A COMPARISON BETWEEN THE CANADIAN AREA FARM SURVEY AND THE CENSUS OF AGRICULTURE DATA

STATISTICS CANADA STATISTIQUE CANADA
OCT 31 1996
LIBRARY BIBLIOTHÈQUE

Alain Théberge
Statistics Canada

ABSTRACT

In 1993, the agricultural area survey was redesigned following the 1991 Census of Agriculture and was called the Area Farm Survey (AFS). The new design which has made use of geographical information systems is described. A comparison of the 1993 estimates from the newly redesigned Area Farm Survey and from a major multiframe survey is made with the latest 1991 Census numbers for total farm area and for number of farms by region and by agricultural product. Recommendations for improvements based on the results of these comparisons are put forth.

1. INTRODUCTION

For more than twenty years Statistics Canada has been conducting an annual area based sample survey to collect agricultural data. From 1983 to 1992, when a multiple frame design was used for the general purpose National Farm Survey, the primary purpose of the area survey was to provide coverage for farms that were not on the list frame. Under that design, land segments for the area survey were drawn in two stages. The first stage was the selection of enumeration areas (EA's) which correspond to the area canvassed by a Census enumerator. In the second stage, each selected EA was partitioned into a number of segments from which one or two segments were chosen for complete enumeration. The total farm area was asked on the questionnaire which was administered only to the farms of the list sample and those of the area sample not on the National Farm Survey list frame. This was sufficient for the production of multiframe estimates, but no estimates based on the area sample alone could be produced. A more detailed description of the design of the National Farm Survey can be found in Ingram and Davidson [1], and in Julien and Maranda [2].

In 1993, the area survey was redesigned, as were other major agricultural surveys, following the 1991 Census of Agriculture and was called the Area Farm Survey. Technological advances now provide the opportunity to segment the whole country, stratify the segments, and therefore to select a sample of them in one stage. This one stage sample allows better control over the inclusion probabilities, thus increasing the efficiency of the sample and making high weights due to nonresponse less likely. The biggest advantage may lie in having segments independent of EA boundaries, thus making sample rotation as well as future redesigns much less labour intensive. Under the previous design, rotation would often be limited to the second stage sample, in order to avoid constructing new segments.

The AFS sample can be used to complement the list samples of various other agricultural surveys with sometimes different list frames. A limited number of estimates based on the area sample alone can now be produced in addition to the usual multiframe estimates since all area sample farms are weighted. This permits a comparison with Census numbers without the inherent errors of multiframe estimates, namely the matching of the area sample farms to the list frames, that the production of these estimates requires.

A description of the AFS design is presented in the following section. In Section 3, AFS provincial estimates of number of farms and total farm area are compared with the Census figures. A similar comparison is done in Section 4, but by category of agricultural products. Estimates of number of farms by the multiframe Livestock Survey are compared with the AFS estimates in Section 5. Finally, Section 6 contains recommendations for improvements based on the results of these comparisons, as well as possible areas of investigation for future research.

2. THE AREA FARM SURVEY

The AFS frame covers nine of the ten provinces of Canada, with the province of Newfoundland being excluded because of the relatively low levels of agricultural activity there. In the Prairie provinces (Manitoba, Saskatchewan, Alberta) and in part of the Peace River District in the province of British Columbia, regular cells of 3 miles × 1 mile were created. These correspond to the fairly regular legal land description in use there. Elsewhere, cells of 3 km × 2 km were constructed using the Universal Transverse Mercator grid. The digitized boundaries of the cells were then overlapped with those of the EA's using geographical information systems. The area of the resulting polygons was also calculated. Cells that do not overlap EA's with farm headquarters were automatically discarded. The farm headquarter is defined as the residence of the farmer when it is located on the farm, or else as the farm main gate. Other cells were discarded after being looked up on a topographical map if they were completely inside a national or provincial park for example. At the same time, cells were combined when, for example a large part of them was water. This combining of cells was coded, captured, and a computer program ensured that the size of each group was reasonable, that the cells were adjacent and that they satisfied other desirable conditions. Each remaining cell, or group of cells if combining was done, corresponds to one of the segments that make up the survey frame.

