The Canadian Longitudinal Census of Agricultural File – a Tool to Better Understand Structural Change of Canadian Farms

Katrin Nagelschmitz[1]*, Jean-Francois Frenette[1], Mark Brown[2], Sandrine Prasil[2]

[1] Agriculture and Agri-Food Canada 1341 Baseline Road
Ottawa, Canada

[2] Statistics Canada
170 Tunney's Pasture Driveway
Ottawa, Canada

Katrin.Nagelschmitz@agr.gc.ca Jean-Francois.Frenette@agr.gc.ca

Mark.Brown@canada.ca
Sandrine.Prasil@canada.ca

DOI: 10.1481/icasVII.2016.g45

ABSTRACT

The structure of the agricultural sector evolves through the responses of individual farms to technological innovation, economic shifts and demographic trends, among other factors. In 2014, the Canadian Longitudinal Census of Agriculture (CL-CEAG) dataset was created to provide an additional tool for micro-level farm analysis. This initiative followed the example of the U.S. Census of Agriculture Longitudinal File. This paper describes the CL-CEAG dataset, the methodology for the record linkage and models of farm exits and farm profitability as examples of analysis using the CL-CEAG. Discussion on the features of the file and on how to access it concludes.

Keywords: Census of Agriculture, longitudinal data, farm exit, farm profitability

* Corresponding Author

1. Introduction[1]

[1] The authors thank their colleagues for valuable comments and suggestions, in particular Ray Bollman.

The primary purpose of the Census of Agriculture (CEAG) is to create a statistical profile of Canadian census-farms[2] at one moment in time. However, given that the CEAG includes the population of census-farms and that it provides a broad range of farm and operator characteristics with a high degree of continuity over time, the CEAG presents a unique opportunity for the creation of a longitudinal dataset for farm-level analysis of structural trends and the lagged impact of policies and external shocks. Thus, following the US example, where the US Census of Agriculture Longitudinal file was created in 1997, linking census records back to 1978[3], Statistics Canada created in 2014 the Canadian Longitudinal Census of Agriculture (CL-CEAG) dataset. The purpose of the paper is to present the CL-CEAG dataset as a new analytical tool, describing in particular the methodology for the record linkage, as well as providing models of farm exits and farm profitability as examples of analysis using the CL-CEAG. Discussion on how the dataset can be accessed while ensuring data confidentiality concludes.

2. Existing longitudinal files for the Canadian agricultural sector

The CL-CEAG micro-dataset complements already existing longitudinal data on Canadian farms from administrative data sources and surveys. Housed at Agriculture and Agri-Food Canada (AAFC), administrative data of from AgriStability and AgriInvest, two Canadian agricultural risk management programs, provide detailed longitudinal farm business revenue and expense data of over 100,000 participating farms, as well as detailed inventory information for about half of these, beginning in 2005. The population is limited to program participants and does not include all Canadian farms. The file does not include farm physical information or farm operator characteristics.

The Longitudinal Administrative Databank (LAD), housed at Statistics Canada, consists of detailed income, social transfer and demographic information of 20 percent of Canadian income tax filers and their families, beginning in 1982. Farm families can be identified as those with a non-zero value for Gross Farming Income and/or Net Farming Income. The LAD has been used to analyze the persistence of negative net farming income and the prevalence of low family income of farm families compared to non-farm families (Culver, 2012; Nagelschmitz, 2007). The file does not include any information on the farm business.

The sample overlap of the biennial Farm Financial Survey, which collects information on farm assets, liabilities, capital purchases, and capital sales, as well as some operator demographic information and program participation information, has been used for analysis, such as of the year-to-year changes in level of farm debt (Culver, 2012).

[2] In Canada, a census-farm - or "agricultural holding" - is any operation producing agricultural products that are intended for sale, which in addition to farms producing field and livestock products crops also includes mushroom houses and nurseries; farms producing Christmas trees, fur, game, sod, maple syrup or fruit and berries; beekeeping and poultry hatchery operations; operations with alternative livestock; as well as operations involved in boarding horses, riding stables and stables for housing and/or training horses (see http://www.statcan.gc.ca/eng/ca2011/gloss) (See http://www.statcan.gc.ca/pub/95-629-x/2007000/4123857-eng.htm). In this paper the focus is on all census-farms, and we use the term "farm" and "census-farm" interchangeably, always mindful that the majority of census-farms are small operations , which provide at most only a very small contribution to their farm operator's income.

