Session Number: Parallel Session 2B
Time: Monday, August 23, PM

Paper Prepared for the 31st General Conference of
The International Association for Research in Income and Wealth

St. Gallen, Switzerland, August 22-28, 2010

Immigrant Earnings Growth: Selection Bias or Real Progress?

Garnett Picot and Patrizio Piraino

For additional information please contact:

Name: Garnett Picot
Affiliation: Statistics Canada

Email Address: picogar@StatCan.ca

PRELIMINARY AND INCOMPLETE DRAFT
NOT FOR CIRCULATION OR CITATION
WITHOUT AUTHORS' PERMISSION

Immigrant Earnings Growth: Selection Bias or Real Progress?

Garnett Picot and Patrizio Piraino

Statistics Canada

July 2010

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.

JEL Classifications: J31; J61
Keywords: Immigration, assimilation, longitudinal data, selection bias.

1. Introduction

Entering immigrants have always earned less than the native born during their first few years in the host country. However, their relative earnings rise with years since migration, as they obtain host country experience, acquire useful language skills and learn the local labour market customs (Chiswick, 1978; Meng, 1987; Borjas, 1999). In Canada, immigrants entering during the late 1970s earned 85% of that of their native born counterparts during the first five years in the host country, and after 11 to 15 years in Canada they earned 92% of that of the native

born. The comparable numbers for the early 1990s entering cohort were 60% at time of entry, and 78% after 11 to 15 years in Canada (Frenette and Morrisette, 2003).[1]

The study of the earnings trajectory of immigrants in successive entering cohorts would ideally be based on longitudinal data. 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 the native born population. While some recent studies have used longitudinal administrative data (Hu, 1999; Edin et al, 2000; Duleep and Dowan, 2002; Green and Worswick, 2004; Lubotsky, 2007; Aydemir and Robinson, 2008), most existing research on entry earnings and the earnings trajectories of immigrants is based on census data. This is because longitudinal data have only recently become available, and they contain relative few covariates of interest. Notably, one cannot control for educational differences between immigrants and the native born.

Typically, researchers turn to repeated cross-sections from the Census to form a pseudo-longitudinal panel of data. For example, immigrants to Canada entering during the 1991 to 1995 period will be captured in the 1996 census following 1 to 5 years since arrival. Immigrants in this cohort who remain in Canada will be captured in the 2001 census after 6 to 10 years in the host country, and in the 2006 census after 11 to 15 years, and so on. On such a basis, both the earnings growth and the change in the immigrant-native born wage gap for various entering immigrant cohorts have been estimated.

However, the samples in such pseudo-longitudinal cohort panels change over time, as many immigrants exit the host country. For Canada, Aydemir and Robinson (2008) focused on young male immigrants, a very mobile group, and estimated that about one-third leave during the first twenty years, with more than half doing so in the first year in the host country. Exit rates among recent immigrant cohorts as a whole will no doubt be lower, but likely still substantial. This may introduce a bias in the earnings trajectories estimated from cross section data. If, for example, those who exit are more likely to have poorer labour market outcomes than those who stay (and hence have an incentive to leave), then the earnings trajectory based on pseudo-longitudinal cross section data will be biased upwards. As more time passes, any pseudo-panel cohort will increasingly consist of "successful" immigrants from the original cohort, those with higher earnings. Hence, much of the progress in earnings (with years since migration) may result

[1] More precisely, these numbers represent log earnings ratios (immigrant earnings to the Canadian born). A number of studies have looked at the decline in relative entry earnings for successive entering cohorts of immigrants in Canada (Bloom and Gunderson, 1991; Abbott and Beach 1993; Mcdonald and Worswick, 1998; Baker and Benjamin, 1994, and Grant, 1999). More recently, researchers have focused on the causes of the decline in entry earnings (see Picot and Sweetman, 2005 for a review).

from a change in the composition of the cohort over time, a form of sample selection bias, not a real increases in earnings.

This is exactly the result found by Hu (2000) and Lubotsky (2007) in the United States. Longitudinal earnings data showed that 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 U.S. census. Lubotsky concludes that the higher probability of out-migration by low-wage immigrants systematically led past researchers to overestimate the wage progress of immigrants remaining in the U.S. These findings paint a less optimistic picture of the degree to which immigrants are able to assimilate into the U.S. labour market. In fact, contributors to the immigration policy debate in the United States often cite the Canadian experience as a system where high-skilled immigration is actively encouraged. Establishing whether immigrant earnings growth in Canada is overestimated, as it appears to be the case in the U.S., will thus help inform policymakers in both countries.