The same subprovincial regions as those for the crop surveys were used in stratifying the segments of each province. A composite measure of agricultural activity based on the total numbers of cattle, pigs, and sheep, total farm area, and number of farms was computed to form strata within each subprovincial region. Statistics for the segments on the frame were arrived at by allocating EA Census totals to the overlapping segments in proportion to the area of the overlap. The number of strata is such that there are on average 25 sampled segments per stratum. Stratification in the province of Prince Edward Island was different and made use of digital satellite imagery to obtain more precise land use statistics for the segments. Images were available to cover this small province.

A sample of two 100 segments was allocated to provinces roughly in proportion to their number of segments on the frame while taking into account the sample sizes under the previous design and respecting a minimum size. Sample allocation to the strata was proportional to the square root of the stratum size. Each stratum's segments were then sorted randomly within Census Subdivision (usually municipalities) before systematic sampling was performed. To implement sample rotation, a second non-overlapping sample was selected in a similar manner and each of the two samples was broken into four stratified subsamples. Replacement of one of the subsamples can provide a rotation rate of 25%.

The sampled segments are plotted on topographical maps (scale of 1 to 250,000) which are sent to seven regional offices for the formation of interviewer assignments. A computer program can graph the assignments as well as provide various statistics to help ensure the assignment quality. The interviewers are provided with topographical maps (scale of 1 to 50,000) to help in locating the segments in their assignment. The interviewers must account for all the land within a segment. This is done by delineating the land of the various operators on an aerial photograph or a large scale map. All farm operators with land in a selected segment are asked identification information which will help in the matching operation. Total farm area and dichotomous variables on the agricultural products of the farm are also collected. The area of the farm within the segment is measured by the interviewer with the help of a grid. Under the old design, when the segments were selected in two stages, the segment boundaries would often be chosen so as to be easily recognizable in the field. This permitted the respondents to estimate themselves their farm area within the segment. However, a study by Bélanger [3] where a sample of segments had been digitized, compared the responses given by farmers to the digitized areas. It was found that on average, the responses underestimated the digitized value by 3% when the whole farm was inside the segment, and by 20% when only part of the farm was within the segment. Having the interviewer use a grid should thus provide better data. In much of the Prairie provinces and in the Peace River District of British Columbia where the regularity of land plots permits it, the interviews are conducted by phone, a more cost effective method than personal interviews which are used elsewhere.

The approximately 13,000 farms found within sampled segments are unduplicated against a register of known farms which comprises those enumerated in the 1991 Census of Agriculture. Most farm surveys use a subset of the 1991 Census farms as their list sampling frame. Farms of the AFS sample which cannot be matched to the list frame of a given multiframe survey are surveyed as part of the data collection activities of that multiframe survey in order to account for all farms not on the list frame.

The fraction of the total farm area inside the segment is used to arrive at segment totals. The estimates are then computed using the stratified estimator based on the segment totals. Another estimation method using directly measured values for the segment totals can be produced only for total farm area. Both of these estimation methods, weighted estimators and closed estimators, are used by the National Agricultural Statistics Service (NASS) of the United States. Bush and House [4] describe the design of the NASS area frame as well as other area frames. Although closed estimators are more precise for seeded areas, they require collecting additional data for all farms in the area sample when only those not on the list frame are of interest. The relative importance of the Canadian list frames compared to the American list frames can explain the differing strategies.

3. ESTIMATES BY PROVINCE

The estimates of the number of farms and of the total farm area with their estimated coefficient of variation (cv) are given in Table 1. Note that the total farm area is the total area operated by farmers, including wasteland, woodland, etc. To make these numbers comparable to those of the Census, the estimates have been adjusted to include farms on Indian reserves and institutional farms which are not part of the AFS survey population. The raw Census figures have been used for this adjustment.

