative data. Some examples of this include slaughter data for hogs and cattle, milk production for dairy cows, crushings for soybeans and sales of cotton. Even though all of the commodity may not be accounted for through one source or process, a limited number of possibilities affords the opportunity to construct a "balance sheet" to account for total production.

For example, survey measurements of soybean production for 1987 plus carryover soybean stocks in storage establish the total available soybeans for the 1988 marketing season. By monitoring exports, soybean processing, seed use, imports and remaining stocks in storage at the end of the cycle a reasonable accounting of the uses for the total available soybeans can be made. Because of sampling errors in surveys and imperfections in administrative records there will be "residual" or unexplained differences between supply and use. However, differences exist within reasonable limits and can be monitored over time. Problems in survey coverage become readily apparent with this type of check data.

Another useful comparison is with the Census of Agriculture at five year intervals. This census of agricultural operators provides inventory numbers for a specific date (December 31) and production statistics for the census year. The census has its own problems in achieving complete coverage and is of course subject to nonsampling errors during data collection. However, the target propulation is the same as for the QAS and differences between the two measurements leads to useful analysis for evaluating coverage.

The checks and balances which exist for the QAS estimates subject the results of this survey to a scrutiny by the data users which is rare among government surveys. Measurements for given dates are verifiable by subsequent events. This puts the issue of coverage as a very high priority concern at the National Agricultural Statistics Service.