

Description of Underlying Data, Sampling, and Variables for Computing Statistics on Canadian Inequality, Volatility, and Mobility: 1983–2016

May 14, 2021

The statistics available on this webpage were calculated using the Canadian Employer–Employee Dynamics Database (CEEDD).¹ Many, but not all, of the statistics appear in [Bowlus et al. \(2021\)](#), or its online appendix (see this article for additional details).

The individual-level data used to compute all statistics are drawn from the T1 Personal Master File (T1PMF), which contains annual personal income tax records for all Canadian tax filers who filed their tax returns before a specified cut-off date. While the exact cut-off date varies over time, it is usually sometime in December one year after the tax reference year ([Messacar, 2017](#)). Only about 3.5–4.8% of all tax filers do not file tax returns before this date.² These late filers are not included in T1PMF.³ The statistics available on this webpage currently cover the period of 1983 to 2016. Over this period, the T1PMF includes records each year for 90–94% of all 25- to 55-year-old Canadians.

Sample selection and earnings measures

This section briefly describes the sample and variables used to generate all reported statistics on earnings inequality, volatility, and mobility.

The underlying T1PMF sample contains all Canadian residents ages 25–55 and uses annual individual earnings as measured by the sum of employment income (wages, salaries, and commissions) reported on T4 slips from all jobs and other taxable receipts from employment (e.g., tips, gratuities, or director’s fees) that

¹The CEEDD is a linkable environment developed by Statistics Canada that consists of several administrative and tax files, including individual tax files (T1 General – Income Tax and Benefit Return), individual employment remuneration files (T4 Statement of Remuneration Paid), individual records of employment (Record of Employment (ROE)), the Longitudinal Immigration Database (IMDB), corporation tax files (T2 Corporation Income Tax Return), and unincorporated business tax files (T1 Business Declarations). Altogether, it provides rich information on individual demographics, employment, job mobility, self-employment and entrepreneurship, and firm characteristics.

²The cut-off date before which a tax return may be included in T1PMF differs from the date by which an individual must submit a return to the Canada Revenue Agency (CRA) or after which interest charges or penalties begin to accrue on outstanding tax balances owed. The latter date is typically April 30 in Canada.

³While [Messacar \(2017\)](#) finds that late tax filers tend to be more prevalent among young individuals, non-residents, emigrants, very low earners, and those with final tax balances close to zero, this has little effect on estimated earnings distributions, including top-earnings percentiles. [Messacar \(2017\)](#) also finds that the likelihood of delayed tax filing is only weakly correlated with changes in individual circumstances (e.g., changes in employment earnings or receiving unemployment benefits).

are not reported on T4 slips. Self-employment income is not considered. All earnings from 1983 to 2016 are denominated in 2018 Canadian dollars using the Canadian Consumer Price Index. Denote these earnings for individual i in year t by $\tilde{y}_{i,t}$.

Not every individual appears in the data every year. Individuals may file late, or they may not file taxes at all during a year in which they have little or no income. Some individuals may be out of the country. Due to high rates of immigration to and emigration out of Canada, the sample omits (or leaves as missing) any earnings measures that are missing before the first year someone is observed filing their taxes or any earnings measures after the last year they are observed filing; otherwise, any missing earnings observations (i.e., in-between years with non-missing earnings) are treated as zeros (i.e., set $\tilde{y}_{i,t} = 0$).⁴ The latter implicitly assumes that missing observations in the middle of someone's time series of tax filings reflect years with no earnings. Although this could be problematic for individuals who simply filed later than the T1PMF cut-off date, these and other very low earnings measures are trimmed in nearly all of the analysis to minimize the influence of temporary employment and weak labor market attachment.⁵

Define a minimum earnings threshold in year t , y_t , as the amount a worker could earn by working part time (20 hours per week) for a quarter of the year (13 weeks) at the real minimum wage for that year. Because Canada does not have a federal minimum wage, the imposed minimum wage is assumed to be the lowest among all provincial minimum wages at the time.⁶ The measure of annual earnings used in the analysis, $y_{i,t}$, equals $\tilde{y}_{i,t}$ if $\tilde{y}_{i,t} > y_t$; otherwise, it is set to missing. Much of the analysis uses the natural logarithm of this annual earnings measure, $\log(y_{i,t})$. Residualized log earnings, $\varepsilon_{i,t}$, are also calculated from regressions of $\log(y_{i,t})$ on annual age dummies separately by gender and year, thereby removing average lifecycle effects on earnings. (Note that these residuals are also missing whenever earnings are below the threshold.)