The cv's show that estimates of farm area are in general more precise than estimates of number of farms. This is because most segments in a stratum have the same area, either 3 mi^2 or 6 km^2, and thus sampled segments are more likely to have similar areas operated by farmers than similar numbers of farms. This is especially true in the Prairies where most of the time all of the segment land is operated by farmers.

The AFS 1993 estimates are compared with the 1991 Census of Agriculture numbers in Table 2. We see for example, that in total, for all nine provinces covered by the AFS, the coverage rate of the Census is 90% in terms of number of farms and 99% in terms of total farm area. This is of course under the assumption that the number of farms and the total farm area have not changed between June 1991, when the Census was conducted, and May 1993, when the first AFS was conducted. The cv's given in Table 1 can be used as approximations for the precision of the ratios presented in Table 2, as linearization of the inverse of a random variable will show.

In the Prairie provinces (Manitoba, Saskatchewan, Alberta), 98% of farms were covered by the Census, but the area of the missing 2% is negligible. The undercoverage for the number of farms in these provinces is not statistically significant. The land in the Prairie provinces is broken into regular easily identifiable pieces, and this along with the large average farm acreage, probably helps reduce undercoverage.

According to the AFS estimates, in Québec and Ontario the total farm area is 10% higher than the Census figures. This is statistically significant, given the cv's of 4.6% and 3.2% for these estimates in these provinces. Contrary to the other provinces, in Québec the percentage of farms missed by Census is less than the percentage of missed area, which would indicate that the missed farms are larger than average in this province.

In British Columbia, there would appear to be nearly 40% more farms than the Census indicates. In terms of area, however, the Census figure is higher than the AFS estimate. This phenomenon can be explained by the numerous small operations/households in the province which are marginally involved in farming and were apparently counted by the AFS but not by the Census. Since the 1991 Census, there is no longer a minimum sales requirement to qualify as a farm. This makes a correct classification more difficult.

In Saskatchewan, Alberta and British Columbia, the AFS estimate of total farm area is slightly lower than the Census figure. The precision of the estimates is sufficient to explain the differences. However, the possibility of undercoverage of the area frame or non-identification of land belonging to farm operations should be kept in mind, especially if in future years similar results are obtained. The problems in identifying farmland could be caused by the use of telephone interviews in Saskatchewan and Alberta or incorrect classification of land where the part within the segment is not cultivated, even though the overall operation is agricultural.

There were some problems with the data collected in the Maritime provinces (Prince Edward Island, Nova Scotia, New Brunswick) which will be discussed below. It is unlikely that more than one farm in three was missed in New Brunswick, or that there is 23% overcoverage in Nova Scotia in terms of total farm area, and at the same time 14% undercoverage in terms of number of farms.

For Table 3, the farms in the AFS were matched with the farms in the Census and the total areas reported by the group of farms common to both occasions were compared. The table shows that there is very little difference in the responses to the two surveys in the Prairies. However, in Québec, Ontario and British Columbia, the same farms reported areas approximately 5% lower to the Census than to the AFS. Thus, part of the Census undercoverage of areas in Quebec and Ontario could be due to response error. This response error would lie mostly with Census, because as the following paragraph will explain, the AFS estimate is fairly Immune to over-reporting of total farm area.

The weight of a sampled farm contains the factor (S_i / S_t), where S_i is the area of the farm inside the segment and S_t the total area of the farm. Consequently, the more a farmer overestimates the total farm area, the smaller the weight given to the farm becomes, and the reduction in the weight offsets exactly the overestimate of total area. In fact, when the total farm area is estimated, only the areas within the segments are actually used. It is difficult to systematically overestimate the areas within the segments because the interviewer knows the total area of the segment (usually 1,920 acres in the Prairie provinces and 1,480 acres elsewhere) and must make sure that the total of the areas inside the segment is within 200 acres of the actual known total. To ensure that the 200-acre margin did not cause a significant bias, the estimates were recalculated after the areas inside the segments had been adjusted proportionally so that their totals equalled the actual known totals. Except for Prince Edward Island, where the estimate of total farm area decreases by 0.4%, all the adjusted estimates are higher than the unadjusted estimates, with Ontario showing the highest increase (2.6%).

