REDESIGN OF THE CANADIAN AGRICULTURE COMMODITY SURVEYS

J. Trépanier and A. Théberge, Statistics Canada
J. Trépanier, Statistics Canada, Tunney's Pasture, Ottawa (Ontario), Canada, K1A 0T6

1. INTRODUCTION

From 1983 to 1992, the Canadian National Farm Survey (NFS) has produced estimates on cropland areas, livestock and farmers' expenses and income. The NFS was the second most important agricultural survey after the quinquennial Census of Agriculture. It was a multiframe survey and was conducted in all provinces of Canada except Newfoundland. The major agricultural surveys have traditionally been redesigned following each Census of Agriculture which provides information on a multitude of farms' agricultural activities. Consequently, the last Census of Agriculture that took place in 1991 gave rise to a complete redesign of the NFS based on the 1991 Census data. An evaluation of the Agriculture Statistics Program and an analysis of its probable evolution over the next few years suggested that the NFS be replaced by three separate surveys. Starting in 1993, surveys covering respectively crops, livestock and farm financial data will be conducted and supplemented by a common area sample collected by the Area Farm Survey (AFS).

Since 1988, the NFS design relied on one list frame and list sample for the Maritime provinces (Prince Edward Island, Nova Scotia and New Brunswick), Québec, Ontario and British Columbia (except the Peace River district). Two list frames, each with their own sample, were used for the Canadian Wheat Board (CWB) area, that is the Prairie provinces (Manitoba, Saskatchewan and Alberta) and the Peace River district in British Columbia (Julien and Maranda, 1990). Note that the Peace River district is agriculturally similar to the Prairie provinces. Both NFS list frames included 1986 Census farms except those on Indian reserves and institutional farms. Small farms in terms of cropland areas were also excluded in the CWB area. The list frames were complemented by an area frame in the CWB area, Ontario and Québec. In the CWB area, income and expense estimates made use of only one of the two list frames, the one with the largest farms, and the area frame.

Under the new design, the crop and livestock surveys are both multiple frame surveys with a list and area frames. They share the same frames. The new list frame includes 1991 Census farms of all provinces except Newfoundland. Only farms on Indian reserves and institutional farms are excluded. These exclusions represent only 0.3% of the total number of farms in the target population. The crop component of the NFS which dealt with cultivated areas is now integrated into an already existing group of crop surveys currently dealing with seeding intentions, yields and stocks of grain, giving a new series of six crop surveys conducted throughout the year. The crop samples are based on univariate stratification method. A master sample is selected and a predetermined subsample of it is taken for each crop survey. On the other hand, the livestock surveys will be conducted in July and January. A sample is first selected for the July survey and then subsampled for the January survey. Multivariate stratification methods are used. Even though the strata are the same, sample allocation is performed separately for the July and January surveys. Methods have been implemented to reduce the overlap between the crop and livestock survey samples in order to control respondent burden.

As in previous designs, an area frame survey is used to account for new operations and those missed in the Census. The new AFS uses a one stage design with stratified random sampling. This new method requires much of Canada's land to be segmented using geographical information systems.

2. CROP SURVEYS

Under the previous design, estimates of cropland areas were produced from NFS data. However, estimates of yields, stocks of grain and seeding intentions were provided by a series of seven crop surveys conducted over the course of the year. For more information on these surveys see Bélanger (1990). It was decided to combine the crop part of NFS and this series of seven crop surveys. The new series of crop surveys consists of six surveys: December (stocks of grain), March (seeding intentions and stocks of grain), June (crop areas), July (stocks of grain), September (yields) and November (yields).

The design of the previous series inspired the sampling design of the new crop surveys. Even if six surveys are conducted, the list frame described above was stratified only once. A sample, called the master

sample, was selected and randomly partitioned into a predetermined number of subsamples. Each crop survey's sample is a union of one or more of these subsamples. This method makes it possible to control the overlap between each survey as we will see in more detail in what follows.

2.1. Stratification

For the crop surveys, the first level of stratification is the province. To facilitate the work, the Peace River district and the rest of British Columbia are treated as two distinct provinces. Within each province, several farms with specific characteristics are selected with probability one. All other farms within the province are stratified and samples are selected within strata.

Because of their large size, some farms must be in the sample of more than one of the six individual crop surveys. These farms are grouped in non partitioned take-all strata which permit control over the inclusion of them in any of the subsamples. There are three such strata for the crop surveys. Some Census farms on the frame are part of farming enterprises with complex operating arrangements. These are called multiholding farms and special data collection procedures have been put in place for them. For this reason, and because they are farming operations with frequently changing structures, all multiholding farms are grouped in a take-all stratum. Since community pastures are also treated separately at the time of data collection, they are also grouped in a take-all stratum.

