

"Are Census-Based Estimates of Immigrants Earnings Growth Biased Upwards? Comparisons with a linked longitudinal data set".

By Garnett Picot and Patrizio Piraino

Description

We use earnings data from both a linked longitudinal administrative file and the census to assess the potential bias in estimates of immigrants' earnings growth based on repeated cross-sectional data. Providing reliable estimates of the immigrants' earnings trajectory is important because the earnings trajectory (with years since immigration) has shifted downwards for each successive immigrant entering cohorts since the 1980s. This disquieting trend has captured the attention of researchers and policy analysts alike.

Longitudinal data are best suited to study the earnings growth of immigrants in successive entering cohorts. However, such work requires very large sample sizes to allow for cohort effects, and information on a large number of covariates to control for differences between the immigrant and Canadian born population. By necessity, most researchers have turned to census data and estimated both conditional and unconditional earnings trajectories for various entering immigrant cohorts based on repeated cross-sections. But the sample in this quasi longitudinal cohort panel changes over time since many immigrant exit Canada.

Aydemir and Robinson (2006), focusing on young male immigrants, estimated that about one-third leave Canada during the first twenty years, with more than half doing so in the first year. This may introduce a bias in the earnings trajectories estimates based on cross-sectional data. If, for example, those who exit tend to have poorer labour market outcomes, and lower earnings than those who stay, then the earnings trajectory based on quasi-longitudinal cross-sectional data will be biased upwards. Hence, much of the progress in earnings (with years since migration) may result from a form of sample selection bias, not real increases in earnings.

Lubotsky (2007) found exactly this result in the U.S. data. The immigrant-native born earnings gap closed only one-half as fast in the true longitudinal data as in the repeated cross sections from the decennial census.

We ask whether a similar bias is observed in the Canadian results. We use longitudinal data based on annual individual taxation returns that links individuals returns over time. These returns are further linked to immigrant landing records to both identify immigrants in the larger file, and obtain the personal characteristics of immigrants. We have a large representative sample of all Canadian workers, immigrants and Canadian-born alike. From this we can estimate the absolute and relative (to Canadian born) earnings trajectories (growth) of immigrants from various entering cohorts since the early 1980s. These data allow us to estimate such trajectories, and the change in the immigrant-native wage gap, based on both true longitudinal as well as representative repeated cross

sections from the same data source. Hence, none of the difference in the results between the repeated cross-sectional and longitudinal formats can stem from differences in the data sources related to collection modes and procedures. This is particularly important if comparing results from administrative (here taxation) data with survey data (the census).

In addition, we use repeated cross sections from census data to estimate conditional and unconditional earnings trajectories (growth), and the change in the immigrant-native wage gap with years since migration. We then compare the earnings trajectories and the change in the immigrant-native wage gap from the longitudinal and cross-sectional data sources to assess if a bias is evident. Preliminary results suggest that the bias is much less evident in the Canadian than American census results, suggesting that immigrants leaving Canada have earnings similar to those who stay.