README: Replication Files
July 2025

July 24, 2025

1. Overview

This README file contains instructions for replicating the results in Adnan et al. (2025).

2. Data Access

2.1 Confidential Data Files

The data for this project are confidential microdata from Statistics Canada. To gain access to the microdata, follow the directions here on how to write a proposal for access to the data via a Research Data Center: https://www.statcan.gc.ca/en/microdata/data-centres/access. You will need to request the following files:

- *Longitudinal Immigration Database*: These files contain landing file information for immigrants along with linkages between parent and child immigrants.
- *1981, 1986, 1991, 1996, 2001 Census*
- *Longitudinal Adminstrative Databank (LAD)*: these files contain household and individual income information from tax files for twenty percent of the Canadian population. We use the files from the years 1982 to 2019.
- *Longitudinal Survey of Immigrants to Canada (LSIC)*

3. Directory Setup

There are four main folders in the directory.

1. "Raw Data": contains the publicly available data files
    - "world population.csv": population by country and year from Organization for Economic Cooperation and Development (n.d.)

- "asylum.csv": number of refugees admitted by country (UNHCR, 2025) [1]
- "canada.csv": refugees admitted in Canada per year (Immigration and Canada, 2003, 2014)
- "frac_censusdivision.csv": information on intergenerational mobility in census divisions from Corak (2020)[2]
- "cpi2002.csv": Consumer Price Index, annual average, not seasonally adjusted, Table: 18-10-0005-01 (Statistics Canada, 2025)
- "conflictdata.csv": data on conflict intensity from the Global Database of Events, Language and Tone (The GDELT Project, 2025)

2. "Program Files": contains all the coding files

Note: the confidential data files are accessed from their original file path which is provided by Statistics Canada.

4. Program Files

4.1 Raw Data

IPUMS

- Go to IPUMS INTERNATIONAL (https://international.ipums.org/international-action/samples) and download the following datasets into "Raw Data": for each dataset choose the following variables from the individual tab: age, edattain
  - Philippines, 1995, → save as "phillippines1995.dta"
  - United Kingdom, 2001, → save as "uk2001.dta"
  - China, 2000, → save as "china2000.dta"
  - Poland, 1988, → save as "poland1988.dta"
  - Jamaica, 1991, → save as "jamaica1991.dta"
  - Pakistan, 1998, → save as "pakistan1998.dta"
  - Vietnam, 1989, → save as "vietnam1989.dta"
  - United States, 1980, → save as "us1980.dta"
  - Romania, 1992, → save as "romania1992.dta"
  - Portugal, 1991, → save as "portugal1991.dta"
  - Haiti, 2003, → save as "haiti2003.dta"
  - Trinidad and Tobago, 1990, → save as "trinidad1990.dta"

[1] https://www.unhcr.org/refugee-statistics/insights/explainers/forcibly-displaced-flow-data.html

[2] https://milescorak.com/wp-content/uploads/2017/06/corak-divided-landscapes-geography-economic-oppor xlsx

  - South Africa, 1996, → save as "southafrica1996.dta"
  - Russia, 2002, → save as "russia2002.dta"
  - Ukraine, 2001, → save as "ukraine2001.dta"
  - France, 1990, → save as "france1990.dta"
  - West Germany, 1987, → save as "westgermany1987.dta"
  - Salvador, 1992, → save as "salvador1992.dta"
  - Iran, 2006, → save as "iran2006.dta"
  - Thailand, 1990, → save as "thailand1990.dta"
  - Laos, 2005, → save as "laos2005.dta"
  - Cambodia, 1998, → save as "cambodia1998.dta"
  - Iraq, 1997, → save as "iraq1997.dta"
  - Guatemala, 1994, → save as "guatemala1994.dta"
  - Nicaragua, 1995, → save as "nicaragua1995.dta"
  - Ethiopia, 1994, → save as "ethiopia1994.dta"
  - Ghana, 2000, → save as "ghana2000.dta"
  - Hungary, 1990, → save as "hungary1990.dta"
  - Chile, 1992, → save as "chile1992.dta"
  - Bangladesh, 2001, → save as "bangladesh2001.dta"
  - US, 1990, → save as "us1990.dta"

- The file *ttest.do* in *Programs/Raw Data/ipums* cleans the data and calculates the education shares for stayers in Table 2

Longitudinal Administrative Databank

The files in *Programs/Raw Data/LAD* clean the LAD data.

- *Step1_calculateLADpercentiles.do* calculates individual and household income percentiles for the Canadian population
- *Step2_mergeallLADfiles.R* cleans the percentile files and appends them together
- *Step3_LADpct_bootstrap.do* prepares the LAD data for the bootstrap exercise.
- *Step4_merge_lad_boostrap.R* merges all the bootstrapped LAD data.