[3] See Hoppe and Korb (2006) for a description of the US Census of Agriculture Longitudinal file dataset; Ahearn et al (2009) and Katchova and Ahearn (2015) use the file for analysis of farm level dynamics.

In the past, CEAG files have been linked at the micro level in pairs of two succeeding census iterations to analyze the characteristics of exiting, entering and continuing census-farms (Kapitany and Bollman, 1983; Ehrensaft et al, 1984; Bollman and Ehrensaft, 1988; Bollman et al. 1994; Kimhi and Bollman, 1999). In addition, Shapiro et al. (1987) used as sample of micro-linked census-farm records from 1966 to 1981 to examine the dynamics of farm concentration.

The CL-CEAG complements these existing longitudinal administrative and survey data sources, by providing longitudinal data for all census-farms, as well as the possibility of including a wide range of farm and operator characteristics variables in micro-level models examining farm behaviour for up to 25 years.

3. The Canadian Longitudinal Census of Agriculture dataset

The Canadian Longitudinal Census of Agriculture (CL-CEAG) dataset links the micro-data of census-farms across the six censuses from 1986 to 2011, using the unique Agricultural Operation Identifier (AGOPID), which is part of each census-farm record. The AGOPID makes it possible to track the change in characteristics of individual farms over time, as well as to identify census-farm entrants and exits between collection years to measure the dynamics of structural change. The dataset provides a consistent set of industry- and geography-based classifying variables.[4] For the latter, the 2011 Standard Geographic Classification is applied, with the fundamental geographic unit being Census Consolidated Subdivisions (CCSs). CCSs can be aggregated, to form Census Divisions, Census Agricultural Regions, and Economic Regions, providing a high degree of flexibility.[5] The CL-CEAG dataset currently includes a number of variables that describe farm and farm operator characteristics that are consistent over the period covered by the data. Many more census variables can be added as research projects are developed.[6]

The AGOPID is fundamental to the creation of the CL-CEAG dataset and the interpretation of analysis that uses the data. The AGOPID is attached to each census-farm, which has been identified based on the definition of "agricultural holding" (see footnote above) and is therefore included in the Census of Agriculture. The AGOPID is largely based on the location of the farm's headquarter. A census-farm is treated as "continuing" if it responds to censuses under the same AGOPID across census years. By the same token, if a respondent associated with an AGOPID indicates that there is no longer an operation that produces agricultural products for sale or with the intent to sale, the census-farm with that AGOPID is considered to have exited[7].

Changes in farm ownership (whether it be an intergenerational transfer or an arms-length purchase from outside the family), change in operators, or change in headquarter location have the following impacts on the AGOPID: If a census-farm is sold or otherwise transferred as an on-going operation and the new operator's information (i.e. name and age) is available and is associated with the farm's location, the farm is considered continuing and its AGOPID is maintained. However, if a census-farm is bought by another existing farm, the farm is most likely treated as part of the new owner's existing operation and the farm's land, building and inventory are recorded in the Census of Agriculture questionnaires as an expansion of the buyer's farm. In that case, the AGOPID of the

[4] The CL-CEAG dataset includes industry variables of longitudinal farm type based on the North American Industry Classification System (NAICS; http://www.statcan.gc.ca/eng/subjects/standard/naics/2012/introduction#a8).
[5] http://www23.statcan.gc.ca/imdb/p3VD.pl?Function=getVDPage1&db=imdb&dis=2&adm=8&TVD=116940
[6] Some CEAG variables have evolved over time. Where feasible, a consistent variable is created. Some variables may not be found in all years, but may nonetheless be useful for analytical purposes.
[7] Note that a farm is not considered active if all land is rented out.

Table 1. Factors Related to Farm Exit between 2006 and 2011	Likelihood of exit	Signifi-cance
One-period farm and operator characteristics [Reference in brackets]		
Macro-Region [Ontario]		
Atlantic		
Quebec		
Prairies	++	***
Alberta	+	***
BC		
Farm Type [Grain and Oilseed]		
Beef Cattle	---	***
Hogs	+++	***
Horticulture		
Mixed		
Revenue Class [Under $50K]		
Between $50K and $250K	-	***
Between $250K and $999K	--	***
Over $1M	---	***
Age of 1st Operator [Under 30 years of age]		
30 to 39 years of age	-	**
40 to 49 years of age	--	***
50 to 59 years of age		
60 years of age and older	+++	***
Off-Farm Employment [No Off-Farm Employment]		
Part-Time Off-Farm Employment		
Full Time Off-Farm Employment	+	*
Number of Operators [One operator]		
Two	--	***
Three	---	***
Multi-period Variables (1996-2006)		
Change in Value of Machinery and Equipment (Bottom tercile [reference]		
Middle tercile	--	***
Top tercile	-	***
Change in Farm Size [Bottom tercile]		
Middle tercile	-	***
Top tercile	--	***
Change in Gross Farm Receipts [Bottom tercile]		
Middle tercile	--	***
Top tercile	-	***

