



Potential uses of tax data in the Canadian census of agriculture

P. Hunsberger Statistics Canada | Ottawa | Canada

L. O'Neill Statistics Canada | Ottawa | Canada

DOI: 10.1481/icasVII.2016.pl2

ABSTRACT

As in many statistical organizations, Statistics Canada is examining ways to reduce the amount of survey response burden placed on its respondents while maintaining a high level of quality for its statistics. Statistics Canada's Agriculture Statistics Program has placed a high importance in recent years on exploring the expanded and improved use of farming revenue and expense data reported by farm operators to Canada's tax agency when filing their tax returns.

The Canadian Census of Agriculture is no exception to this initiative. In addition to questions on agricultural characteristics such as livestock inventories, crop acreages and agricultural practices, the Canadian Census of Agriculture has traditionally collected information from farm operators related to revenue and detailed expenses. The possibility of removing these detailed, high response- burden financial questions from the Census and replacing them with income tax data was explored following the 2011 census occasion as part of a feasibility study. The study concluded that such a replacement would work well for the vast majority of farms, but there were certain sub-populations in which it would not perform well.

Agricultural tax data can also be used in the processing of other variables collected by the Census of Agriculture. Tax data will be used in the data validation stage of the 2016 Census of Agriculture, to corroborate the responses provided to census questions such as agricultural commodities. They will also be used to improve the nearest neighbour donor imputation of total non-response.

A further use of tax data has been to help define the frame of farm operations to be used for the mail out of the Census. Detailed agricultural tax data are used to add and remove units from the agricultural population.

The paper will describe the studies undertaken after the 2011 Census of Agriculture to assess the feasibility of using tax data for the replacement of respondent-provided information and frame generation. It will also describe the new agricultural tax data environment developed at Statistics Canada which can be used by both the Census of Agriculture and other agricultural programs, including the Agriculture Taxation Data Program. Finally it will discuss the manner in which tax data will be used for the Census of Agriculture in 2016.

Keywords: Data replacement, agricultural frame, data validation, feasibility study

PAPER

1. Introduction

Statistics Canada's Agriculture Statistics Program has a long history of using administrative data. These data are obtained from other federal and provincial government departments as well as non-government organizations, and are incorporated into the Agriculture Taxation Data Program (ATDP), the quinquennial Canadian Census of Agriculture (CEAG), the Remote Sensing and Geospatial Analysis project and ongoing and special agricultural surveys. This paper will discuss recent initiatives undertaken by Statistics Canada to increase the use of agricultural tax data obtained from Canada's tax agency, the Canada Revenue Agency, in the construction of the agricultural survey frame, the redesign of the agricultural tax data environment, the processing of the 2016 Census of Agriculture and the potential replacement of existing variables in a future Census of Agriculture. Adding to the challenge of achieving these goals is the fact that multiple agricultural projects have been historically using tax data in different ways, and the new standardized processing and frame methodology must now meet their various needs at the same time.

2. Building the Agricultural Survey Frame

Prior to 2012, Statistics Canada maintained a list of agricultural operations solely for the purposes of the Agriculture Statistics Program called the Farm Register (FR). The CEAG and agricultural surveys

derived their survey frames from the FR, which included baseline information necessary for survey design. The FR in turn was largely updated in terms of additions and deletions of operations and contact information primarily by conducting the CEAG once every five years, and to a lesser extent throughout each year based on feedback from agricultural surveys.

That changed in 2012 when Statistics Canada made the decision to use the Business Register (BR) to maintain its frame of agricultural operations for its Agriculture Statistics Program, similar to most other business survey programs at the Agency, in an effort to reduce redundancy. The BR is a complete list of Canadian businesses primarily maintained by reported tax data and other associated administrative signals, such as indications of cessation of business activity, structural changes, and shifts in primary business activity type. A key field on the BR is the Business Number, a nine-digit code that facilitates interaction with some government departments and that has been incorporated into the CEAG. The post-2011 CEAG reconciliation of the FR with the BR involved major challenges; for more information see Dongmo Jiongo et al (2013). Tax data are typically received for existing farm businesses every year and in some cases every month. The direct connection between the BR and tax data allows the BR, and hence the agriculture survey frame, to be updated more fully and more frequently.