There are other farms on the frame that can be very large. These farms are often very different from the majority of the farms. In order to avoid the undesirable situation of stratum jumping due to large farms changing hands, which can lead to very high weights, these farms are also placed in a take-all stratum within each province. Special procedures ensure that they remain in their stratum. These largest crop farms are identified by what is called the Sigma-Gap Rule, and are called the Sigma-Gap farms. Total cropland area, as defined below, is the Sigma-Gap variable. The Sigma-Gap Rule can be summarized as follows. Within a given province, let X_i be the value of the Sigma-Gap variable X for farm i. Let P be the set $[X_i : X_i > 0]$, M be the median of X on P, σ be the standard error of X on P, and $X_{(1)}$, $X_{(2)},\ldots,X_{(N)}$ be the ordered values of P. Now let k be the smallest number, if it exists, where $X_{(k)} > M$ and $X_{(k)} - X_{(k-1)} > \sigma$. All farms with $X_i \geq X_{(k)}$ are identified as the Sigma-Gap farms. For the crop surveys, Sigma-Gap farms represent 0.02% of the number of farms on the list frame. All the remaining farms were stratified using the method described in the next two paragraphs.

In Canada, estimates of crops are calculated at both the provincial and subprovincial levels but published at the provincial level only. These subprovincial levels correspond to agricultural regions within provinces. The number of agricultural regions in a province varies from one to twenty. Provincial authorities are increasingly more interested in subprovincial estimates, but budget, and consequently the sample size, are often a constraint. Some of the crop surveys have a sample size of approximately 11,000 units for a population of 280,000 units. Agricultural regions could form the second level of stratification within the province for these surveys and thus provide improved subprovincial estimates. A study was conducted before the redesign to analyze this possibility. Stratification where strata are created within the agricultural regions was compared to stratification where strata are created within the province. If the total number of strata to be formed is fixed within a province, stratification within the agricultural regions showed that improvements to the estimates (reduction in coefficients of variation) at the agricultural region levels could not justify the use of the method because of the deterioration of the provincial estimates (increase in coefficients of variation). A compromise was investigated and made. Similar agricultural regions were grouped. The number of subprovincial regions was thus approximately cut in half. These new subprovincial regions became the second level of stratification, even though estimates are still to be computed at the finer regional level. Strata were then formed within each province and subprovincial region. The number of strata created in each subprovincial region varies from 4 to 10 depending on the size of the region.

The new series of crop surveys covers cropland area, seeding intentions, yields and stocks of grain. Census information does not include variables on intentions, yields and stocks. Thus the 1991 Census total cropland area was the only variable used for stratification. Despite the fact that hay is not a variable of major interest by itself, the Census total cropland area includes hay area. Canadian farmers (except in the CWB area) use hay as part of their crop rotation; what is hay acreage one year can easily become grain the year after. Thus hay acreage was kept in total cropland area except in the CWB area where it was excluded. Total cropland, as modified above, is the stratification variable. The boundaries of

the strata were calculated with Sethi's algorithm when Neyman's optimal sample allocation is used (Sethi, 1963).

2.2. Master Sample and its Subsamples

Canada, for the purpose of the crop surveys, is divided into three major areas: Maritime provinces and British Columbia (excluding Peace River district), Québec and Ontario, and CWB area. There are six crop surveys involved in this redesign. Each survey's sample size was fixed for each of the areas defined above as well as the size of the overlap, if any, between the survey samples. For instance, the March survey sample and the July survey sample have no common element, but together they also form the sample for the December survey. Thus, the size of the master sample in each major area was determined. Simultaneously, the number of subsamples needed and their sizes were fixed. Remember that the master sample is partitioned into a number of subsamples. The union of some of these subsamples forms the sample of each survey. The final size of the sample must be close to the target size. The size of each subsample was then allocated to the provinces within the major area. Allocation to the strata that are not take-all was performed using Neyman's optimal sample allocation. Total cropland area (excluding hay acreage in the CWB area) was the allocation variable. The size of the master sample is approximately 98,000 units.

Keeping in mind that good estimates are wanted at the agricultural region level, proportional representation was ensured by sorting the observations of all strata by agricultural region and selecting the units of the master sample using a circular systematic method. Units within an agricultural region were randomly sorted to ensure that samples within regions are the equivalent of simple random samples.

