READ-ME

Permanent and Transitory Responses to Capital Gains Taxes: Evidence from a Lifetime Exemption in Canada

Review of Economics and Statistics

Adam M. Lavecchia & Alisa Tazhitdinova

February 2024

Overview

This replication package contains the Stata files that (i) construct the analysis datasets using the raw data files from the 1982-2016 years from the Longitudinal Administrative Databank (LAD) and (ii) conduct the analysis in the paper and the online appendices (including all figures and tables). This replication package is organized using three folders. The first folder, labelled “Do files" contains the Stata .do files that load the raw Stata files from the LAD, construct the analysis datasets and produce the analysis in the manuscript and Appendix. The second folder, labelled "Outputs" contains the regression output and summary statistics generated using the files in the first folder. The “Outputs" folder also contains other publicly available data on the Canadian Consumer Price Index, the Toronto Stock Exchange Index and Canadian bond yields that are used in the paper. The third folder, labelled "Figures and Tables" contains the tables and figures from the manuscript and online appendix in .pdf form. The analysis in the paper was produced using Stata inside of the Statistics Canada Research Data Centre (RDC) at McMaster University (Hamilton, Ontario).

Data Availability and Provenance Statements

This analysis in this paper uses the following datasets:

1. Restricted-access microdata files from the Longitudinal Administrative Databank (LAD) covering the years 1982-2016 (inclusive) (Statistics Canada, 2021a). This data is accessible through the network of Statistics Canada Research Data Centres (RDCs). To obtain access to the data, prospective users must submit a proposal through the Statistics Canada Microdata Access Portal: https://www.statcan.gc.ca/rdc-cdr/eng/user/login. Information about the RDCs can be found here: https://www.statcan.gc.ca/en/microdata/data-centres.

2. Public use data on the Consumer Price Index (all items or Total CPI) from Statistics Canada, accessible using this link: https://www150.statcan.gc.ca/t1/tbl1/en/cv.action?pid=1810000501 (Statistics Canada, 2021b). The analysis in this paper uses the annual average (not seasonally adjusted) values of the CPI for all items. We divide all values of the annual CPI by the 2016 value.

3. Monthly data for yield of 10-year bonds for the Government of Canada from the Federal Reserve Bank of St. Louis. The data is accessible using this link:
https://fred.stlouisfed.org/series/IRLTLT01CAM156N.

4. Monthly data on the closing values of the Toronto Stock Exchange Index from Statistics Canada. The data is accessible using this link: https://www150.statcan.gc.ca/t1/tbl1/en/tv.action?pid=1010012501.

5. Figure 1 in the manuscript reports the effective marginal tax rates on capital gains income in select English-speaking countries from 1982-2016 (inclusive). The tax rates for Australia are accessible at the following link: https://www.ato.gov.au/tax-rates-and-codes/tax-rates-australian-residents. The tax rates for the United Kingdom are accessible at the following link: https://webarchive.nationalarchives.gov.uk/ukgwa/+/ http://www.hmrc.gov.uk/stats/tax$_s$*tructure/menu.htm*. The tax rates for the United States are accessible at the following link: https://www.taxpolicycenter.org/statistics/historical-capital-gains-and-taxes.

Statement about Rights

X I certify that the author(s) of the manuscript have legitimate access to and permission to use the data used in this manuscript.
X I certify that the author(s) of the manuscript have documented permission to redistribute/publish the data contained within this replication package.

Summary of Availability

☐ All data **are** publicly available.
X Some data **cannot be made** publicly available.
☐ **No data can be made** publicly available.

Details on each Data Source

Data.Name	Data.Files	Source	Provided	Notes
"LAD file for YYYY year"	lad'YYYY'.dta	Statistics Canada	No	Confidential
"LAD Register"	reg2016.dta	Statistics Canada	No	Confidential
"CPI Dataset"	inflation.dta	Statistics Canada (2021b)	Yes	-
"GOC 10-Year Bonds"	Canada_10YrBond_IRLTLT01 CAM156N.xls	Statistics Canada (2021b)	Yes	-

Data.Name	Data.Files	Source	Provided	Notes
"TSE Index"	Canada_TSE_index_GSPTSE.xlsx	Statistics Canada (2021b)	Yes	-
"Country CG EMTRs"	Capital_gains_MTRs_Other_countries.xlsx	Statistics Canada (2021b)	Yes	-

Computational requirements

Software Requirements

- Stata (code was last run with version 17)
  - `outreg2` (as of 2023-04-20, Wada (2005))

The required files are included in the "${path}Programs/ado" subfolder.

Memory and Runtime Requirements

Summary Approximate time needed to reproduce the analyses on a standard (2023) desktop machine:

☐ <10 minutes
☐ 10-60 minutes
☐ 1-2 hours
☐ 2-8 hours
☐ 8-24 hours
X 1-3 days
☐ 3-14 days
☐ > 14 days
☐ Not feasible to run on a desktop machine, as described below.