There had not been an agricultural area survey in the Maritime provinces for ten years. Yet the training, although decentralized, was similar across the country in terms of content and time allocated. This could explain why the data collected in the Maritime provinces were often inconsistent. As mentioned above, the interviewers had to make sure that the total of the areas inside the segment was within 200 acres of the actual total. A high number of the Maritime segments did not meet this requirement. In addition, the interviewer had to make sure that the total area of the farm was not smaller than the area inside the segment. Many farms also failed this check. This appears to be due to the fact that in many cases, only the cultivated area was counted as the total area of the farm. This also explains why, in the Maritimes, the area reported to the AFS is less than the Census figure, as shown in Table 3. The relatively low level of agricultural activity in these provinces could have contributed to the lower data quality.

4. ESTIMATES BY CATEGORY OF AGRICULTURAL PRODUCTS

Table 4 shows the estimates with cv's for the number of farms and the total area of farms producing various agricultural products. The areas are the total farm areas; for example, the total area of farms growing potatoes is estimated to be 2,6 million acres. This is not the area where only potatoes are grown.

Table 5 compares the 1993 AFS estimates with the 1991 Census figures. For every product category except potatoes, the AFS found fewer farms than did the Census. This is particularly surprising given that the AFS estimate for the total number of farms is higher than the Census number. The operators are thus reporting less product categories in the AFS than in the Census.

For Table 6, the farms in the AFS were matched with the farms in the Census and the numbers of farms producing each category of products were compared for the group of farms common to both surveys. The table indicates there would be a very large response error, always assuming that the number of farms for each category remained unchanged between June 1991 and May 1993. In the group of farms common to both surveys, the AFS finds about four times fewer farms than the Census that produce eggs or have livestock other than cattle, pigs or sheep. There seems to be a problem distinguishing the categories "other livestock" and "other agricultural products." Lists are given in the Census questionnaire for these two categories, whereas space prevents the inclusion of such lists in the AFS question. Thirty-five per cent of the matched farms classified as having "other agricultural products" in the AFS were classified in the Census as having livestock other than cattle, pigs or sheep. The interviewers or the operators are likely reluctant to consider some animals as livestock. More general explanations for the response error would include interviewers marking only the agricultural products inside the segment rather than all products of the entire farm. The interviewers may have felt it was sufficient to mark only the main products of the farm, or they may have marked only the products they could see without questioning the respondent. For the eggs category, it should be noted that on the Census questionnaire, this is the only category where it is explicitly asked to include the production for home use. Much of the difference in response could be due to a difference in the way products for home use are treated by the two surveys.

5. LIVESTOCK SURVEY ESTIMATES OF NUMBER OF FARMS

The Livestock Survey (LS) was conducted in early summer 1993. It is a multiframe survey. The list frame consists of all farms enumerated in the 1991 Census with the exception of farms on Indian reserves and institutional farms. The list frame is thus quite extensive when compared to those used in agricultural surveys of the United States for example. A more complete description of the Livestock Survey is given in Trépanier and Théberge [5]. The area frame of the AFS is used to identify farms that are not on the list frame either because of Census undercoverage or because they have started operating after the Census was conducted. The first column in Table 7 gives the estimate of the number of farms on the list frame, or more precisely the number farms on the list frame that are still in business at the time the LS was conducted. This estimate is derived from the LS list sample, the sample selected from the list frame. The second column gives the estimate of the number of in-business farms not on the list frame. This estimate comes from the area sample, where sampled farms that are also on the list frame contribute zero. Coverage of the list frame (column 3) is obtained by dividing the estimated number of active farms on the list frame (column 1) by the AFS estimate of the number of farms from Table 1, and coverage of the multiple frame (column 4) is obtained by dividing the LS estimate of the number of farms (the sum of the first two columns) by the AFS estimate of the number of farms.

