#### The Intergenerational Income Database (IID): a historical perspective and some applications

Y. Ostrovsky, Statistics Canada

QICSS workshop, Montreal, June 11



#### **STATISTICS CANADA ONE HUNDRED YEARS AND COUNTING**







### Background

- 1980s-early 1990s: rapid developments in the literature on intergenerational income mobility.
- Emergence of data suitable for intergenerational analysis, i.e., panels of sufficient length.
- United States: Panel Study of Income Dynamics (PSID) (since 1967) and National Longitudinal Survey (NLS) (since 1966).
- European (Scandinavian) countries: administrative files.

#### Canadian studies

- No economics studies had been done before the existence of the IID.
- Some studies on social mobility (education, occupations, etc.).
- Not much was known about the intergenerational income transmission in Canada.
- "The lack of research in the area of intergenerational income mobility in Canada suggests that almost any discussion of the issue will be a significant contribution to our understanding of the way the Canadian labour market functions." (1994 Corak and Heisz's linkage proposal).

## Statistics Canada T1 Family Files

- T1 Family Files (T1FF of TIFF): constructed annually since 1982.
- TIFF is based on the *census family* concept, which is specific to Statistics Canada.
- <u>Census families</u> are married couples or couples living common law with or without children, or lone parents with at least one child living in the same dwelling.
- Related to TIFF: Longitudinal Administrative Database (LAD) was a 10% longitudinal sample of the TIFF (currently 20%).

### TIFF

- Taxfilers from the same *census family*, including children, are matched using common links (e.g., spousal SIN, same name, and same address). Prior to 1993, non-filing children were identified from information on their parents' tax form and using information from the Federal Family Allowance
- Program.
- Since 1993, children are added to the census family by using the Canada Child Tax Benefit file, provincial births files and previous years of the TIFF.

# IID

- By the 1990s: a need for a database suitable for intergenerational income transmission studies in Canada.
- The idea was to use the LAD and additional T1 data for 1978-1981 + 1991 +1992.
- The proposal mentioned only for sons-fathers links.
- Kids: 16 to 19 in 1982.
- The income of children would be measured at 26 to 29 years of age.

Business and Labour Market Analysis Division: Miles Corak and Andrew Heisz.

b) No. 003-95: Intergenerational Income Mobility in Canada

> This proposal would involve selecting a sample of sons aged 16 to 19 in 1982 from the T1 Family File (T1FF) and extracting their SINs and those of their fathers from the T1FF along with the demographic and income variables for the years 1982 to 1990. It would also be necessary to link

Excerpt from the original 1995 linkage proposal

#### Canada

the sample of SINs to the 1978, 1981, 1991 and 1992 T1 files to obtain the same information to extend the length of the study panel.

Policy Committee approved the request. It was also agreed that a substantive letter from the Chief Statistician to the Deputy Minister of Revenue Canada should be prepared and sent at the time of the public release of the output resulting from any record linkage activity involving exploitation of tax records. Louise

 $(\mathfrak{B})$ 

#### - 2 -

#### 7. FREQUENCY

This would be a series of one-time linkages and would be conducted under the terms of this analytical project only.

#### 8. FILE RETENTION

#### The files created would be retained for a period of five years.

- The proposal was approved in February 1995.
- was too small?).

First iteration: used 1982 TIFF instead of the LAD (the LAD-based sample)

Second iteration (addendum): used a 5-year window to select a cohort.

#### Earlier studies

- First study: Corak and Heisz, "The Intergenerational Income Mobility of Canadian Men." *Canadian Business Economics*, 1995 (?)
- A paper with the same title was published in January 1996 in the Analytical Studies Branch Research Paper Series (No. 89).
- The first major IID-based study published in academic literature was Corak and Heisz (1999).
- Computed IGE (intergenerational income elasticity) for Canada.
- Discussed IID advantages and limitations.

## IID issues discussed in Corak and Heisz (1999)

- <u>Representativeness</u> of the IID sample.
- We'd want the sample to be representative of all children in a given birth cohort.
- To be in the IID sample, however, children.

✓ had to have a SIN in the year in which the link with their family was established (i.e., when they were 16-19)\*.

- $\checkmark$  had to live with their parents (at least one parent) for the link to be established.
- $\checkmark$  had to have a father present.

- Concerns:
  - during their teenage years.
  - attachment.
  - in their adult years.

✓IID may underrepresent children with no attachment to the labour market

 $\checkmark$ IID may underrepresent children who left home before establishing an

✓IID may underrepresent children who had little labour market attachment

Overall, IID may underrepresent children who become low-income adults.

- Incentive to underreport income to the tax authorities.
- Specific questions:

 $\checkmark$  Are these incentives strong?

 $\checkmark$  Do they vary across the income distribution?

 $\checkmark$  Do they affect parents and children differently?

Ten income brackets in the late 1970s and the early and mid 1980s.

Another set of concerns is related to the <u>accuracy</u> of reported income.

#### Addressing IID limitations

- CH point out that these problems are not unique to the IID.
- Re: accuracy. Census (and SLID at the time) link respondents to their tax data.
- CH compare tax data and the Survey of Consumer Finances.
- Compare various points of income and earnings distributions: most of the differences are in the tails.
- CH further explore sensitivity of their results to the exclusion of children for whom a link to their parents (father in their study) could not be established.
- Conclude: "our sample selection rules do not lend a major bias to the findings."

#### 1998

- A new linkage proposal.
- From the new linkage request:

The analysis to date has established the viability of using administrative data to examine issues of this kind, and offered some important findings that speak to current public policy concerns. However, it has also highlighted certain limitations of the data. The objective of the current proposal is to overcome these limitations and extend the ability of the data to address more detailed research questions in this general area.