The largest portion of the runtime comes from the construction of the analysis dataset using the raw files of the LAD. I suggest that prospective replicators use a computer with at least 64 GB of RAM to load and save the annual LAD files.

Description of programs/code

Analysis Using the LAD (Do files/RESTAT_Replication_Programs_Feb_14_2024/Final folder)

Since the annual LAD files are very large (and take a long time to load into Stata), I extract only a few variables needed at a time and later merge the smaller datasets as required.

- The program `0_Set_Directories.do` sets the directories for the various folders needed to run the programs, save the results etc.

- The program `1a_Extract_Capital_Income.do` extracts the capital income variables (notably realized capital gains) used in the analysis from the raw LAD files.
- The program `1b_Extract_MiniDemographics.do` extracts the time-varying demmographic variables (e.g. age, family size) used in the analysis from the raw LAD files.
- The program `1c_Extract_Demographics_permanent.do` extracts the permanent demographic variables (e.g. year of birth) used in the analysis from the raw LAD files.
- The program `1c_Extract_Demographics.do` extracts additional time-varying demographic variables (e.g. family size, postal code, census division) used in the analysis from the raw LAD files.
- The program `1d_Extract_age.do` extracts the age variable (in a calendar year) used in the analysis from the raw LAD files.
- The program `1e_Extract_weight.do` extracts the disturbance noise variable from the raw LAD files. This variable is used to protect the anonymity of tax filers in the LAD when constructing summary statistics.
- The program `1f_Extract_Death.do` extracts the year of death variable from the raw LAD files.
- The program `1g_Extract_TaxProvince.do` extracts the province of residence for tax filing purposes from the raw LAD files.
- The program `1h_Extract_SpousalCG.do` extracts the realized capital gains of spouses of tax filers from the raw LAD files.
- The program `1i_Extract_LossDeductions.do` extracts the interest and carrying charges variable and the non-capital loss varaible from the raw LAD files.
- The program `2_Organize_data.do` merges the smaller data sets created by the previous programs and generates the dependent variables of interest.
- The program `2c_Business_Owners.do` extracts self-employment income variables from the raw LAD files and identifies those with business income in the current tax year as well as the previous and future years.
- The program `2e_Calculate_Exemption_Left.do` calculates the remaining Lifetime Capital Gains Exemption (LCGE) room remaining available to a tax filer.
- The program `2f_CTACs_TaxRates_ForMerge.do` uses the CTaCS tax simulator to compute effective marginal tax rates on capital gains income by province, year and taxable income level.
- The program `2g_Calculate_TaxRates.do` merges the simulated effective tax rates on capital gains income to the analysis dataset.
- The program `2h_Regressions_Demographics_Prep.do` creates the covariates/control variables used in the analysis.
- The program `3a_Predicted2.do` calculates the cumulative capital gains realized by tax filers from 1985-1993 (which is subsequently used to define the treatment and control groups for the analysis).
- The program `3h_Baseline_Regression_Specifications_Set_Sample.do` creates the variables used in the baseline regression specifications.

- The program `4a_Regressions_DiffnDiff_Assemble.do` creates the variable used to determine whether a tax filer is assigned to the treatment or the control group.
- The program `4b_Baseline_Regression_Specifications_Set_Sample.do` sets the analysis sample for the regression (this makes the dataset smaller so that it runs more quickly).
- The program `RESTAT_MainText.do` runs the regression models to create the figures and tables in the main manuscript (i.e. Figures 2-9 and Table 2 in the manuscript).
- The program `RESTAT_OnlineAppendix.do` runs the regression models to create the figures and tables in the Online Appendix.
- The programs `DiffnDiff_Save.do` and `DiffnDiff_Save_'variable'.do` are called in the `RESTAT_MainText.do` and `RESTAT_OnlineAppendix.do` and save/store the regression estiamtes. The 'variable' suffix refers to one of 15 versions of this program that are used for various alternative dependent variables or regression specifications

Creating the Figures/Tables in the Paper from the RDC Output (Do files folder)

The analysis for our paper is done in Statistics Canada Research Data Centre (RDC). Output that is released from the RDC is vetted to ensure it complies with Statistics Canada's data confidentiality rules and procedures. One consequence of these procedures is that the vetting and release of regression model output is much easier than descriptive statistics. Furthermore, creating figures in the RDC is made difficult by the fact that vetting would require producing counts and various other statistics that support each point in a figure. As a result of this, we run the regression analysis in the RDC and then create the tables/figures that appear in the paper from the vetted/released output. The vetted/released output from the RDC is saved in the "Outputs" sub-folder in this replication package. The following programs produce the final figures and tables that appear in the manuscript and the Online Appendix using the output from the RDC.