Notes: +/- indicates the direction and strength of the influence of the variable on the likelihood of exit and are only reported when statistically significant. *,**,*** indicate significance at .01, .001 and .0001 levels, respectively. For categorical variables, reference groups are noted in parentheses.

purchased farm is terminated and no new AGOPID is created. If a farm is sold and bought but the new operator or head quarter location cannot be identified with the farm under the previous owner, the old AGOPID is inactivated and a new AGOPID is created.

Entry and exit rates generated by the CL-CEAG are qualitatively similar both to entry and exit rates generated by the CEAG going back to the early 1900s of between about 35 and 50 percent over 10 years, and to the U.S. (Bahar and Brown 2014). They are also are in line with manufacturing firm 10-year entry and exit rates of about 40 percent (see Baldwin 1998).

4. Examples of Modelling using the CL-CEAG

Longitudinal datasets are useful tools for many types of analysis. Here we first highlight the use of lagged and multi-period variables in the modelling of farm dynamics, specifically farm exists, and then the analysis of the factors associated with change over time in farm performance (i.e. profit margins). It is important to keep in mind that, while instructive, the results of these simple models presented here should be viewed as correlative and not aimed at identifying causal relationships.

Example A: Farm Exits

Farm exit and entry dynamics in Canada have been modelled with various datasets in the past, including multiple panels of two-period CEAG panel data and a sample of longitudinal CEAG records (Ehrensaft et al. 1984; Kimhi & Bollman, 1999). The CL-CEAG allows the use of micro-multi-period variables of farm dynamics and behaviour on the whole population of census-farms, which could be related to future continuation or exit decisions of census-farms.

Building on Freshwater (2015), we use a logit model to estimate the relationship of farm and farm operator characteristics on farm exits, including three multi-period variables identifying the farms' growth behavior. Table 1 shows the results for census-farm exits between 2006 and 2011, (i.e., of AGOPIDs which existed in 2006 but not in 2011). The first multi-period variable accounts

for the change of the value of machinery and equipment between 1996 and 2006, namely the ten years leading up to the anchor year (2006). The change in the value of machinery and equipment is assumed to be partly an observation of investment behaviour and reflective of the medium to long-term business objectives of the farm. The variable was created using terciles rankings within each farm type and farm size group in the anchor year, assigning each farm the value of Top, Middle, and Bottom Tercile, from greatest percent increase to greatest percent decrease in the value of machinery and equipment. The hypothesis is that farms planning on exiting are less likely to increase their investment in years prior to exiting than are farms in the other two categories. We also construct a similar variable of terciles of percentage change for physical farm size, using the number of acres for crop farms and herd size for livestock enterprises between 1996 and 2006, with similar a hypothesis.[8] Finally, a variable for change in annual gross farm receipts between 1996 and 2006 (in 2001 dollars) is included in the model to captures the trend of farm revenues in the years prior to the decision to exit or continue, which may have a different trend than physical size of the operation. In addition to these multi-period variables, the model also includes standard single-period variables in the farm exit model: region, farm type, gross farm receipts, off-farm employment and farm operator age.[9]

Given that the purpose of the paper is to illustrate the use of the dataset, it is sufficient to show the general tendencies rather than exact coefficient. This approach limits the impact of this analysis on the confidentiality of the dataset, as will be discussed further, below. The longitudinal farm growth indicators all have the expected signs with regards to the probability that the farm exits. Farms that were growing are less likely to exit than those that are decreasing in size. At the same time, the single-period variable farm revenue class (using farm gross receipts) is the strongest predictor of farm exits, and operator employment off the farm and the number of operators are also significant variables. The CL-CEAG enabled the modeling of the role of multi-period trends at the census-farm level on farm behaviour.