Some of the analysis focuses on (residual) log earnings growth based on $\varepsilon_{i,t}$. To that end, annual and 5-year earnings growth measures are constructed: $\Delta^1 \varepsilon_{i,t} \equiv \varepsilon_{i,t+1} - \varepsilon_{i,t}$ and $\Delta^5 \varepsilon_{i,t} \equiv \varepsilon_{i,t+2} - \varepsilon_{i,t-3}$, where the latter is roughly centered on year t .⁷

Finally, two measures of “permanent” earnings (actually 3-year averages) are calculated to minimize the effects of year-to-year fluctuations. The main measure of permanent earnings is designed to focus on the distribution and dynamics of earnings for individuals with moderate to strong labor market attachment. An alternative measure is also considered for an analysis of long-term earnings mobility that incorporates workers who may be less attached, including those with very low or zero earnings for multiple years. In both

⁴This approach implicitly assumes that individuals are out of the country prior to their first observed filing year or after their last observed filing year.

⁵By itself, the exclusion of late filers has little effect on earnings distributions (Messacar, 2017). The analysis of annual earnings measures trims very low earnings observations, so setting missing observations to zero has no effect on most of statistics. Treatment of missing observations only affects the calculation of “permanent” earnings measures, which are primarily used in our analysis of earnings mobility. Alternative treatments of missing observations yield nearly identical mobility patterns for permanent earnings.

⁶See the Appendix of Bowlus et al. (2021) for minimum wages for each Canadian province over time.

⁷When calculating these earnings growth measures, earnings in the earlier period must exceed the earnings threshold for that year, while earnings in the later year must only exceed one-third of that year's threshold. More precisely, the later residuals ($\varepsilon_{i,t+1}$ or $\varepsilon_{i,t+2}$) are estimated using the sample for which earnings exceed one-third the threshold for that year (i.e., $\tilde{y}_{i,t+1} > y_{t+1}/3$). Thus, two sets of residuals are actually calculated for every year based on the two different restrictions on minimum earnings.

cases, $P_{i,t}$ is first calculated as the average of non-missing earnings over 3 years ($\tilde{y}_{i,t}$, $\tilde{y}_{i,t-1}$, and $\tilde{y}_{i,t-2}$). The main permanent earnings measure reflects this average whenever (i) $\tilde{y}_{i,t}$ is non-missing (i.e., all years from first to last observed filing) and (ii) at least two of the earnings measures are above the earnings threshold for their respective years; otherwise, this measure is set to missing. The second restriction ensures that this measure is always positive. Inequality and volatility statistics are based on *residualized permanent earnings*, $\varepsilon_{i,t}^P$, after removing lifecycle effects via regressions of $\log(P_{i,t})$ on annual age dummies separately by gender and year. The analysis of long-term earnings mobility considers an alternative permanent earnings measure that also begins with $P_{i,t}$, but this measure requires only 1 of the 3 earnings measures to be above the minimum earnings threshold. Further, it is not adjusted for age, gender, or year effects. Let $\tilde{P}_{i,t}$ denote this *alternative permanent earnings* measure.

References

- Bowlus, Audra, Émilien Gouin-Bonenfant, Huju Liu, Lance Lochner, and Youngmin Park. 2021. “Four Decades of Canadian Earnings Inequality and Dynamics across Workers and Firms.” CHCP Working Paper 2021–1.
- Messacar, Derek. 2017. “Big tax data and economic analysis: Effects of personal income tax reassessments and delayed tax filing.” *Canadian Public Policy* 43 (3):261–283.