Using Statistics Canada's Microdata to Evaluate Policy Impacts: The AgriInnovation Stream C - Commercialization Program

Nov 2, 2023

Motivation

The AgriInnovation-Stream C and AgriInnovate Programs

AgriInnovation-Stream C (2013-14 to 2017-18) and AgriInnovate (2018-19 to 2022-23) aimed to accelerate the commercialization phase of the innovation process in the sector using interest-free, repayable contributions for eligible innovation projects.

Economic theory

Government innovation subsidies incentivize firms to invest more in such activities, thereby improving the performance of the individual firms as well as the broader economy.

Research question:

Are AAFC's AgriInnovation Stream C and AgriInnovate Programs effective in improving the economic performance of the recipient firms?

Data

Source: Statistics Canada's Linkable File Environment (LFE) and Diversity and Skills Database (DSD)

Time frame: 2005-2020, some of the main series end in 2017.

This database contains **firm-level**:

- Financial information: Income tax data
- Employment information: Payroll deduction accounts
- Other variables: Location, research and development and gender of owner

Table 1. Program participants, AgriInnovation-Stream C and AgriInnovate programs, 2013-2020

Program	AgriInnovation-Stream C					AgriInnovate			Total
Year	2013	2014	2015	2016	2017	2018	2019	2020	
Number of participants	9	10	9	12	5	6	8	5	**64**

Table 2. Number of observations and firms in the dataset, 2005-2020

	Participants		Non-participants	
	Firms	Obs.	Firms	Obs.
All (in scope)	64	809	106,147	945,383
Used in main regressions	39	377	676	1,160

Note: Some series end in 2017. As such, the 19 firms that participated in the program after 2017 could not be used in the main regressions. Similarly, good matches could not be found for six firms, leaving 39 in-scope firms for the main model.
Source: Statistics Canada's LFE and authors' estimations.

Methodology

Objectives:

- To investigate a causal link from participation in the program to net income; and
- To explore factors that affect program participation.

Approach:

Step 1: Build control group

Matching: each participant was matched with a non-participant of similar size (and other characteristics) from the same industry category and year (see Annex 4).

Step 2: Estimate program impact

Apply a difference-in-difference (DID) regression model to the matched observations:

- DID refers to the difference of differences in the outcomes of participants and non-participants before and after the program (see Annex 2).
- **Net income** estimated as a function of: **program participation**, labour, capital, and other important variables.

Advantages:

- Matching: Assurance of random selection – i.e., no false comparisons
- DID: Control for some firm-specific factors – e.g., managerial talent
- Together, matching and DID mostly address self-selection bias (we took additional measures to address remaining concerns)

Summary Statistics

Table 3. Distribution of small, medium, and large firms in the sample, before matching, 2010-2019

	Participants		Non-participants	
Number of Employees	Obs.	Share (%)	Obs.	Share (%)
Small (1 to 99)	412	58.2	454,281	97.3
Medium (100 to 499)	229	32.3	11,223	2.4
Large (>500)	67	9.5	1,388	0.3
Total	708		466,892	

Table 4. Distribution of firms across industries, before matching, 2010-2019

	Participants		Non-participants	
Industry	Obs.	Share (%)	Obs.	Share (%)
Farming	201	19.3	689,911	71.4
Manufacturing	477	45.9	111,769	11.6
Wholesale	102	9.8	143,108	14.8
Other*	260	25.0	21,827	2.3
Total	1,040		966,615	

Notes: 1. "Other" includes smaller categories such as real estate, finance, insurance, supporting services, etc.; 2. Number of observations are different in the two tables. This is because they are obtained from different sources and thus their missing observations do not necessarily align with one another – e.g., while Table 3 is based on *Number of employees*, which originates from payroll deduction accounts and ends in 2017, Table 4 is based on North American Industry Classification System (NAICS) codes, which originate from income tax filings and end in 2020. *Source*: Statistics Canada's LFE.

