Asia and Pacific Commission on Agricultural Statistics

Twenty-fourth Session
Da Lat, Viet Nam, 8-12 October 2012
Agenda Item 7
United States- Agriculture Census and Surveys

**2012 United States
Census of Agriculture**

Robert Hale, Mathematical Statistician
International Programs Office
National Agricultural Statistics Service
United States Department of Agriculture
robert.hale@nass.usda.gov

The United States has been fortunate to be able to conduct a census of agriculture on a regular basis for many years. Currently the National Agricultural Statistics Service (NASS) of the United States Department of Agriculture (USDA) has that responsibility. The Census of Agriculture is the only source of such complete agricultural information at the lowest administrative level available in the US. The Census of Agriculture is conducted on a five year cycle with reference years ending in 2 and 7. 2012 fits that criterion. NASS has been preparing for the Census of Agriculture for a few years. Questionnaire design and testing, updates to our list of farm operations, and much of the software testing is complete. The data collection process will begin in January 2013. Several methods of data collection will be used as well as different forms of data entry. Data processing and analysis will continue through the summer and autumn of 2013. The 2012 Census of Agriculture will be published in February 2014.

This paper provides an overview of the steps NASS will take to conduct the Census of Agriculture. Key milestones will be noted for the preparations, data collection, summarization, and dissemination of the US Census of Agriculture.

Introduction

The United States Department of Agriculture (USDA) National Agricultural Statistics Service (NASS) is responsible for conducting the US Census of Agriculture. NASS is also responsible for all follow-on censuses related to agriculture. The Census Bureau of the US Department of Commerce had this responsibility until 1997. At that time there were discussions within the Census Bureau to change the official definition of a farm by raising the minimum sales qualification. The USDA said we would conduct the Census of Agriculture and maintain the same farm definition. For that reason, the US Congress decided to move the responsibility of the Census of Agriculture to the USDA.

The official definition of a farm in the United States is: "Any place that produced and sold, or normally would have produced and sold, $1,000 or more of agricultural products during the calendar year." Note there are no minimum qualifications for the area of the farm or the number of livestock. Given this definition, there are about 2,200,000 farms in the US.

The US Census of Agriculture is conducted every five years, with reference years ending with either 2 or 7. All information collected is for the entire calendar year except current livestock inventories. A 31 December reference date is used for livestock inventories. The next Census of Agriculture will have a 2012 reference year. The focus of this report will be the preparations for the 2012 US Census of Agriculture.

Data collection for the US Census of Agriculture is different from most countries. Most data are collected through the mail. We mail the questionnaire to the farm operator, with some instructions and ask the farm operator to complete the questionnaire and return it to us by the mail. We will follow-up with phone calls to those who do not respond. Our last choice is a personal interview. We will make special arrangements for the very large farms to collect the data when and how they choose. Another difference from many countries is that the US Census of Agriculture is not conducted following a population census. The US Population Census is conducted on a ten year cycle in years ending 0.

Since the Census of Agriculture is conducted every five years, we have developed a five year cycle for our work. During year 1 we evaluate the processes used in the previous census and start to plan for the next census. In year 2 we build the specifications for the next census and set the content of the questionnaire. In year 3 we develop and test the questionnaire and computer systems needed to edit, summarize and publish the data. In year 4 we collect and process the data. And in year 5 we publish the data. Also in year 5 we begin to collect data for the follow-on censuses. The follow-on's are spread out over several years and will overlap with the cycle for the next Census of Agriculture.

We will conduct follow-on censuses in six topic areas. Each of these topic areas has a section in the regular Census of Agriculture. 100% of the farm operators that report positive information in these areas are automatically is included in the more detailed follow-on census. We will conduct follow-on census for Farm and Ranch Irrigation, Renewable Energy, and Aquaculture in 2013. In 2014 we will conduct follow-on censuses for Land Tenure and Horticulture. A follow-on census on Organic Agriculture will be conducted in 2015.

Frames

Since the US Census of Agriculture is not associated with a population census, we have to use other sources for the names of farm operators to contact. NASS has two frames for our yearly operational program. First, we use a list of names and addresses of all of the farm operations we know in the United States. This is what we call our List Frame. Second, we have all of the land in the US separated into blocks that are categorized into groups based on the likelihood of finding agriculture. These blocks are sub-sampled and canvassed to find farming operations. This is what we call our Area Frame. NASS uses both of these frames to estimate the total agricultural output of the United States.