Even in the Prairie provinces, where the Census enjoys very good coverage (Table 2), the coverage of the list frame is down to 95%, two years after the Census. The overcoverage of the multiple frame is very disappointing. In the Prairies, it is as great or even greater than undercoverage. Overcoverage means that much of the overlap between the area sample and the list frame is going undetected. Part of the problem stems from the imprecision in the way farms are identified. Currently, a farm is identified through its name, in the few cases where there is one. In most cases the names of the farm managers are used. An exception is made when a farm is bought by someone who does not already own a farm; in such a case, to prevent deterioration of the list frames, the new owner replaces the old owner as the identifier of the same statistical unit rather than deathing the old farm and birthing a new one, as should be done if a farm were defined by its management. The current method of dealing with changes of operator is a factor in making detection of overlap between the area sample and the various list frames difficult. Most list frames are so large that it is too costly to maintain them in order to know the current managers of each farm. It is important to bear in mind that the problem arises every time farms are matched and every time a search is made for a farm on the register. Such searches are necessary at present to properly handle changes in farm managers at the estimation stage of every survey.

6. CONCLUSION

The next AFS will be conducted in the spring of 1994, but will not include the Maritime provinces in order to cut costs. Various improvements to the AFS are being considered for 1994. These include clarifying the definition of a farm, how to arrive at the total farm area, use of the phrase "other animals" rather than "other livestock", and instructing the interviewers to read the list of agricultural product categories to the respondent. The procedures to maintain the farm register are being reviewed, although no major changes are expected to be made in time for the 1994 AFS.

At the time the area frame was constructed, the use of telephone interviewing in the Prairie provinces was only a possibility. The data collection methodology should be taken into account in future designs of the area frame. Stratification of segments according to data collection methodology would permit better control of costs, and segments to be done by phone could be much smaller, thus improving the efficiency of the sample by reducing clustering of farms.

If there is response error with the total farm area reported at Census in the provinces of Québec, Ontario and British Columbia, and no such error in the Prairie provinces, perhaps the problem is with the reporting of non cultivated land. The percentage of non cultivated land is much lower in the Prairies than in the rest of the country. Some efforts could be made to improve the reporting of such land at Census. To improve Census coverage, making better use of the previous Census records could be contemplated. It should be noted that many of the AFS sample farms that we could not match to a 1991 Census record could be matched to a 1986 or 1981 Census record.

Future research could try to focus on the characteristics of the farms missed by Census. Are they located in more heavily populated areas or in areas with few farms? Can we measure the importance of the missed farms for variables other than total farm area while avoiding the problems of multiframe estimates? It would also be interesting to know more about the AFS sample farms that should match to the list frame, but do not. For example, what is the impact of multiframe overcoverage on variables other than number of farms? A measure of the problem of false matches is also needed.

REFERENCES

INGRAM, S. and DAVIDSON, G. (1983), "Methods Used in Designing the National Farm Survey," *Proceedings of the Section on Survey Research Methods*, American Statistical Association, pp. 220-225.

JULIEN, C. and MARANDA, F. (1990), "Sample design of the 1988 National Farm Survey," *Survey Methodology* 16, pp. 117-129.

BÉLANGER, Y. (1992), "Étude sur l'utilisation d'une technique de numérisation pour l'Enquête nationale sur les fermes," technical document, Statistics Canada.

BUSH, J. and HOUSE, C. (1993), "The Area Frame: A Sampling Base for Establishment Surveys," *Proceedings of the International Conference on Establishment Surveys*, American Statistical Association, pp. 335-344.

TRÉPANIER, J. and THÉBERGE, A. (1993), "Redesign of the Canadian Agriculture Commodity Surveys," *Proceedings of the International Conference on Establishment Surveys*, American Statistical Association, pp. 688-692.