In a given province, the master sample was randomly divided into a certain number of subsamples of equal size. This partition was performed within each stratum. The representativity of the population by each subsample was verified by computing sample statistics for many of the variables of interest. A number of subsamples is assigned to each of the six surveys. If an overlap is desired between two surveys, the same subsamples are used. Otherwise, different subsamples are assigned to minimize respondent burden. The number of subsamples in a province varies from 5 to 18.

3. LIVESTOCK SURVEYS

The NFS livestock data used to be collected annually in July. In January, a subsample of the July sample was contacted for the January Farm Survey (JFS) and used to provide another series of livestock estimates. This redesign retains the same basic idea. See Julien and Maranda (1990) for more details on the NFS design. Starting in 1993, a July Livestock Survey and a January Livestock Survey are being conducted. Once again, these surveys use the list frame described in Section 1.

3.1. Stratification

For the livestock surveys, the province is the first level of stratification. Unlike the crop surveys, the livestock surveys do not treat the Peace River district separately from the rest of British Columbia. Within a province, the multiholding farms were again grouped into a take-all stratum as were the community pastures. The Sigma-Gap Rule was applied to the remainder of the frame for each of the following variables: beef cows, milk cows, sows, total number of cattle, total number of pigs and total number of sheep. If a farm happened to be a Sigma-Gap farm for any one of these variables, it was included in the take-all stratum created for the Sigma-Gap farms. Sigma-Gap farms represent 0.08% of the total number of farms on the list frame.

A significant number of Census farms on the list frame do not have any livestock (in terms of cattle, pigs and sheep). These can represent up to one half of the total number of farms in a province. Because many farms can have no livestock during the season when Census is done, but yet have some during winter when the January Livestock Survey is conducted, these "zero" farms were kept on the frame. Retaining the "zero" farms also ensures that farms that were misclassified at Census or have since started breeding livestock are also covered. The Census farms with no livestock were grouped in one provincial stratum. All the other farms were stratified using the multivariate procedures Fastclus and Cluster of the SAS software (SAS Institute Inc, 1985). The stratification variables were beef cows, milk cows, sows, total number of cattle, total number of pigs and total number of sheep. These were standardized to have zero mean and unit variance before using the clustering algorithms. Empirical investigation showed that standardized stratification variables achieved a better balance of coefficients of variation than unscaled variables.

The number of farms to be stratified was quite high. Due to limited computer resources, the Fastclus procedure was first performed in order to produce an initial clustering of 150 clusters. (The Cluster procedure produces a hierarchical stratification starting its number of strata with the number of observations in the population and ending with one stratum regrouping all population units. It can be very expensive when the number of units in the population is very large.) Then the Cluster procedure was applied to these 150 clusters using Ward's minimum variance method. The final number of clusters varies from 11 to 42 depending on the province.

3.2. Samples

The sample size for the July Livestock Survey is about 30,000 units whereas that of the January survey is approximately 13,000 units. The provincial sample sizes were fixed. Part of the sample size is taken up by the three special take-all strata. For both surveys, a sampling rate of 1/60 was used in the stratum of farms with no livestock. Then, independently for the January and July surveys, allocation of the remaining sample size to the other strata was performed. Both allocations were done to optimize the precision of the estimates of the same livestock variables: beef cows, milk cows, sows, total number of cattle, total number of pigs and total number of sheep.

The 1993 July sample was selected using stratified simple random sampling without replacement. For January, a stratified simple random sample was drawn from the July sample according to its own allocation. In order to reduce respondent burden, in subsequent years, these two samples will be rotated annually at a rate of 50% for the stratum of Census farms with no livestock and 20% for the other strata.

3.3. Overlap Reduction between the Crop Master Sample and the July Livestock Sample

Controlling respondent burden has always been an important issue at Statistics Canada. Each year, a number of agricultural surveys are conducted and consequently the respondent burden can be significant. The crop and livestock surveys are a major part of the agricultural program and thus it was decided to minimize the overlap between the crop master sample and the July livestock sample. A method proposed by Kish and Scott (1971) was chosen and adapted for this purpose. In each intersection of crop and livestock strata, the units in both the crop and livestock samples are identified. As many of these units as possible are removed from the livestock sample and replaced by non sampled units in the intersection. It can be mathematically described as follows:

Let $U_{c.}$ and $U_{.l}$ be respectively the crop stratum and the livestock stratum. Let also $U_{cl} = U_{c.} \cap U_{.l}$ and let

s_{cl} = units of U_{cl} in both the crop and livestock samples
n_{cl} = number of units in s_{cl}
U'_{cl} = units of U_{cl} that are in neither the crop sample nor the livestock sample
N'_{cl} = number of units in U'_{cl}

Two different actions are possible:
(1) If $N'_{cl} \geq n_{cl}$, then n_{cl} units are selected from U'_{cl} by simple random sampling. These selected units replace the units of s_{cl} in the livestock sample.
(2) If $N'_{cl} < n_{cl}$, then N'_{cl} units are selected from s_{cl} by simple random sampling. These selected units are replaced by the units of U'_{cl} in the livestock sample.
In both cases the crop sample remains unchanged.