Output files:

- *Derived Data/LAD/finaldist.csv* : individual and household income percentiles for Canadian population
- *Derived Data/LAD/all_lad_bootstrap*: bootstrapped individual and household income percentiles for Canadian population

IMDB Database

The files in *Programs/Raw Data/IMDB* clean the IMDB Database files.

- *InstallPackages.R* installs the R packages needed.
- *0_child_parent.R* creates a child-parent linked dataset from the IMDB. Output: *Derived Data/child_year_panel2025.csv*
- *1_CleanData.R* uses *Derived Data/child_year_panel2025.csv* and calculates parent and child income percentiles using the LAD income distribution in *Derived Data/LAD/finaldist.csv* . The program creates *Derived Data/IGM_dataset_2025.csv*
- *2_CleanData_bootstrap.R* merges IMDB parent and child information and creates bootstrapped income percentiles.
- *3_SummaryStatistics.R* creates the derived dataset "cohort_2025.csv " - this is the main file used for analysis.
- *4_CreateConnollyVersion.R* creates the derived dataset using the income definitions from Connolly et al. (2023)

LSIC

- *LSIC/0_importlsic.R* cleans the Longitudinal Survey of Immigrants dataset.

Census

- The file *Census/immenclaves.R* creates the enclave definitions.
- *Census/Clean_Census_earnings.do* cleans the Canadian census files and harmonizes the occupation codes with the IMDB.
- The file *Census/LandingCensusTract.R* gathers information on the first census tract of each immigrant and merges them to enclave tract information.

4.2 Output Files

These files are located in *Programs/Output* and create the output for the paper.

- *Exhibit_Table1.R* calculates the summary statistics for Table 1.
- *Exhibit_Figure1.R* creates Figure 1.
- *Exhibit_Figure2_CoefPlots.R* creates Figure 2.
- *Exhibit_Table2.R* calculates the results in Table 2
- *Exhibit_Table3.R* calculates regressions for Table 3 and Table B.1.

- *Exhibit_Figure3.do* and *Exhibit_Figure3_CoefPlots.R* create Figure 3
- *Exhibit_Table_Figure4.R* calculates income prediction exercise and creates Table 4, Table B.5, and Table B.6
  - The code calls *IncomePrediction_Plots.R* which makes and Figure 4 and Figure B.10
- *FigureB1.R* creates FigureB1 using "asylum.csv" and "canada.csv"
- *Exhibit_FigureB2.R* creates Figure B.2.
- *Exhibit_FigureB4.R* creates Figure B.4.
- *Exhibit_FigureB5.R* creates Figure B.5
- *CreateFigureB10.R* creates Figure B.10
- *Exhibit_FigureB8_11.R* creates Figure B.8 and Figure B.11
- *Exhibit_TableB2.R* calculates the boostrapped standard errors for Table B2
- *Figure B3.R* create Figure B3
- *Exhibit_TableB3.R* creates Table B.3
- *Exhibit_TableB4.R* calculates the estimates in Table B.4

References

Adnan, Wifag, Jonathan Zhang, and Angela Zheng, "Intergenerational Mobility of Immigrants by Refugee Status: An Analysis of Linked Landing Files and Tax Records," 2025.

Connolly, Marie, Catherine Haeck, and Anne Mei Le Bourdais-Coffey, "Age at Immigration and the Intergenerational Income Mobility of the 1.5 Generation," Technical Report 2023.

Corak, Miles, "The Canadian geography of intergenerational income mobility," *The Economic Journal*, 2020, *130* (631), 2134–2174.

Immigration, Refugees and Citizenship Canada, "Facts and Figures 2003:Immigration Overview-Permanent & Temporary Residents - "Canada - Permanent Residents by Category. 1980 to 2003."," https://publications.gc.ca/collections/collection_2010/cic/Ci1-8-2003-eng.pdf 2003.

_ **and** _ , "Facts and Figures 2014:Immigration Overview-Permanent & Temporary Residents - "Canada - Permanent Residents by Category. 1989 to 2014."," https://publications.gc.ca/collections/collection_2015/cic/Ci1-8-9-2014-eng.pdf 2014.

Organization for Economic Cooperation and Development, "Population," data.oecd.org.

Statistics Canada, "Consumer Price Index, annual average, not seasonally adjusted," https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1810000501 2025.

The GDELT Project, "GDELT 1.0 Event Database," https://www.gdeltproject.org/data.html#rawdatafiles 2025.

UNHCR, "Forced displacement flow dataset," https://www.unhcr.org/refugee-statistics/insights/explainers/forcibly-displaced-flow-data.html 2025.