The data used in this study are described in detail below. Essentially, we use longitudinal data created by linking annual individual tax returns over time, which are in turn linked to the immigrant landing records to obtain the personal characteristics of immigrants. We have a large representative sample of all workers, immigrants and native born alike. We can estimate the earnings growth of immigrants, and the change with years since migration in the immigrant-native born wage gap for three entering cohorts since the early 1980s. These data allow us to estimate such trajectories based on both true longitudinal data, as well as representative repeated cross sections from the *same data source.* The fact that we can obtain both cross section and longitudinal results from the same data source is important. It eliminates differences in the estimates that may stem from variation in collection modes and procedures across datasets. This is particularly relevant if comparing results from administrative (here tax returns) sources with survey data (the census). In order to more closely relate our results to the existing literature, we also estimate immigrant earnings trajectories with years since migration using repeated cross sections from census data.

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. 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 extent. 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. This is in sharp contrast with the

existing evidence from the United States, suggesting the potential role played by differing labour market institutions and immigration policies in the two countries.

The rest of the paper proceeds as follows. Section 2 explains the nature of the potential bias in immigrant earnings growth and provides a review of the small empirical literature on the issue. Section 3 presents the administrative database used in this study and describes the analytical advantages it offers, as well as its shortcomings. The empirical findings are presented and discussed in Section 4 and 5. Section 6 concludes.

2. The Issue: bias in immigrant earnings growth from repeated cross-sections

The main goal of this study is to assess the bias in cross-section estimates of immigrants' earnings trajectories. Table 1 explains the difference in the measurement of immigrant earnings growth in longitudinal and repeated cross-sectional data, as it applies to the Canadian census waves. The rows of the table indicate the year of arrival for three selected cohorts of immigrants (1985, 1990, 1995), while the columns show the year in which earnings are measured. In each cell, E(w) represents the average earnings measured at time c (column) for the cohort of immigrant who arrived in Canada at time r (row).

Panel A clarifies that the cross-sectional samples will lead to estimates of immigrant earnings growth that will depend on the nature of immigrant exits from the labour force. For instance, the first row shows that immigrant earnings for the 1985 cohort will be measured in 1990 with the average earnings over the subset of immigrants who are still in Canada after five years. In 2005, the estimated average earnings for the same cohort will be conditional on still being in Canada after 20 years since migration. Hence, each immigrant will contribute to the estimated earnings growth rates of her/his arrival cohort for as long as she/he is in the data. If those who leave tend to be a non-random sub-sample of those who initially entered, a composition bias in the estimated earnings trajectory will occur.

In panel B (longitudinal data), we can restrict the sample to those immigrants who are captured in the latest year of data. This allows, for each immigrant cohort, the estimation of average earnings over the *same* subset of individuals in all years of observation. As a result, the immigrant earnings growth measured in the longitudinal sample will provide an unbiased estimate of the earnings growth among immigrants who remain in the sample until the latest year of data.[2] This is not the same as estimating the earnings growth of the entering cohort had they *all* stayed until 2005. The latter could be obtained from the

[2] This does not exclude the possibility that in some years they may be absent.

longitudinal data on those who remained only if we are willing to assume that out-migration is based on permanent attributes that are not related to immigrant earnings growth over time. We do not attempt this interpretation, as the focus of our paper is to test whether existing estimates of immigrant earnings growth obtained using repeated cross sections from the census are biased.[3]

Table 1
Measures of average immigrant earnings: longitudinal vs. cross-sectional data

	Year of observation			
	1990	1995	2000	2005
Year of arrival	A. Repeated Cross Section			
1985	E(w| 5 years)	E(w|10 years)	E(w|15 years)	E(w|20 years)
1990		E(w| 5 years)	E(w|10 years)	E(w|15 years)
1995			E(w| 5 years)	E(w|10 years)
	B. Longitudinal data			
1985	E(w|20 years)	E(w|20 years)	E(w|20 years)	E(w|20 years)
1990		E(w|15 years)	E(w|15 years)	E(w|15 years)
1995			E(w|10 years)	E(w|10 years)