TABLE 1. Area Farm Survey Estimates of Number of Farms and Total Farm Area by Province

Province	Number of farms (in thousands)	CV (%)	Total farm area (millions of acres)	CV (%)
Prince Edward Island	2.7	14.5	0.6	10.7
Nova Scotia	4.6	16.8	0.8	16.6
New Brunswick	5.0	14.7	1.2	13.8
Québec	40.7	4.6	9.4	4.6
Ontario	78.9	3.9	14.8	3.2
Manitoba	26.3	3.8	19.2	1.9
Saskatchewan	62.2	2.1	65.8	0.8
Alberta	58.3	2.9	51.1	1.1
British Columbia	31.2	14.2	5.7	7.2
Total	310.0	2.0	168.7	0.7

TABLE 2. Ratios of Census Figures to Area Farm Survey Estimates by Province

Province	Ratio of number of farms	Ratio of total farm areas
Prince Edward Island	0.87	1.04
Nova Scotia	0.86	1.23
New Brunswick	0.65	0.78
Québec	0.93	0.90
Ontario	0.87	0.91
Manitoba	0.98	1.00
Saskatchewan	0.98	1.01
Alberta	0.98	1.01
British Columbia	0.62	1.03
Total	0.90	0.99

TABLE 3. Ratios by Province of Census Figures to Area Farm Survey Estimates Among Matched Farms

Province	Ratio of total farm areas
Prince Edward Island	1.08
Nova Scotia	1.09
New Brunswick	1.07
Québec	0.94
Ontario	0.95
Manitoba	1.01
Saskatchewan	0.98
Alberta	1.01
British Columbia	0.96
Total	0.99

TABLE 4. Area Farm Survey Estimates by Product Category

Product category	Number of farms (in thousands)	CV (%)	Total farm area (millions of acres)	CV (%)
Greenhouses and nurseries	6.0	12.1	0.5	17.8
Potatoes	5.3	11.4	2.6	12.5
Fruits and vegetables	19.4	11.0	3.8	9.1
Other crops	218.0	1.7	140.0	1.0
Eggs	6.6	11.8	2.6	13.0
Cattle, pigs and sheep	156.0	2.4	95.7	1.5
Other livestock	37.8	8.7	11.8	5.9

TABLE 5. Ratio of Census Figures to Area Farms Survey Estimates by Product Category

Product category	Ratio of number of farms	Ratio of total farm areas
Greenhouses and nurseries	1.27	1.30
Potatoes	0.86	0.89
Fruits and vegetables	1.06	1.00
Other crops	1.07	1.12
Eggs	3.37	4.78
Cattle, pigs and sheep	1.03	1.13
Other livestock	2.54	5.36

TABLE 6. Ratio by Product Category of Census Figures to Area Farm Survey Estimates Among Matched Farms

Product category	Ratio of number of farms
Greenhouses and nurseries	1.47
Potatoes	0.99
Fruits and vegetables	1.20
Other crops	1.13
Eggs	4.17
Cattle, pigs and sheep	1.16
Other livestock	3.88

TABLE 7. Livestock Survey Estimates of Number of Farms and Ratios to Area Survey Estimates

Province	Number of farms (in thousands) On the list frame	Not on the list frame	Undercoverage of the list frame	Overcoverage of the multiple frame
Prince Edward Island	2.3	0.4	0.85	1.01
Nova Scotia	3.8	2.2	0.82	1.29
New Brunswick	3.0	2.2	0.59	1.03
Québec	36.2	6.8	0.89	1.05
Ontario	66.4	18.1	0.84	1.07
Manitoba	24.8	3.6	0.94	1.08
Saskatchewan	59.3	5.7	0.95	1.05
Alberta	55.3	9.6	0.95	1.11
British Columbia	18.5	13.7	0.59	1.03
Total	269.5	62.2	0.87	1.07

STATISTICS CANADA LIBRARY
BIBLIOTHEQUE STATISTIQUE CANADA
1010225800

c.2