Summary Statistics

Table 5. Summary statistics, financial and socio-demographic variables, before matching

Variable	Participants		Non-participants	
	Mean	Obs.	Mean	Obs.
Total assets (million $)	144	809	3.2	945,383
Total revenue (million $)	184	809	3.2	945,383
Total expenses (million $)	175	809	3.0	945,383
Net income (loss) before tax (million $)	9.3	804	0.2	927,275
Gross profit (loss) (million $)	25.1	802	0.5	892,129
Share of immigrant employees in workforce (%)	20	594	12	463,879
Business share held by immigrants (%)	22	315	9	598,220
Share of female employees in workforce (%)	23	594	35	463,879
Business share held by women (%)*	41	183	34	307,672
Average age of all paid employees (years)	38	594	42	463,552
Average age across all owners (years)	49	315	53	597,014
Share of single majority owners (%)	36	315	55	598,781

Note: Number of observations are different for different variables. This is because they are obtained from different sources and as such their missing observations do not necessarily align with one another – e.g., while the top 5 rows (financial variables) are obtained from the LFE and end in 2020, the bottom 7 rows (socio-demographic variables) originate from the DSD and end in 2017. *Source*: Statistics Canada's LFE and DSD.

Estimated benefits

- **Average program impact ($)**: a **$33 million** increase in the total revenues and a **$28 million** increase in the total expenses, leading to a **$5 million** increase in the net income before tax of an average participant over the 2013-2017 period.
- **Aggregate benefit**: around **$200 million** increase in net income before tax of the 39 participants over the 2013-2017 period, or $40 million per year.

Table 6. Summary of main findings: average program impact, 2013-2017.

	Program impact on an average participant	
Financial outcome	**Average effect (million $)**	**Marginal effect ($) (effect of a dollar of funding)**
Total revenue	33	8.8
Total expenses	28	7.7
Net income (loss) before tax	5	1.1

Note: To calculate the aggregate benefit, we use the average effect ($5.12 million × 39=$199.9 million) rather than the marginal effect. This is because the latter is only valid for the interpretation of small changes.

Estimated costs

- **$137.3 million in interest-free loans** were made to the participants during the 2013-2017 period.
- It takes an average participant **13 years to fully repay AAFC**:
  - 2 year completion time + 1 year repayment lag + 10 year repayment schedule = 13 years
- The **opportunity cost** of the $137.3 million interest-free loans is approximately **$50 million:**
  - Interest cost of $137.3 million @ (5% interest, 13 years repayment, monthly payments)= $50 million

Effectiveness: Public perspective

Scenario I (base case):

Over the 2013-2017 period,

- **Aggregate benefit**: around **$200 million** increase in net income before tax of the 39 participants.
- The **opportunity cost** of the $137.3 million interest-free loans is approximately **$50 million.**

Aggregate benefit ($200 million) > Opportunity cost ($50 million)

Caveat:

- Program contributions could not exceed 50 percent of eligible project costs. Since program contributions are approximately $137.3 million, the participants must have spent at least another $137.3 million, leading to a total cost of at least $274.6 million.
- While non-capital expenses are fully reflected in the participants' net income before tax, capital expenses may only be partially reflected because they cannot be claimed for tax purposes all at once.
- If there are unclaimed capital costs, then the $200 million estimated program benefit may be an overestimation.

Effectiveness: Public perspective

Scenario II (65% potentially unclaimed capital costs):

- Assume that 65% of the total eligible costs – i.e., $178.5 million – is spent on Class 6 capital items such as buildings, which have one of the lowest depreciation rates (10%).
- Even for this capital class, on average, firms could claim up to 22% of the costs as capital cost allowance (CCA) within five years. The remaining 78% potentially unclaimed costs amount to $139 million.

Aggregate benefit - Potentially unclaimed capital costs > Opportunity cost

($200 million- $139 million= $61 million> $50 million)

- Even under the most pessimistic assumptions, the program has been welfare-improving from a public perspective.