- The program `0_Set_Directories.do` sets the directories for the various folders needed to run the programs, save the results etc.
- The program `3a_Graphs_MainText.do` produces the final version of the tables and figures that appear in the main manuscript (i.e. Figures 2-9 and Table 2 in the manuscript).
- The program `3d_Graphs_Appendix.do` produces the final version of the tables and figures that appear in the Online Appendix.

Instructions to Replicators

- Edit all programs in the `Do files/RESTAT_Replication_Programs_Feb_14_2024/Final` and `Do files` folders to adjust the paths for the raw data and the folders where the data, tables and figures are saved (based on the folder structure at your local Statistics Canada RDC).

- For the analysis using the LAD, replicators may be required to "uncomment" lines 14-26 in the
- Aside from running the programs in the replication folders in the order in which they are listed, researchers need to transfer the CPI data to their local RDC.

List of tables and programs

The provided code reproduces:

☐ All numbers provided in text in the paper
X All tables and figures in the paper
☐ Selected tables and figures in the paper, as explained and justified below.

Exhibit	Program	Line Number	Dataset
Table 2	Do files/3a_Graphs_MainText.do	21-90	LAD
Table 3	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do	298-479	LAD
Table A3	Do files/3a_Graphs_MainText.do	21-90	LAD
Figure 1	Do files/3a_Graphs_MainText.do	95-112	Country
Figure 2	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do & Do files/3a_Graphs_MainText.do	34-138 & 116-146	LAD
Figure 3	Do files/3a_Graphs_MainText.do	148-196	-
Figure 4	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do & Do files/3a_Graphs_MainText.do	141-201 & 200-231	LAD
Figure 5	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do & Do files/3a_Graphs_MainText.do	204-295 & 234-257	LAD
Figure 6	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do & Do files/3a_Graphs_MainText.do	298-479 & 264-320	LAD
Figure 7	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do & Do files/3a_Graphs_MainText.do	481-748 & 325-382	LAD
Figure 8	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do & Do files/3a_Graphs_MainText.do	750-1029 & 385-413	LAD
Figure 9	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_MainText.do & Do files/3a_Graphs_MainText.do	1033-1176 & 446-520	LAD
Table B1	Do files/3d_Graphs_Appendix.do	22-41	LAD
Figure E1	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		

	Do files/3d_Graphs_Appendix.do	50-114 & 47-145	LAD
Figure E2	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	117-187 & 47-145	LAD
Figure E3	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	191-337 & 150-209	LAD
Figure E4	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	340-422 & 212-270	LAD
Figure E5	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	426-444 & 274-453	LAD
Figure E6	Do files/3d_Graphs_Appendix.do	457-487	LAD
Figure F1	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	448-673 & 483-545	LAD
Figure F2	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	678-861 & 552-607	LAD
Figure F3	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	865-1071 & 610-670	LAD
Figure F4	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	1075-1268 & 672-730	LAD
Figure F5	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	1272-2104 & 732-888	LAD
Figure F6	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	2107-2403 & 892-949	LAD
Figure F7	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	2407-2617 & 954-1013	LAD
Figure F8	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	2623-2913 & 1022-1074	LAD
Figure F9	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	2918-3105 & 1080-1133	LAD
Figure F10	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	3108-3373 & 1140-1196	LAD
Figure F11	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		

	Do files/3d_Graphs_Appendix.do	3376-3548 & 1200-1259	LAD
Figure F12	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	3551-3904 & 1261-1311	LAD
Figure F13	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	3908-4075 & 1314-1381	LAD
Figure G1	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	4485-4590 & 1385-1505	LAD
Figure G2	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	4594-4612 & 1510-1633	LAD
Figure G3	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	4615-4616 & 1385-1633	LAD
Figure H1	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	4619-4702 & 1635-1689	LAD
Figure H2	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	4705-4917 & 1692-1726	LAD
Figure H3	Do files/RESTAT_Replication_Programs_Feb_14_2024 /Final/RESTAT_OnlineAppendix.do &		
	Do files/3d_Graphs_Appendix.do	4920-5026 & 1732-1764	LAD

References

Statistics Canada. (2017). Survey of Financial Security (SFS). Restricted access microdata. Reference periods 1999, 2005, 2012 and 2016.

Statistics Canada. (2021a). Longitudinal Administrative Databank. Restricted access microdata. Annual files, 1999-2018.

Statistics Canada. (2021b). *Table 18-10-0005-01. Consumer Price Index, annual average, not seasonally adjusted.* (Last Accessed: February 2021).

Wada, Roy. (2005). OUTREG2: Stata module to arrange regression outputs into an illustrative table, Statistics Software Components S456416, Boston College Department of Economics, revised 17 August 2014.

Acknowledgements

The content on this page was copied or adapted from https://social-science-data-editors.github.io/template_README/.