This method was applied to every possible U_{cl} and resulted in a reduction of approximately 60% in the overlap between the crop master sample and the July livestock sample. After rotation is applied to the livestock sample, reduction of the overlap is performed using the same method with U'_{cl} = units of U_{cl} that are in neither the crop sample nor the livestock previous and current samples.

4. AREA FARM SURVEY

The purpose of the Area Farm Survey (AFS) is to complement the list frames of various agricultural surveys, in particular the common list frames of the livestock surveys and the crop surveys. Farms that are not on this list frame include those that were missed by the 1991 Census and new farms that started operating after the Census. It will also be possible to produce from the AFS, estimates of number of farms and of total farm area.

4.1. Frame

Under the previous design, the area sample, which was a component of the NFS, was selected in two stages. The first stage was the selection of Enumeration Areas (EA) which correspond to the area that was canvassed by a Census enumerator. The selected EA's were then manually partitioned into land segments from which a sample was drawn.

Technological advances have now permitted to segment the whole country automatically, and therefore to select the sampled segments in one stage. The AFS, just as the list frame, covers all provinces except Newfoundland whereas the previous design covered only Québec, Ontario, and the CWB area.

In the Prairie provinces and in part of the Peace River district, 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 cells are then intersected with the EA's. EA's with no farm headquarters cover very large areas and have little agricultural activity, therefore cells that do not intersect EA's with farm headquarters were discarded automatically. Other cells were discarded after being examined using a topographical map if, for example, they were completely inside a national park. At the same time, cells were combined when, for example a large part of them was water. Each remaining cell, or group of cells if combining was done, corresponds to one of the segments that make up the frame.

4.2. Stratification

The same subprovincial regions as those used for stratification in the crop surveys form the second level of stratification after the province. Statistics for the segments on the frame were arrived at by allocating EA Census totals to the intersecting segments in proportion to the area of intersection. 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 and used as input into the Fastclus procedure of SAS to form strata within the subprovincial regions. The number of strata was such that we had on average a sample of 25 segments per stratum.

4.3. Sampling and Estimation

A sample of 2,100 segments was allocated to provinces roughly in proportion to size (total number of segments in the province) while taking into account the sample sizes under the previous design and respecting a minimum size of sampled segments in each province. Sample allocation to the strata was proportional to the square root of the stratum size (total number of segments in the stratum). Each stratum's segments were then sorted randomly within Census Subdivision (usually municipalities) before systematic random sampling was employed to ensure a good coverage of the land. A rotation rate of 25% is planned for sampled segments in each subsequent year.

All farms with land in a selected segment are enumerated. These farms are unduplicated against a given survey's list frame and non listed farms are surveyed as part of the data collection activities of that given survey. The fraction of the farm that lies inside the segment is used to arrive at segment totals.

5. CONCLUSION

The first crop and livestock surveys using the new list samples were respectively the 1992 December Farm Survey and the 1993 January Livestock Survey. Furthermore, the Area Farm Survey was first conducted in April and May 1993. Up to now, few results on the evaluation of this redesign are available, but will be produced shortly.

REFERENCES

Bélanger, Y. (1990), "Redesign of the Crop Surveys at Statistics Canada", Statistics Canada, Methodology Branch Working Paper, No. BSMD-90-014E/F.

Julien, C. and Maranda, F. (1990), "Sample Design of the 1988 National Farm Survey", *Survey Methodology*, 16, 117-129.

Kish, L. and Scott, A. (1971), "Retaining Units after Changing Strata and Probabilities", *Journal of the American Statistical Association*, 66, 461-470.

SAS Institute Inc. (1985), *SAS User's Guide: Statistics*, Version 5, Cary, North Carolina: SAS Institute Inc.

Sethi, V. K. (1963), "A Note on Optimum Stratification of Populations for Estimating the Population Means". *Australian Journal of Statistics*, 5, 20-33.