Note that low-wage immigrants need not disproportionately emigrate *from* a country for the bias to arise. To assess the extent of the bias, data on emigration rates is not necessary. The major concern is disproportionate exit from employment (not the country) by low-earnings immigrants. Whether they leave the country or not is irrelevant. The sample of interest is the employed, and hence it is the exit (and re-entry) pattern from employment that is of concern. In fact, when estimating immigrant earnings assimilation using pooled Census waves, only observations with positive earnings in each cross-section are typically used. In order to assess whether a bias exists in the estimated immigrant earnings progress from repeated cross-sections, we will condition the immigrant cohorts in the longitudinal samples on being employed after a number of years since migration (but they need not be continuously employed). In Canada, estimating emigration rates is plagued with data difficulties, as in most other countries, but reliable longitudinal data on the dynamics of employment and earnings are available.

[3] Existing studies estimate the earnings trajectories of immigrants who stayed in Canada over their study periods. They do not estimate the trajectory of the entering cohorts, had they all stayed in Canada.

2.1 Empirical evidence from previous studies

There is a very small literature asking whether selective out-migration of immigrants results in a bias in cross-sectional estimates of immigrant economic assimilation. Based on U.S. data, Hu (2000) and more recently Lubotsky (2007) conclude that selective emigration results in an overestimation of the economic assimilation of immigrants.[4] In particular, Lubotsky uses longitudinal earnings data for the 1951 to 1997 period from Social Security records and shows that 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. As Lubotsky points out, however, this effect was not consistent across all entering cohorts, being more evident among the 1970-79 arriving cohort, and only marginally observed among the cohorts entering in the 1960s and 1980s.

Duleep and Regets (1997) and Duleep and Dowhan (2002) also perform longitudinal analyses of immigrant earnings using U.S. data. Their focus, however, is on relaxing a different assumption in cross-sectional analyses: the assumption that immigrant earnings profiles are stationary across cohorts. They document important intercohort variation in earnings growth, and find an inverse relationship between immigrants' entry earnings and earnings growth. On the other hand, Borjas (1999) shows that this correlation is positive when education is not held constant, and argues that declining entry wages are not compensated by steeper earnings profiles. While these papers recognize that out-migration can also affect estimates of immigrants' economic assimilation in cross-sectional studies, they do not provide empirical evidence on the issue.

Two papers utilize Canadian data to address this issue, although in a less direct manner than the U.S. research where the results from longitudinal and cross-sectional data are compared directly, as we do in this paper. Both Canadian papers are based on the Survey of Labour and Income Dynamics (SLID), a six year longitudinal panel of Canadian workers, in which immigrants can be identified. Hum and Simpson (2000), exploring earnings growth over the 1993 to 1997 period, find that even in the raw, unadjusted longitudinal data, little economic assimilation is observed among male immigrants. That is, there was little change in the immigrant-native born wage gap among males, as earnings growth was about the same for immigrant and Canadian-born men over the five year study period. Among women, an increase, rather than a decline, in the (unadjusted) wage gap was observed, as earnings growth was greater among the Canadian born than among immigrants. Employing a fixed effects model, they conclude that there is no evidence of economic assimilation (i.e. a closing of the

[4] A similar conclusion is reached by Edin, LaLonde and Aslund (2000) in their analysis of Swedish data.

wage gap) for foreign born males.[5] This is in contrast with virtually all existing Canadian studies based on repeated cross-sectional census data, which find significant economic assimilation among immigrants. Hum and Simpson conclude that their results "provide a warning that evidence from cross-sectional data, which may be prone to bias resulting from unobserved worker heterogeneity, should be interpreted cautiously."