Notes:

- Depreciation rates for Class 8 (e.g., equipment without motors) and Class 10 (e.g., machinery with motors) are 20 and 30 percent, respectively. Class 12 items (e.g., kitchen utensils and computer software) depreciate at 100 percent.
- In Canada, for the first year, only half of the 10 percent Class 6 depreciation rate can be claimed as CCA. By the end of the second, third, fourth, and fifth years, firms are able to claim 15, 23, 31, and 38 percent of the cost for a Class 6 item, respectively.

Conclusions

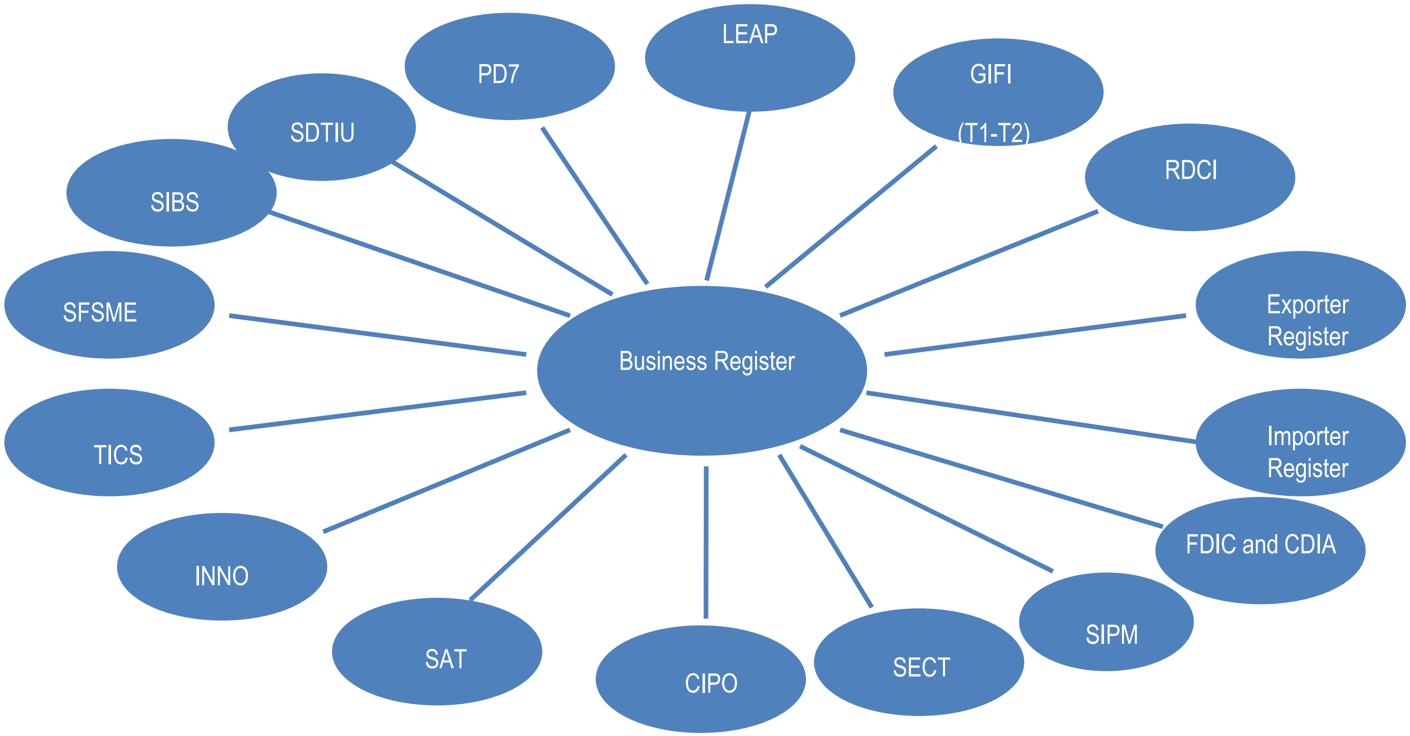
- From a private perspective, the $50 million opportunity cost is irrelevant. Thus, on average, the program is more effective from a private perspective than from a public perspective.
- The benefits of the program to the economy could go beyond the benefits experienced by the recipients in the regression analysis:
  - The estimated benefit ($200 million) applies to the 39 recipients in the regression models, not all the 45 recipients;
  - The projects could benefit consumers by offering higher-quality, low-cost, or a broader variety of products; and,
  - Other firms in the sector (downstream and upstream) could benefit from the program.
- After the 5-year study period, there could be more costs (e.g., interest costs) and benefits.
- Future research could re-examine the effectiveness of the programs with an updated dataset to:
  - Capture the benefits of the programs after the 5-year study period; and
  - Include the AgriInnovate Program participants.