A key annual use of tax data is in the selection of which businesses should be added into (birthed), or removed from (deathed), the agricultural survey frame. Initially, the BR identifies a unit as a birth or a death based on the appearance or long-term disappearance of tax data. The agricultural survey frame, as a subset of units on the BR, will then reflect these additions or removals. New units added to the BR are included in the agricultural survey frame based on tax indications of agricultural activity. However, it is not prudent to include all new businesses that have reported agricultural activity. Filtering has proved necessary for a variety of reasons. For example, some BR units that file agricultural tax information are out of scope to the Agriculture Statistics Program. Out of scope agricultural tax units may include landlords who rent out farmland to others, but do not actively operate a farm themselves, yet still identify themselves when filing taxes as being an operator of a business of an agricultural nature and with agriculture-related revenue. Conversely, there are other businesses that report agricultural tax activity that are suspected, but not yet confirmed, to be in a partnership of some sort with an existing unit on the agricultural survey frame. Their automatic addition would generate duplication. A filter is therefore applied to only incorporate tax births that exhibit significant, unique, detailed and commodity-specific agricultural revenue and/or expenses to the agricultural survey frame.

Finally, tax information, in the form of commodity-specific agricultural revenue and expense information supplied by the tax filer, is used to derive measures of commodity-specific agricultural production that allow agricultural surveys to better identify their survey populations and develop efficient survey designs. Currently these production measures are generated for previously unknown (birthed) frame units on an annual basis, and so overall the agricultural survey frame now has a better, more recent picture of the agricultural activities of Canadian farm operations during the inter-censal period.

3. The Redesigned Agricultural Tax Data Environment

For over 25 years the ATDP has tabulated statistics based on agricultural revenue and expense information reported on farm operators' Canada Revenue Agency tax returns. Prior to 2016, the ATDP received a sample of tax records from the Canada Revenue Agency that contained some indication of agricultural revenue. The ATDP processed the tax fields obtained from this sample using custom rules and methodology. Starting in 2016, the ATDP will be defining their population of interest using the BR in a similar manner to that used by the CEAG and agricultural survey programs. Tax data will be obtained for all units resulting in a census of in-scope agricultural tax filers rather than a sample. This transition faces some challenges. Statistics Canada has a system in place for the processing of tax data for businesses in all industries including those outside of the agriculture sector, but this system does not include some specialized rules and variables necessary for both the ATDP and other uses within the agriculture program. It was therefore necessary to build a new hybrid system, consisting of the general Statistics Canada tax processing rules plus some extra agriculture-specific requirements. Furthermore, the standardized tax data processing does not uniformly produce the highly specific commodity-level revenue and expense details required by ATDP, and so a new imputation system for agricultural tax data was implemented.

This updating of the agricultural tax data processing environment was undertaken collaboratively, with ATDP contributing their tax data processing techniques and CEAG their agricultural frame expertise. The ultimate goal of these efforts is to produce one high-quality Agricultural Tax (Ag-Tax) database that will be capable of supplying all agricultural statistics programs in Statistics Canada with highly detailed, processed financial information on a sub-annual basis. The Ag-Tax database resulting from this initiative will be made available to a variety of agricultural statistics programs, including the CEAG, for the first time in 2016 for the 2015 tax year.

With both the ATDP and the CEAG now defining their populations based on the BR and using the same set of processed agricultural tax data, the potential benefits are great. ATDP can now produce financial statistics based on a census of agricultural operations rather than a sample. The resulting Ag-Tax

database from the redesigned tax system reduces the need for the CEAG to collect this data from respondents directly, and makes it feasible to directly replace detailed expenses in a future Census or other agricultural surveys which require financial information. In addition to the reduction in survey response burden, it results in more coherence between the financial statistics from different products within the Agriculture Statistics Program.

4. Use of Agricultural Tax data in the 2016 Census of Agriculture

Drawing directly from the newly created Ag-Tax database described above, the 2016 CEAG will for the first time have a high-quality comparative dataset of agricultural tax revenues and expenses available at a commodity-specific level of detail. The most extensive use of this database will be made by CEAG data validation, although it will also be applied to total non-response imputation.

Ag-Tax data, as well as other sources of administrative data, will play an important role in the validation of CEAG data. They will be used to confront and validate both the data provided by respondents, as well as data which has been imputed. For tax data in particular, CEAG will use total farm revenue and total operating expenses tax fields to validate the gross farm receipts and total operating expenses questions on the CEAG questionnaire (Statistics Canada, 2015). CEAG will also make use of other tax variables in the Ag-Tax database to improve the data quality in other sections of the CEAG questionnaire. For example, the detailed revenue Ag-Tax variables will be used to validate the presence of corresponding commodities as reported by the respondent on the Census questionnaire. CEAG validators will also use tax data to validate the value of land and buildings and machinery values provided by respondents on their census form. Another important use of tax data in validation will be to ensure that the values reported by the respondent do not include components that are considered to be out of scope for the Census such as capital sales and goods purchased only for resale. Goods purchased for resale can be a difficult concept to identify in the gross farm receipts value provided by the CEAG respondent, so CEAG validators employ ratios using detailed expense items and other relevant questions on the questionnaire. For example, seed purchases in relation to greenhouse area under glass to remove resales in greenhouse operations; and livestock purchases in relation to feed purchases to eliminate sales from livestock dealing activities. Finally, using detailed revenues from tax data will also allow analysts to identify non- agricultural receipts, such as revenue from trucking, rental income or oil leases.