Example B: Farm Profit Margins

While the previous example shows the benefit of the CL-CEAG in enabling the use of multi-period variables for analysis in traditional regression models, this second example shows how the longitudinal dataset enables the modelling of change of farm attributes over the entire period (i.e. 1986 to 2011). This example models the factors impacting farm profit margins across the period.

The general linear mixed longitudinal model (GLMLM) is used because it is better suited to longitudinal analysis than the more common generalized linear model (GLM). While the GLM assumes that observations are independent of each other, the GLMLM takes accounts the fixed effects impacting observations of each subject, in addition to the random effects of each observation (Breslow and Clayton, 1993).

The dependent variable of the model is 'profit margin', calculated as Total Gross Farm Receipts minus Total Farm Business Operating Expenses plus Interest Expenses divided by Total

[8] The variables are change in farm acres (for grain, tobacco and horticulture farms); head of cattle (beef cattle farms); and number of pigs (hog farms).

[9] Since 1991, the CEAG allows for multiple operators, with no difference in status, i.e. there is no 'primary operator' designation. However, for this analysis we use the first operator that is identified, and assume that the first operator is most involved in the operator of the farm, i.e. the "primary" operator.

Gross Farm Receipts.[10] In addition to the one-period farm and farm operator characteristics, as in the farm exit model, the model includes CEAG variables associated with two farm management practices as a proxy for adoption of innovation: 'Use of computer for farm management' and 'Use of tillage practices', which references the adoption of no-till technology.

The results of the estimation are as follows. The coefficients of the farm and farm operator variables have broadly speaking the expected signs: Larger farms with older farm operators that dedicate their time to farm operations tend to have higher profit margin, after taking into account their farm type and location.[11] Farms with two operators are more likely profitable than those with either one or three operators. This might be a sign of the most stable farm management and farm size structure, perhaps due to husband and wife partnerships[12]. The variable 'Use of computer for farm management' has a positive and marginally significant effect of farm profit margins, while the variable 'Use of tillage practices' has no significant effect. Census year dummy variables were included in the model to account for external factors, such as market conditions and weather patterns; they were found not to be significant.

The model provides largely an intuitive set of associations. More work is required to develop a fully formed identification strategy. The longitudinal nature of the file, however, opens up several avenues of econometric analysis to pursue this, including the development of internal instruments, for instance using Systems- GMM (generalized method of moments) estimation.

5. Discussion

The CL-CEAG dataset has been developed to provide an additional tool for the analysis of the Canadian agricultural industry, given its broad range of farm characteristics and now the ability to use multi-period and lagged variables to identify the drivers of farm behaviour, including responses to macro-factors, such as changes in the infrastructure and cost of transportation, regional economies, and policies and regulations.[13]

However, the CL-CEAG is not free of limitations: The linking of farm records over survey time, like all firm-based data files, is not always clear-cut. Some changes occurring to the farm business require special consideration and judgement as to whether they represent the death and birth of a new business or the continuation of an on-going farm operation, such as the sale of the farm to new owner but without change of headquarter location or structure of the operation. Also, given the quinquennial nature of the CEAG, farm entries and exits which occur within the five year period are not observed and thus entry and exit rates may be underestimated (Katchova and Ahearn, 2015). In addition, the analysis of multi-period variables or the observation of dynamic trends is limited to those farm businesses that are in existence for at least ten years to be part of three censuses. Lastly, since the primary purpose of the CEAG is to provide information on Canadian

[10] Gross farm receipts include revenues from the sale of agricultural and forest products and agricultural custom work and machine rental, as well as payments from government programs and insurance proceeds. Operating expenses include business costs in the production of agricultural products plus wages and salaries paid to children and spouses. For incorporated farms, operating expenses may also include wages and salaries or rent paid to owner/operators (i.e. shareholders), while labour remuneration owner/operator of unincorporated farms is part of net operating income.
[11] The relationship with employment of other family members is not modeled and might have a different effect.
[12] Detailed results are available from the authors upon request.
[13] For example, Ferguson and Olfert (2015) used information for Census Consolidated Subdivisions in their work on the impact of the elimination of a grain transportation subsidy. Now, with the CL-CEAG the impact of individual farm characteristics can be analyzed.

census-farms at a point in time, the questions are not designed with the longitudinal dataset in mind. However, the development of future censuses might now take the longitudinal CL-CEAG dataset into consideration, so that lagged and multi-period variables could potentially become farm typology indicators.