Thank you for listening.

Questions or comments?

ANNEXES

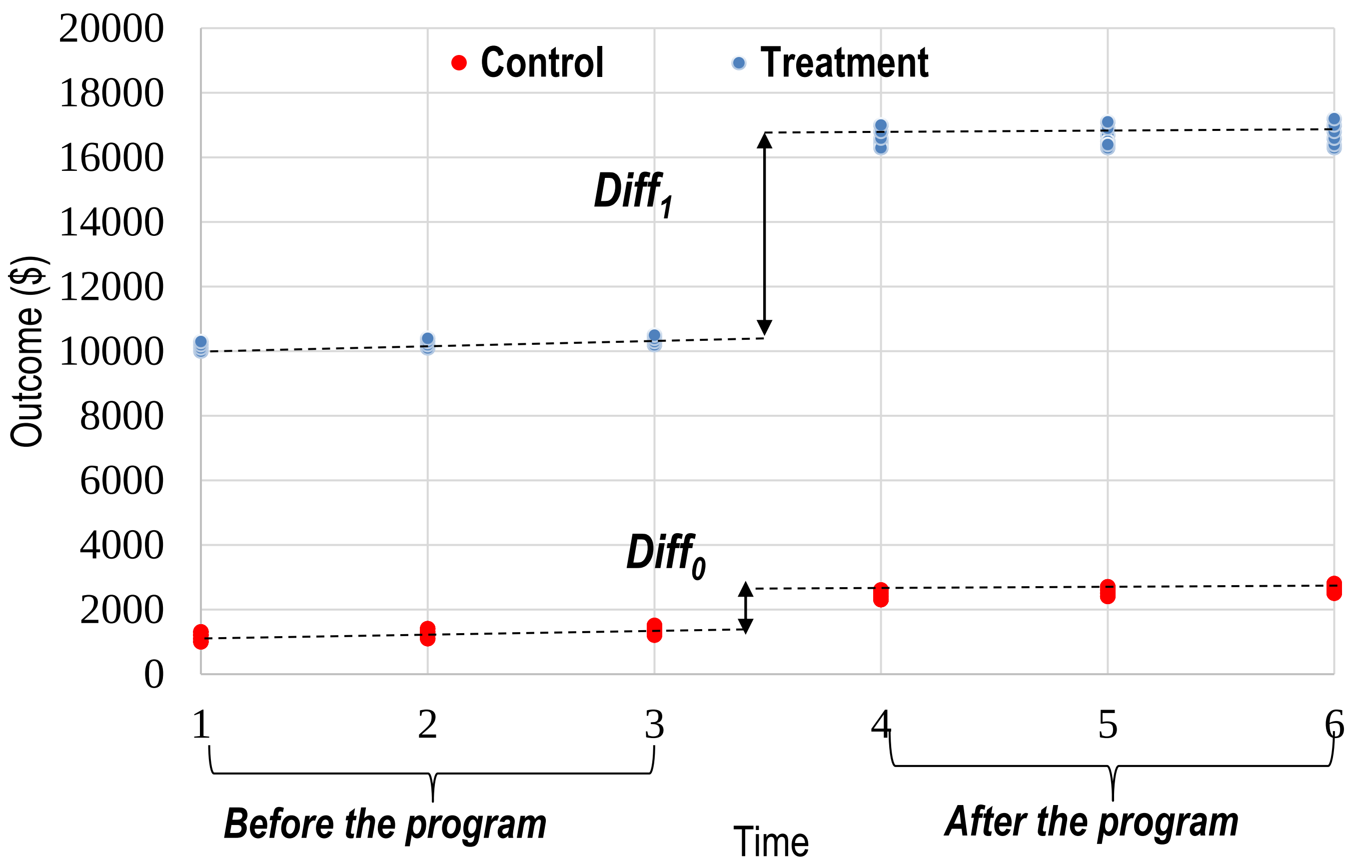
Annex 1: Linkable File Environment Galaxy