In a more recent paper, Skuterud and Su (2009), pool four panels of the SLID collected between 1993 and 2004 in order to augment the longitudinal sample of immigrants and Canadian-born. Contrary to Hum and Simpson (2000), they find evidence of considerable economic assimilation of immigrants. More relevant to our discussion, Skuterud and Su also try to address the issue of a bias in immigrant wage assimilation. Since the panels in their data are quite short, they utilize a substantially different approach than the one used in this paper, or by Lubotsky (2007). They employ a fixed effects model to eliminate, to the extent possible, the effect of unobserved individual effects on both emigration and wage growth (i.e. the effects of selective out-migration on wage growth). They conclude that the fixed effects approach changes the estimates of wage growth relatively little and that it "does *not* imply substantially lower immigrant wage growth in longitudinal data as the US literature has tended to find (e.g. Lubotsky, 2007)." Their results are consistent with the notion that the nature of out-migration is different in Canada, and that we should not expect an upward bias in the existing cross-section estimates of economic assimilation of immigrants.

By taking advantage of higher quality administrative data, the present study can help shed light on these contrasting Canadian results. Moreover, given the longer nature of our panels we can focus on the effect of selective 'exits' and can compare the Canadian results to the findings from the U.S. research. We do this by adopting the same approach used in Hu (2000) and Lubotsky (2007), which consists of conditioning the samples of immigrants on reaching a certain level of years since migration and examine their earnings trajectories over this period.

3. Data

This study uses three data sources: the Longitudinal Administrative Databank (LAD), the Longitudinal Immigration Database (IMDB), and Census of population data. The LAD is a random, 20% subset of the T1 Family File (T1FF), which is a yearly cross-sectional file of all individual tax-filers and their families. Although one has to file an individual income tax return to be captured in the

[5] This finding is re-obtained in a successive study, which is based on the same dataset (Hum and Simpson, 2004).

T1FF (and hence the LAD), the population coverage is very high (around 95% for the working age population), because of tax rebate incentives which encourage individuals with no taxable incomes to file a return. Individuals in the LAD are selected randomly, based on a unique identification number generated from the Social Insurance Number (SIN) and are linked across years to create a longitudinal profile. The LAD is augmented each year with a sample of new tax filers so that it consists of approximately 20% of tax filers for every year. In addition to annual earnings in each year, the data contains information on individuals' date of birth and gender.[6]

The IMDB merges immigrant landing records with taxation records. The former provide information on immigrant characteristics, the latter provides detailed longitudinal information on employment earnings in particular. Given the near-universal coverage of tax files, this data source allows detailed tracking of earnings trajectories of entering cohorts of immigrants since the early 1980s up to 2005. In this paper, we utilize a linked LAD-IMDB data set. The linkage is possible due to an individual's unique longitudinal identifier. Until recently, it was not possible to identify immigrants in the LAD files, and hence potentially important immigrant research was precluded.[7]

Our empirical analysis will focus on three successive cohorts of immigrants: 1985-89, 1990-94, 1995-99. Since we are using earnings observations up to the year 2005, the three cohorts will differ in the time-span over which we will be able to analyze their earnings trajectories. That is, we can estimate their earnings growth up to twenty, fifteen and ten years after migration, respectively. We analyze immigrant earnings over time both in *absolute* terms and *relative* to the Canadian born (i.e. the immigrant-native born wage gap).[8]

Similar to previous studies, we focus only on men, in order to avoid complications from selective labour force participation. Immigrants are defined as foreign individuals who were 25–44 years of age at the time of arrival in Canada, as reported by their 'landing record'.[9] In order to generate earnings trajectories for

[6] The definition of earnings includes wages, salaries, and commissions, before deductions, as well as taxable receipts from employment other than wages, salaries and commissions (e.g. tips, gratuities, or director's fees). It excludes self employment income. More details on the dataset are available in Statistics Canada (2009).

[7] The possibility to link the LAD with IMDB files has supported some recent work on the economic assimilation of immigrants entering Canada (e.g. Picot and Hou, 2009).

[8] Note that we have more covariates available when estimating immigrant absolute earnings trajectories. In particular, we know the educational attainment at entry of immigrants—while we do not have such information for the Canadian born.

[9] Immigrants who arrived outside this age range, as well as temporary foreign workers are dropped from the analysis. The lower age limit is imposed because the labour market experience of very young immigrants is likely to be more similar to that of Canadian-born workers than to that of adult immigrants. The upper age limit serves to focus on immigrants with a potential of higher

the native born that match those of the immigrants, the Canadian comparison groups are formed from the same birth cohorts as the immigrants. Finally, the overall sample is restricted to person-year observations ages 25 to 64.