Another application of the Ag-Tax database in the 2016 CEAG will be to use variables from this database as matching variables to assist in the identification of nearest neighbours in the CEAG total non-response imputation stage. Ag-Tax variables used in this manner are total revenue, total expenses and a field describing the main agricultural activity of the farm. As has traditionally been the case, the CEAG uses a nearest-neighbour hot-deck imputation system, imputing missing information from a valid CEAG donor record using Statistics Canada's BANFF generalized imputation system (Statistics Canada, 2014). It is expected that the application of tax data here will lead to improved total non-response imputation results, with recent Ag-Tax data now being used instead of five-year old CEAG data.

Finally, tax data will be used to maximize coverage of CEAG farms. While the initial 2016 CEAG frame is derived from the agricultural survey frame, and hence the Business Register, as close to Census day as possible, there is still at that time a volume of new tax units that will not be added to the BR until later in the year. Since the goal of the CEAG is to achieve 100% coverage of all Canadian farms as of Census day, an attempt is made to match this set of new tax units to the Canadian Census of Population, which is carried out concurrently with the CEAG and contains a question asking if anyone in the household is a farm operator. At defined points throughout census collection, a list of Census of Population households with an agricultural operator is generated.

Households from the list that link successfully to a tax unit on the waiting list that has reported agricultural tax data are eligible to be sent a CEAG questionnaire.

5. Potential for Detailed Financial Data Replacement in a Future Census of Agriculture

The 2011 CEAG questionnaire (Statistics Canada, 2010) featured two questions on financial totals (Total Sales and Total Expenses) and seventeen questions on detailed financial expenses, such as fertilizer expenses, livestock purchases, wage expenses and machinery rental. In many (but not all) cases these census questionnaire variables sought the same information as that which was already available on the agricultural tax forms submitted to the Canada Revenue Agency. With the goal of assessing whether a direct replacement of CEAG questionnaire variables with tax data from the proposed Ag-Tax program would be feasible in the future, a comparative analysis was undertaken in conjunction with the 2011 CEAG collection period.

The 2011 CEAG Tax Data Replacement Initiative was a strategic investment of Statistics Canada and was part of the 2011 CEAG program. The tax data replacement feasibility study (TAX) (Statistics Canada, 2013) replaced 2011 CEAG financial data with corresponding tax data from the Canada Revenue Agency.

The purpose of the study was to evaluate the robustness of tax data as a replacement for detailed operating expenses provided by respondents. It has been estimated that the use of administrative data to replace all of the detailed CEAG financial questions would reduce the response burden for CEAG respondents by as much as 13.5% in terms of time spent completing the questionnaire. In addition, their removal would address requests to do so by respondents in previous censuses. Tax data were processed through the existing 2011 CEAG production edit and imputation system to produce a set of financial data for all of the approximately 200,000 Census in- scope records. In this manner CEAG-TAX financial data were fully edited, imputed and consequently validated using the same processing methodology as in 2011 CEAG production, allowing for a fair comparison between 2011 CEAG and CEAG-TAX replacement data.

The results of this comparison, showing the percent relative difference between TAX and CEAG financial data, are shown below in Table 1. Results are grouped into "Standard" and "Non- Standard" farm types. "Non-Standard" here defines the less than 1% of Canadian farms that are either extremely large and/or complex in terms of production and therefore have a critically high impact on the Agricultural sector in Canada, or that are highly unique in terms of their reporting arrangements.

Table 1: Percent Relative Difference between TAX and CEAG: Gross Farm Receipts and Total Operating Expenses, by Farm Type

CANADA	FARM TYPE	% RELATIVE DIFFERENCE (TAX vs. CEAG)
Gross Farm Receipts	All	-2.00%
Gross Farm Receipts	Standard	-0.70%
Gross Farm Receipts	Non-Standard	-8.65%
CANADA	FARM TYPE	
Total Expenses	All	0.60%
Total Expenses	Standard	2.60%
Total Expenses	Non-Standard	-8.63%

Overall, the comparison for all farms between TAX and CEAG for gross farm receipts and total operating expenses at the Canada level was good. However, there were significant differences in the comparability between TAX and CEAG for "Non-standard" CEAG farms. This analysis was also extended to detailed expense items, and a similar pattern of results was found, with most

detailed expense categories comparing well for "Standard" farms and the largest differences occurring for "Non-standard" CEAG farms.

