Canadian Labour Market and Skills Researcher Network

Working Paper No. 69

Immigrant Earnings Growth: Selection Bias or Real Progress?

Garnett Picot
Statistics Canada

Patrizio Piraino
Statistics Canada

December 2010

Immigrant Earnings Growth:
Selection Bias or Real Progress?

Garnett Picot and Patrizio Piraino

Statistics Canada

ABSTRACT

We use longitudinal tax data linked to immigrant landing records to estimate the earnings growth of immigrants from three entering cohorts since the early 1980s. Selective attrition by low-earning immigrants might result in lower earnings growth with years since migration in longitudinal data compared to repeated cross-sections. Existing studies on U.S. data have found exactly this result (Lubotsky 2007, JPE). We ask whether a similar bias is observed in the Canadian data and find that it is not. We show that while low-earnings immigrants are more likely to leave the cross-sectional samples over time, the same is true for the Canadian born population. We conclude that there is no evidence of selective labour force participation patterns among immigrants in Canada compared to the native born population.

Executive Summary

The decline in both entry earnings of immigrants, and their earnings trajectory after entering the Canadian labour market, have been among the most studied topics in immigration research during the past twenty years. Ideally such research would be based on longitudinal data, tracking immigrants and their earnings after they enter Canada. Often, however, because of the richness of the immigration data in the census, and because of the lack of a readily available longitudinal alternative, researchers turn to repeated cross-sections from the census to construct quasi-longitudinal panels. Immigrants entering Canada between, say, 1976 and 1980, are in Canada for between 1 and 5 years in the 1981 census. In the 1986 census, this same entering cohort is observed in Canada after 6 to 10 years, and in the 1991 census, between 11 to 15 years, and so on. These repeated cross-sections can be used to estimate the earning growth, and the earnings gap with the Canadian born for, in this case, the late 1970s entering cohort of immigrants

 Unfortunately, the sample in such quasi-longitudinal panels changes over time. Since these are repeated cross-sections based on a 20% sample of the population, and since many immigrants exit the country each year, the immigrants in the sample after, say, 11 to 15 years may be different from those in the sample during the first 1 to 5 years. More importantly, if the probability of exiting the sample is greater among immigrants who are struggling with low wages in the labour market, and lower among the "successful" immigrants, then an upward bias in the earnings trajectory will result. Through time the cohort will increasingly consist of "successful" immigrants, with higher earnings. The result would be a form of sample selection bias, producing an upward bias in the immigrant earnings pattern, and an increasingly underestimated earnings gap between immigrants and the Canadian born population. An American study found such a bias in the U.S. census based research.

  We use the "Longitudinal Administrative Database (LAD), a true longitudinal source derived from administrative data to determine if such a bias exists in the research based on Canadian census data. We focus on two outcome variables: the earnings growth of immigrants during the first twenty years after entering Canada, and the change in the immigrant-Canadian born earnings gap during the same period. Both could suffer from the sample selection bias mentioned earlier. Using the standard economic integration econometric models, we estimate the change, with years in Canada, in these outcome variables using three different data sources: (1) a quasi-longitudinal data set based on repeated cross-sections from the census, (2) the LAD, a true longitudinal data set based on administrative data, and (3) a quasi longitudinal data set constructed from repeated cross-sections from

the *same* administrative data used in (2), that is the LAD. This last point is important. The fact that we can obtain both cross-sectional and longitudinal results from the same data source eliminates differences in the estimates that may stem from variation in the collection modes and procedures across data sets. We then compare the immigrant earnings trajectories and the change in the immigrant-Canadian born earnings gap from the cross-sectional quasi-longitudinal data, and true longitudinal data, to determine if there is any evidence of a bias.

Our analysis provides little evidence of a significant bias in the immigrant-native born earnings *gap* trajectory computed from repeated cross sections as compared to true longitudinal data. Most earlier research focused on this earnings gap. Although the less successful and lower paid immigrants in the various cohorts are more likely to exit the sample, the same appears to be true for the native born. That is, the earnings growth of both the immigrant and Canadian born cohorts is over-estimated in cross-sectional data, by roughly the same degree. Hence, the "gap" trajectory obtained by estimating the standard assimilation model on longitudinal data points to little bias in previous studies of earnings assimilation in Canada based on census data. This is in contrast with the existing evidence from the United States, although the bias in the U.S case was only observed in one out of three cohorts studied. We do find evidence of an upward bias in the *earnings trajectory* (as opposed to the earnings gap) of immigrants based on repeated census cross-sections.