A nice feature of our data is that it allows us to create both a 'cross-sectional' and a 'longitudinal' sample from the same LAD-IMDB files. Because the data source is updated annually with new observations, the yearly files remain cross-sectionally representative. To obtain the cross-sectional sample, we pool selected yearly files—for comparability with Census waves we choose 1990, 1995, 2000, and 2005—and use all person-year observations with positive earnings.[10] This sample will be used to replicate the standard pseudo-longitudinal approach to the estimation of immigrant earnings growth.

The longitudinal sample uses annual earnings data for each entering cohort of immigrants and the respective comparison group in all available years. The crucial sample restriction is that individuals must appear in the latest year of data to be included in the longitudinal sample. This is defined as having positive earnings in that year. We believe this to be the appropriate definition if the goal is to assess the bias in cross-sectional estimates of immigrant earnings growth. In fact, estimates of immigrant earnings assimilation from Census pooled waves are based on a positive earnings restrictions in each cross-section used.

In order to relate our results to previous cross-section estimates, and to check the comparability between our administrative earnings data and the Census, we also draw a pseudo-longitudinal sample from the quinquennial Canadian Census of population. We use earnings information for the years 1990, 1995, 2000, and 2005. For consistency with the administrative sample, we include only males aged 25-64 and with positive earnings.[11]

Several features of our data offer advantages over previous studies of immigrants' earnings dynamics, in particular compared to the social security earnings records used in the United States (Lubotsky, 2007). First, as already mentioned above, our dataset does not result from a match of administrative records with survey data. This means that we need not worry about the potential bias from non-random matches. Moreover, we can compare the earnings trajectories of immigrants on longitudinal samples with repeated cross sections from the *same data source*. That is, we do not have to deal with comparability issues originating from the use of distinct datasets. A second advantage is that the earnings data employed here are not top-censored. Our estimates are therefore free

---

levels of years since migration. The sensitivity of the main results to this restriction is tested in the appendix.

[10] The actual exclusion rule is earnings>CAD$ 500. Robustness checks are performed on various thresholds with no effect on the paper's main findings.

[11] Also for the sake of consistency, we only consider immigrants who arrived in Canada as adults (25-44 years old).

from concerns related to top-coding of the sample and the associated changes in the earnings ceiling over time. Finally, the data used in this study enable us to differentiate the immigrant status of legally admitted foreign individuals. In particular, we can identify "landed immigrants" (i.e. foreign individuals who were in Canada as permanent residents) and differentiate them from temporary foreign workers.

The LAD-IMDB, however, also has its shortcomings. The most obvious one is that the longitudinal earnings data are available only for individuals filing a tax return (although this represents about 95% of the working age population). When no tax return is observed following a number of years of filing, it is not possible to determine whether this was the result of not being employed but resident in Canada, of leaving Canada, or of simply not reporting earnings (e.g. informal employment). For this reason, we are particularly cautious when interpreting our findings as evidence of specific out-migration patterns as opposed to dynamics in labour market participation. Also, while the LAD-IMDB file contains information on the educational attainment at entry, intended occupation, and other characteristics of entering immigrants, it does not contain such information for the native born. Hence, estimates of conditional immigrant- native born wage gaps may be hampered by this lack of information. We explain later that this shortcoming does not affect our analysis, however.

A peculiar feature of the dataset poses an additional problem when estimating the relative earnings of immigrants. The IMDB only identifies immigrants landed since the year 1980. Foreign individuals who arrived in Canada before 1980 are part of the yearly tax records but cannot be 'flagged' as immigrants. This presents problems in identifying the "comparison group" when estimating the *relative* earnings of immigrants. In any given cross-section of data for calendar year T, the "comparison group" will include not only the native born, but also immigrants who have been in the host country for a number of years greater than T – 1980. For instance, in year 1991, immigrants who have been in Canada for more than eleven years will be included in the "comparison" group, along with the Canadian born. While this does not affect our analysis of immigrant absolute earnings trajectories, it means our estimates of the immigrant-native earnings gap will include a comparison with some "long duration" immigrants as well.