List of acronyms:
RDCI – Research and Development in Canadian Industry
GIFI – CRA T1 and T2 Income statement and Balance sheet
LEAP – Longitudinal Employment Analysis Program
PD7 – Business Payrolls Survey (based on CRA payrolls deduction form PD7)
SDTIU – Survey of Digital Technology and Internet Use
SIBS – Survey of Innovation and Business Strategy
SFSME – Survey on Financing of Small- and Medium-sized Enterprises
TICS – Trade in Commercial Services
INNO – Survey of Innovation
SAT – Survey of Advanced Technology
CIPO – Canadian Intellectual Property Office
SECT – Survey of Electronic Commerce and Technology
SIPM - Survey of Intellectual Property Management
FDIC – Foreign Direct Investment in Canada
CDIA – Canadian Direct Investment Abroad

Annex 2: Difference-in-difference (DID) Models

Figure A1. Treatment effect in difference-in-difference models: a hypothetical case

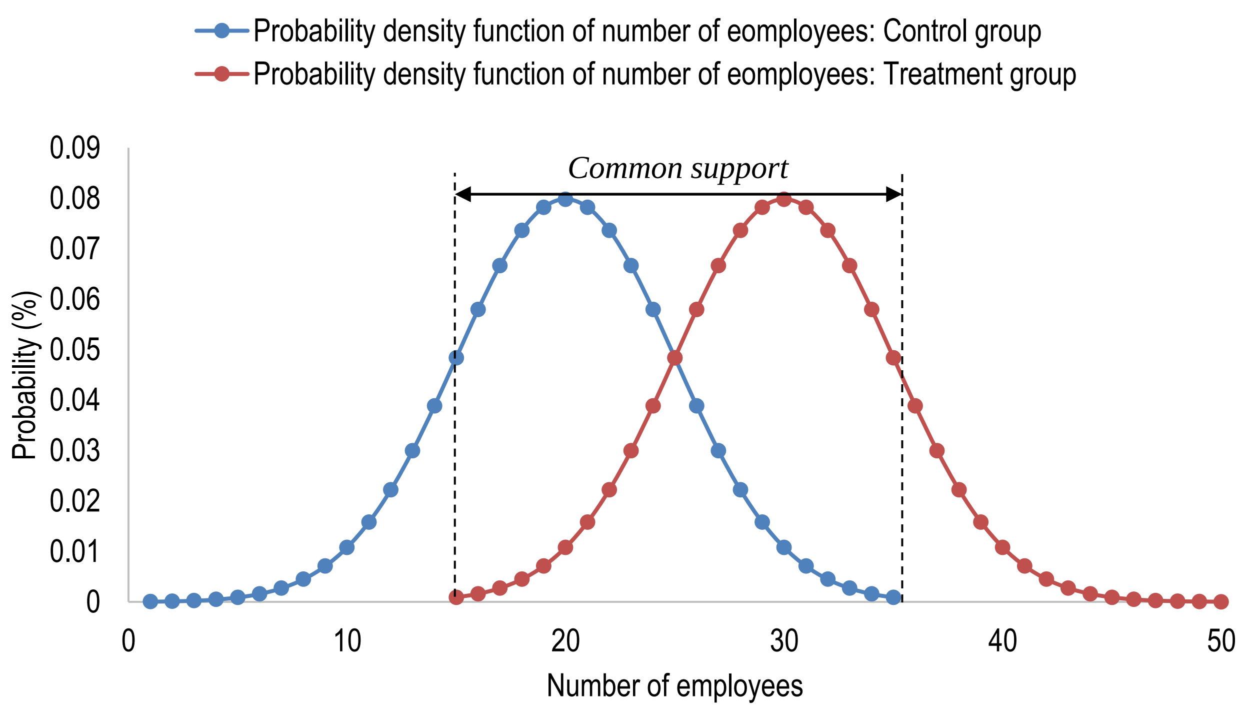


Treatment effect= $Diff_1$ - $Diff_0$

Annex 3: Common support

Common support: the overlapping area of two distributions.
Common support assumption: the treatment and the control group must have common support (i.e., be comparable) for *all* observable characteristics.

Figure A2. Example of common support for number of employees for hypothetical control and treatment groups



Annex 4: Propensity Score Matching

1. To match participants and non-participants, we first **estimate the propensity of participation** (using a logistic regression model) as a function of firms' observable characteristics:
   - *Share of immigrant employees* (%)
   - *Share of female employees* (%)
   - *Average age of employees* (years)
   - *Number of employees*
   - *Salaries and wages* ($)
   - *Total assets* ($)
   - *Year* (2005-2020)
   - *Province* (where the firm is headquartered)
   - *Industry category* (56 distinct categories based on their NAICS codes)
   - *Exporter status* (represents whether the firm is an exporter)
   - *Years since birthdate* (represents age of the business)

2. Using *Greedy matching* algorithm, **each participant is then matched with a non-participant from the same industry category and calendar year** whose propensity score is within a pre-specified threshold (*radius caliper*) of that of the participant.

Annex 5: DID regressions

The **DID** model is estimated as a **fixed effects (FEs) panel** of the following form:

(2) $Y_{it} = \alpha + \beta(\boldsymbol{Treatment_{it}}) + X_{it}\lambda + \gamma_i + v_{it},$