The study arrived at several key conclusions. For "Standard" farms, which make up 99% of the farm population in Canada, the total replacement of CEAG detailed expense items was deemed feasible by the study; however, it was considered essential to keep gross farm receipts, value of forest products sold and total operating expenses fields on future CEAG questionnaires. Gross farm receipts and total operating expenses are considered critical fields for reconciliation between CEAG questionnaires and tax data, and also for imputation for the small number of CEAG units for which linkage to tax data is not possible. Value of forest products sold is not obtainable from tax data with any degree of quality and so it should continue to be asked on the questionnaire.

While it was deemed feasible to use tax data for total wages and salary expenses, it was recommended not to provide the split between family and non-family wages directly through tax replacement. It was recommended that future CEAG questionnaires collect the proportion of family or non-family wages and salaries.

The study concluded that due to their complex structures and importance to the agriculture sector, extremely large agricultural operations should continue to provide their detailed finances via the CEAG questionnaire rather than through tax data. Similarly, due to their unique nature and the poor reconciliation between CEAG questionnaires and tax forms, it was recommended to continue to require direct collection of detailed finances for community pastures, institutional farms and northern farms. To allow for the replacement of financial data with tax data for collective farm operations, it was recommended that research be undertaken to better understand their organization and reporting tendencies. It is worth noting, however, that the more recently developed harmonized Ag-Tax processing environment may provide significant and necessary methodological improvements regarding the processing of raw tax data for the above-mentioned "Non-standard" farms. If this proves to be the case, this recommendation would need to be revisited.

In general, the feasibility study showed shifts in financial variables, with some shifts more unique and impactful than others. Having used the same processing methodology to treat tax data and CEAG financial data alike, the feasibility study recognized that some of these shifts may be due to systematic differences in the way that certain financial information is reported to the Canada Revenue Agency as

compared to CEAG. Consequently it was recommended that CEAG should consult with agricultural accountants and other knowledgeable parties to understand reporting tendencies relating to tax filing and to help provide explanations for these data shifts.

Overall, the results of the feasibility study were encouraging. They align well with Statistics Canada's commitment to increase the use of administrative data to reduce the burden on survey respondents (Statistics Canada, 2016) and the recognition that harmonization of the processing of raw agricultural tax data within the agricultural statistics program would reduce duplication of efforts and also improve the overall quality of financial data. However, while the Ag-Tax database was successfully developed as described above in Section 3, fiscal constraints resulted in the 2016 CEAG cycle being unable to implement the replacement of detailed expenses with tax data. Plans are ongoing to consider having the 2021 CEAG ask only total agricultural sales, total expenses and value of forestry product information as in the 2016 CEAG questionnaire, and use the Ag-Tax database of processed micro-level tax data to fully replace the detailed financial expense questions. At this time, a final decision has not been made. If the decision is made to proceed with tax data replacement in 2021, these financial expense fields would be processed and validated alongside other 2021 CEAG questionnaire fields such as questions on commodity, machinery and land management practices to ensure consistency, and released as part of 2021 CEAG dissemination.

CONCLUSION

Statistics Canada has successfully undertaken the major recent initiative to adopt the Business Register as its agricultural survey frame, which directly enabled linkage to administrative data including tax data. The CEAG and the ATDP are therefore now both running off a similar frame, and the ATDP will for the first time produce a census file of tax records, as opposed to a sample.

The resulting availability of up-to-date, processed tax and administrative data will support the ongoing maintenance of the agricultural survey frame. It will also make it easier to use tax data for the validation and imputation of the 2016 CEAG. Building on the results of a feasibility study, continuing investigations are being made to consider using tax data to replace detailed financial questions on future censuses and agricultural surveys. The resulting efficiencies gained in population coverage and contact success rates as well as the reduction in response burden on Canadian farm operators with regard to the reduced direct collection of financial information should be significant.

REFERENCES

Statistics Canada (2016) Corporate Business Plan, 2015-16 to 2016-17, available on: <http://www.statcan.gc.ca/about-apercu/bp-pe-eng.htm>

Statistics Canada (2015) Questionnaire for the Canadian Census of Agriculture (2016), available on: http://www23.statcan.gc.ca/imdb-bmdi/instrument/3438_Q1_V4-eng.htm

Statistics Canada (2014) Functional Description of the BANFF System for Edit and Imputation, internal document

Dongmo Jiongo J., Émond N. and Lynch J. (2013) Migrating Agricultural Surveys to Statistics Canada's Business Register: Many Challenges, Many Successes and Much Work Ahead, 2013 International Methodology Symposium

Statistics Canada (2013) Census of Agriculture Tax Replacement Project: Feasibility Study, internal document

Statistics Canada (2010) Questionnaire for the Canadian Census of Agriculture (2011), available on: <http://www.statcan.gc.ca/eng/ca2011/110008>