The List Frame is the primary sampling frame used for the NASS operational program. We are constantly working to keep it as complete as possible. Every day addresses change, new operators begin farming and others stop. We work with farm organizations, other federal agencies as well as state and local organizations to gather information to maintain the list. However, in the Census of Agriculture reference year we make a special effort to have our list as complete and up to date as possible. In the first months of the year, NASS conducted the National Agricultural Classification Survey to verify the farm status of questionable farm operations on our list frame. We mailed 1,236,000 questionnaires to possible farm operations on our list to correct as many as possible before the Census of Agriculture questionnaires are mailed. The responses to these questionnaires were used to verify farms, update basic information about these farms and delete the non-farms.

The US Census of Agriculture is essentially a 100% sample from our list frame. The results of the census are used to update the list frame on farm status, location, size and operator characteristics. This information will be used for the next five years in our operational program.

In June 2012 our enumerators visited 14,773 blocks of land randomly selected from our Area Frame. The size of the blocks varied depending on the type of land, however, in the primarily agricultural lands each block averaged one square mile. The enumerator's duty was to account for all of the land in the block and complete a questionnaire on any farm operation with any land inside the block. The questionnaire contained the name and address of each farm operator as well as basic information about the operation. Now we are comparing the names and addresses from the area frame sample to all of the names on the list frame. Any farm operations not on the list are noted and used later. The same process was used for the 2007 Census of Agriculture. That year we found 12,281 farms not on the list out of a total of 2,204,792 farms.

Questionnaire

We have nine different versions of our questionnaire. Seven versions are for different regions of the country. Most of the questionnaire is the same in each version, the differences are in a few sections where common items for that region are printed on the questionnaire rather than hand coded by the farm operator. For example, a cell is coded for cotton in the crops section for the southern regions (2 and 3) but not the northern regions (1 or 4). This is because the growing conditions for cotton are not suitable in the northern regions. However, if a farm operator in the northern regions did grow cotton he could hand code cotton on the questionnaire. There is also a special version of the questionnaire for Puerto Rico in

Spanish. The final version if for American Indian Reservations to handle the special way agriculture is conducted on Indian Reservations. Each questionnaire is 24 pages long and has 37 sections. Since our primary mode of data collection is through the mail, four pages of instructions are included with each questionnaire. All farm operators are requested to complete all 37 sections.

The questionnaire includes multiple sections on land in the operation, crops grown, livestock raised, operator and operation characteristics, and expenses. Questions in these sections cover all of the World Census of Agriculture core census items. Many of the 80 variables included in the supplementary modules are covered in one of the 37 sections.

Process

For the 2007 Census of Agriculture about 83% of the farm operators responded by mail. Mail is our most cost effective way to collect data. Included in each questionnaire is a method to respond via the Internet. The farm operator is given a login ID and password to a secure website which allows the operator to complete the questionnaire online. In 2007 about 4% of the farm operators responded via the Internet. We hope that percentage increases for 2012. If the farm operator does not respond by mail or the Internet, we will follow up with a telephone call. We will use a computer assisted telephone interview (CATI) process to guide the telephone interviewer through the questionnaire to collect and enter the data. We received about 7% of our responses from farm operators over the telephone in 2007. If all else fails, we will send an enumerator to the farm to collect the data face to face. This method is usually reserved for larger farms. We collected information from about 2% of the farm operators using personal interviews in 2007. You will note this totals to only 96%. Adjusting for the missing 4% will be explained later.

Once the questionnaires are returned a multi-step process is used to capture the data. A direct optical character recognition system will not be used. The majority of the questionnaires will be completed by the farm operators rather than by trained interviewers. Therefore, the handwriting and location of the data within a response cell will not be consistent. The first step is to scan the questionnaires for an image. The image is reviewed to verify all entries are legible and then the paper questionnaire is destroyed. The data will be hand entered from the scanned image and stored in a database. There will be a series of edits to check for consistency within the questionnaire. Many of the inconsistencies found are corrected within the computer edits without review. Some require a human review. Many are just confirmed by reviews but some will require further interaction. The reviewers will use the scanned image and the keyed data to determine the proper correction.