where Y_{it} is a measure of financial performance – e.g., net income – of firm *i* in time *t*, $\boldsymbol{Treatment_{it}}$ is a dummy variable that takes the value of 1 for participant firms after participation and 0 otherwise, and the set of other explanatory variables X_{it} is as follows:

- *Share of immigrant employees (%)*
- *Share of female employees (%)*
- *Salaries and wages ($)*
- *Total assets ($)*
- *Year* (2005-2020)
- *Year* × *Province*: interaction of calendar years and provinces
- *Year* × *Industry category*: interaction of calendar years and industry categories

Notes:

1. α is the constant term, β is the parameter that measures the effect of the program, λ is a set of parameters to be estimated, γ_i represents firm-specific (observable and unobservable) time-invariant characteristics or fixed effects, v_{it} reflects the remainder stochastic disturbances.
2. *Province* and *Industry category* are time-invariant for each individual firm and thus their effect is already captured in the firm-specific fixed effects – i.e., including them in the model would be redundant.

Annex 6: Self-selection bias (picking winners)

Problem: there may be something intrinsic about the participating firms or their managers that positively affects both their likelihood of participation and their outcome.

Example: Opportunistic managers vs. less opportunistic managers

Implication: The estimated program impact does not reflect the effect of the program alone but is confounded with the effect of the firms' or their managers' intrinsic abilities.

Solution: If source of bias is

- Observable characteristic (e.g., size, value of assets)

  → **Matching**

- Unobservable time-invariant (e.g., the manager's character)

  → **Panel DID with FEs**

- Unobservable time-variant (e.g., the way managers' expertise differentially evolve over time)

  → Our **supplementary approach**: test whether the rejected applicants experienced a statistically significant increase in their performance after their rejection date and relative to a **control group**.

The **control group** consists of firms that are similar to the rejected applicant with respect to observable characteristics but never applied for the program. The logic behind this approach is simple: if the increase in financial performance is all due to unobservable factors – e.g., having an innovative idea or having an opportunistic manager – and not affected by the program at all, then the rejected applicants would be as likely as the accepted applicants to experience an increase in their performance