#### 1998

- The coverage is expanded to incorporate mothers and daughters. Two more cohorts are added to the IID: 1984 and 1986.
- T1 files: up to 1995.
- Regular updates are planned for T1 files (but not new cohorts). The file is proposed to be retained indefinitely.
- T4, ROE, LEAP, STVC, BNOP...

## IID weights

- Around the same time (late 1990s), IID weights were introduced to correct for potential under-representativeness of the IID.
- Problem: the IID covers <sup>2</sup>/<sub>3</sub> to <sup>3</sup>/<sub>4</sub> of children in each cohort (based on Census data).
- First step: used the LAD (later T1FF) to adjust counts for non-filing kids using parents' market income and place of residence for stratification.
  Second step: used Census data to adjust for T1FF underrepresentation or
- Second step: used Census data to adjust for T1 overrepresentation relative to Census counts.

### 2010

- Added new T1 files (up to 2008).
- Added information about children's spouses.
- Added a few variables.

# Chen, Ostrovsky & Piraino (2017)

- After CH was published in 1999, IID improved in many important ways.
- A chance to re-estimate the IGE for Canada using better data and knowing more about potential pitfalls.
- Build on Mazumder (2005), Grawe (2006), Haider and Solon (2006) and Nybom and Stuhler (2016).

# Chen, Ostrovsky & Piraino (2017)

Main objective is to (re)estimate the IGE for fathers and sons Canada

 $Y_i^s = \alpha$ 

- We consider various income definitions.
- fathers and daughters.
- transmission.

+ 
$$\beta Y_i^f$$
 +  $\varepsilon_i$ 

In addition, we consider intergenerational income transmission between

• We also revisit the issue of nonlinearities in the intergenerational income

- Since the lifetime income is usua incomes as proxies.
- Two main issues related to using proxies for lifetime income: Problem 1. The measurement error in the fathers earnings.
- A well-understood problem leading to attenuation bias.

$$Y_i^s = \alpha + \beta Y_i^f + \varepsilon_{it}$$
$$Y_i^s = \tilde{\alpha} + \tilde{\beta} Y_{it}^f + \tilde{\varepsilon}_{it} = \tilde{\alpha} + \tilde{\beta} \left( Y_i + u_{it} \right) + \tilde{\varepsilon}_{it}$$

Since the lifetime income is usually unknown, most studies use annual

Problem 2. Lifecycle bias.

 This is a particularly serious issue when the incomes of children and parents are measured at different points of their lifecycles.



$$Y_{it} = \lambda_t Y_i + u_{it}$$

#### Lambda changes throughout the lifetime.



## Summary of results

- The IGE for Canada is estimated at about 0.32 (compared to 0.22 in CH).
- The lifecycle bias explains about  $\frac{3}{5}$  to  $\frac{2}{3}$  of the discrepancy.
- The error-in-variables bias explains the rest.
- Significant gender differences with regard to these biases.
- The father-daughter elasticity remains modes irrespective of the ages at which daughters' earnings are measured.
- Interesting patterns of non-linearity: high mobility at the very bottom of the distribution and low mobility at the top.

#### Latest update

- A major IID update in 2016 thanks to the financial support from *Groupe de recherche sur le capital humain* lead by Prof. Connolly and *Fonds de recherche du Québec - Société et culture* (FRQSC) .
- Three new cohorts of children: a chance to look at mobility changes across birth cohorts.
- Ostrovsky (2017) looks at changes in absolute mobility across different cohorts of children.

## Relative vs. absolute mobility

INCOME 0% relative mobility AGE



• Chetty et al. (Science, 2017): a dramatic decline in absolute intergenerational income mobility rates in the U.S in the past several decades.





# Absolute mobility in Canada

- Children aged 16 to 19 in 1982, 84, 86, 91, 96 and 01 are linked to their parents.
- Tax returns from 1978 to 2014.
- Main focus is on children's and parents' family incomes when they were 30 years old.
- Compute the share of children in each cohort whose family income was at least as high as that of their parents

 $M_c = 100\%$  ×

$$\langle N_c^{-1} \sum_i 1 \left\{ y_{ic}^k \ge y_{ic}^p \right\}$$

#### Chart 1

#### Absolute income mobility rate at age 30, by birth cohort of child, adjusted and unadjusted family income, 1970 to 1984

absolute income mobility rate (percent)



Adjusted family income

Note: All incomes were converted to 2015 constant dollars using the all-items Consumer Price Index. Source: Statistics Canada, Intergenerational Income Database.

Unadjusted family income

## How does Canada compare to the US?

 A. Estimated rates of absolute mobility at 30 (unadjusted)

 1970
 1977
 1984

 United
 61%
 55%
 50%

 Canada
 48%
 55%
 52%

C. Estima	ated rate	es of ab	solute	
mobility a	nt age 4	0		
(unadjusted)				
	1963	1967	1974	
United	58%	58%	55%	
Canada	52%	58%	60%	

**B**. Estimated rates of absolute mobility at 30 (adjusted)

	1970	1977	1984	
United				
States	72%	68%	59%	
Canada	60%	66%	64%	
In LLS: family income is divided by				

In US: family income is divided by the sq. root of the family size In Canada: family income divided by the number of adults

## Concluding remarks

- The IID has been the main source of our understanding of intergenerational income mobility in Canada for almost 25 years.
- It is valued internationally, and there is a substantial academic literature based on IID data.
- It has limitations, but it is much richer now than it was at the beginning.
- It is now available at RDCs: data, user guides (in French and English) + full support of the RDC staff.
- Feedback is most welcome.