The summarization of the data is also a multi-step process. The data are summarized to our county level. County is the lowest administrative level in the US. The Census of Agriculture is the only comprehensive source of agricultural information at the county level in the US. The summary has three parts. First, is the total of all of the data reported by the farm operators. This accounts for the vast majority of all of the data. As noted earlier, we received responses from about 96% of the farm operators in 2007. The second part of the total is a non-response adjustment. This accounts for a small part of the total representing only the farm operations that did not report. Finally, an adjustment is made for the farm operations that are not on the list and were not eligible to respond. Again as noted earlier, we use our area frame to estimate the

number of farm operators we do not have on our list. The reported total, non-response adjustment and coverage adjustment are combined and calibrated to produce the final figures at the county, state, and US level.

The two adjustments themselves are each a multi-step process. The non-response adjustment is based on a decision tree analysis within each state. Ten variables will be used to build a decision tree to create non-response groups. There will be several hundred groups in each state. Every completed questionnaire will be assigned to one of the groups. To make the non-response adjustment each non-responding farm is assigned to one of these groups. Then the weight assigned to the reporting farm operators is adjusted to compensate for the missing information.

A different process will be used to adjust for the farm operators not on the list. Several inputs are needed to complete this adjustment. First are the non-response adjustment weights calculated in the previous step. Also needed are calibration targets for each state. Typically there are over 100 targets for each state. Next are a tolerance for the target and an allowable range for changes to the non-response adjustment weights. A very complex system of linear equations is solved to determine the coverage weight. Basically, the equations look at all of the non-response adjustment weights and the 100+ calibration targets and solve for the best solution to meet these targets within the limits specified. This is an iterative procedure which takes several iterations to determine the best solution.

In the end the coverage adjusted estimates for each reported farm operation will be;

$$\hat{t}_y = \sum_k g_k d_k y_k = \sum_k w_k y_k$$

where,
y_k = the reported value for farm k
d_k = non-response adjusted weight for farm k
g_k = coverage adjustment factor for farm k
w_k = final coverage adjusted weight for farm k

A tabulation will be run once the adjustments have been made to each reporting farm operation. A separate publication will be produced for each state and one for the US. There will be about 120 tables in each publication covering crops, livestock, economics, demographics, etc. This will mean there will be 6,120 tables with about 15,000,000 cells. Each will be reviewed for accuracy and checked to verify no individual farm operator's information will be disclosed.

Schedule

The 2012 Census of Agriculture for the US covers the agricultural activity for 2012. Our plan is to begin data collection on about 1 January, 2013 and have the publication released 4 February, 2014. To meet this target publication date we have established the following schedule. On 28 December 2012 we will mail 3,000,000 questionnaires to individuals and enterprises we believe operated farms in 2012. Each day we will track all of the returns by mail. On 2 January, 2013 the electronic data reporting website will be

opened to allow farm operators to respond via the Internet. The website will remain open until 31 May 2013. On 16 January, 2013 we will send a reminder/thank you post card to everyone who has not responded. The post card will thank anyone who may have responded after we created the response list and remind the other to please respond by 4 February, 2013. On 14 February, 2013 we will send a second complete set of materials to everyone who has not responded. A third mailing will be sent 25 March, 2013. On 11 February, 2013 we will begin to call some of the farm operators. Calling will continue through 14 June, 2013. In that same time frame we will start sending some interviewers to visit some of the large farm operators who have not responded and we were not able to contact by phone. The data capture process will begin 23 January, 2013 and continue through 14 June, 2013. Data edits will begin soon after the data capture process begins and continue through 14 June, 2013. By 13 April, 2013 we will have enough data edited and in the system to begin data analysis. The data analysis will continue through 30 September, 2013. An initial summary will be run 25 July, 2013 and a review of the summary will continue with the data analysis. A second summary with no disclosure checks will be run on 30 September, 2013. A complete summary review will be conducted and a final summary with disclosure checks will be run on 25 November, 2013. A final check of the summary will be completed and publication tables checked. The first publication at the county, state, and US levels will be released on 4 February, 2014.

Lessons Learned

This will be the fourth time NASS has the responsibility to conduct the US Census of Agriculture. Each time we learn ways to improve our methods. Changes have ranged from improved list building methods to specialized computer edit and summary systems. One major lesson for us was how to use optical scanning. Optical scanning for data capture did not work for us. That is why we scan for image only and then key from the images as a separate step. The questionnaire needs to be specifically designed for data capture to be successful.

In 2007 we tried using a short and long version of the census questionnaire. The concept was to reduce burden on the farm operators receiving the short version. The way we structured the shortened questionnaire was to include more open tables to report information. This caused multiple problems in processing and editing the data. Although we do have nine version of the questionnaire for 2012, all are structured the same.