The CL-CEAG dataset – like all micro-level data – must be managed in a manner that prevents residual disclosure of records, which repeated use of the data with slightly different methodologies can cause. Longitudinal business micro-datasets are inherently at greater risk of residual disclosure than datasets of populations, because the number of records tends to be smaller and the firms (farms) tend to be more easily identifiable than people, especially the larger firms (farms) in a given region or industry. The CL-CEAG is housed in Statistics Canada's Centre for Data Development and Economic Research (CDER), a centralized Research Data Centre at Statistics Canada headquarters that provides strictly managed access to business economic microdata for analytical research. Projects need to go through an approval process and are subject to extensive security requirements throughout the research process[14,15]. Due to the risk of residual disclosure, i.e. identification of respondents through combination of research results, steps have to be taken to limit the effect of published research projects subsequent work. Best practices, such as minimizing the use of descriptive data, have been employed to ensure the usefulness and availability of the dataset into the future.

6. Conclusion

This paper describes the CL-CEAG dataset and provides examples of this new tool for research of the dynamics of structural change of Canadian farms. The aim of Statistics Canada and Agriculture and Agri-Food Canada is that CL-CEAG will encourage much new research.

REFERENCES:

Ahearn M.C., Korb, P. Yee, J. (2009) Producer Dynamics in Agriculture: Empirical Evidence. In: Dunne, T. et al, eds. Producer Dynamics: New Evidence from Micro Data. University of Chicago Press.

Culver D. (2012). The Use of Panel Data for Policy Analysis in Canada. Presentation to the 2012 AAEA Annual Meeting. Seattle.

Baldwin, J.R. (1998) The Dynamics of Industrial Competition: A North American Perspective. Cambridge University Press.

Bollman R., Whitener L., Tung F.L. (1994). Trends and Patterns of Agricultural Structural Change: A Canada – United States Comparison. Statistics Canada Agriculture Division.

Breslow, N. E., Clayton, D. G. (1993), "Approximate Inference in Generalized Linear Mixed Models", Journal of the American Statistical Association 88 (421): 9–25,

[14] See the CDER website for details http://www.statcan.gc.ca/eng/cder/index

[15] The document 'Providing Access to Agriculture Microdata – A guide' by the Global Strategy (Global Strategy, 2014) provides an overview of the various approaches for access to microdata.

Brown, M. , Bahar S.. (2014). Longitudinal Census of Agriculture Project: Linkage Report. Unpublished.

Ehrensaft P., LaRamée, P., Bollman, R.D. and Buttel, F.H. (1984). The Microdynamics of Farm Structural Change in North America: the Canadian experience and Canada-U.S.A comparison. Can. J. Ag. Econ. 32(824-828)

Ferguson, S., Olfert, R. (2016) "Competitive Pressure and Technology Adoption: Evidence From a Policy Reform In Western Canada. Am. J. Agr. Econ. 98 (2): 422-446.

Freshwater, D. (2015) Factors Influencing Farm Exits: New Analysis using Linked Census of Agriculture Data. Unpublished.

Global Strategy (2014). Providing Access to Agriculture Microdata-A Guide. Global Strategy to improve agricultural and rural statistics (GSARS), FAO Statistics Division, Rome. Online (02-02-2016) http://gsars.org/en/providing-access-to-agriculture-microdata-a-guide/

Hoppe R., Korb P. (2006) Understanding U.S. Farm Exists. USDA/ERS. Online 2015-08-15 www.ers.usda.gov

Kapitany, M., Bollman, R. D. (1983). Entry, Exit, and Structural Change in Agriculture: Summary Results from the 1966 to 1981 Census of Agriculture Match. Proceedings of the Business and Economics Section of the American Statistical Association (p.100-109).

Katchova A.L., Ahearn M.C. (2015). Dynamics of Farmland Ownership and Leasing: Implications for Young and Beginning Farmers. Appl. Econ. Perspect. Pol.

Kimhi A., Bollman R. D. (1999). Family farm dynamics in Canada and Israel: the case for farm exits. Agricultural Economics v.21 (69-79)

Nagelschmitz, K. (2007). Is low income persistent among Canadian farm families? A longitudinal profile, 1983 to 2004. In: Pacioli 14 Changes in farming and the effects on FADNs, Poppe K., Boone K., Teeuwen-Vogelaar C. (Eds.). The Hague, LEI

Shapiro D., Bollman R. D., Ehrensaft, P. (1987). Farm size and growth in Canada. Am. Ag. Econ. V.69(2) (477-83)