In order to gauge the scale of this problem, and to assess the comparability of the LAD-IMDB with the Census, we compute the incidence of immigrants on the total number of observations in appendix table A1. To make the comparison possible, the native born group in the Census is 'augmented' with immigrants landed before 1980.[12] Column 1 and 2 show that the two data sources are largely

[12] This is only done to obtain the descriptive statistics in Table A1. In our empirical analysis that follows, the Census samples do not include immigrants landed before 1980.

consistent, with higher number of immigrants and higher incidence on the population in the two later cohorts. In Column 3, we use the information available in the Census to determine the share of the "comparison group" who are longer term immigrants as opposed to truly native born. For the year 1990-94 cohort, slightly above 3% of the comparison group are longer term immigrants (in this case, in Canada for sixteen years or more), and over 96% are Canadian born.[13] The same share is of course higher in the earlier cohort: not quite 6% of the "comparison group" for the 1985-89 cohort consisted of immigrants who arrived in Canada before 1980. For the latest cohort, the share of immigrants in the comparison group is negligible.

Given the relatively low shares of the comparison group who are longer term immigrants, and their closer economic resemblance to the native born compared to recent immigrants, we do not see this "contamination" issue as being particularly troublesome.[14] As well, the fact that the extent of the contamination varies from cohort to cohort does not concern us, since we are interested only in within cohort comparisons of the wage gap trajectories based on longitudinal and cross-sectional data. Finally, and perhaps most importantly, in our empirical analysis we can compare the *cross-sectional* results based on the LAD-IMDB samples with the estimates obtained from the Census. In the next section, we will show that the Census results, which are not affected by any 'contamination' issue, are consistent with the estimates from the administrative records.

4. Empirical Results

We start by providing some descriptive patterns using the raw data. Table 2 compares the level of both immigrant earnings and the immigrant-native earnings gap by years since migration based on three different samples, and for three different cohorts. The three samples are (i) the longitudinal sample from the LAD-IMDB data set, (ii) the cross-sectional sample from the LAD-IMDB data set, and (ii) the cross-sectional sample from the Census. In essence, Table 2 uses our data to 'fill in' the information outlined in Table 1 from section 2.

We first look at the differences between the longitudinal and cross-sectional samples from the LAD-IMDB data (top two panels in the table). We note that, for all cohorts, the immigrant earnings levels during the first few years

[13] Note that the Canadian born group also includes child migrants (age<18) who arrived in Canada before 1980.

[14] Longer term immigrants resemble the Canadian born in economic terms. For example, the low-income rate among immigrants in Canada for less than five years is 2.5 times that of the Canadian born, but among those in Canada for 11 to 15 years, it is only 1.6 times higher, and among those in Canada for 20 years or more, it is indistinguishable from that of the Canadian born (Picot and Hou, 2003).

in Canada tend to be higher in the longitudinal data. By the end of the study period (e.g. after 20 years in Canada for the 1985-89 cohort) earnings are identical in the two samples. This is by design, because the two samples are identical by this time, consisting of all immigrants who were still in Canada and employed after 20 years. However, since the mean earnings in the raw data tends to be somewhat less upon entry to Canada, and identical by the end of the period, the earnings growth is marginally steeper in the cross-sectional as compared to the longitudinal data. This is what one might have expected to see based on the discussion above. In terms of earnings *gaps*, however, there is little variation across the longitudinal and cross-sectional administrative samples. Some differences emerge for the two more recent cohorts, but these differences are far from important, and for the earliest cohort, 1985-89, it is not observed at all. These patterns anticipate the major finding in our econometric analysis below: while there appears to be some differences in the *absolute* earnings growth between the cross-sectional and the longitudinal samples, the earnings *gap* closes over time at a similar pace in the two samples.

The bottom panel of Table 2 reports the same statistics for the samples drawn from the Census. We note the similarity in the earnings trajectories between the Census samples and the corresponding cross-sections from the administrative records. Immigrant earnings growth, both absolute and relative, is the same in the two cross-sectional data sets for the 1990-94 cohort and very similar for the other two cohorts (the difference is between 2 to 4 log points). For example, for the 1985-89 cohort, the earnings growth over fifteen years was 42 log points (i.e. 10.46-10.04, or roughly 42%) in the cross-sectional LAD-IMDB data, and 38 log points in the census. The consistency across the two cross-sectional data sources in both the absolute and relative (to native born) immigrant earnings *trajectory* is reassuring. On the other hand, there appears to be some differences in the earnings *levels* across the two datasets. Earnings from the census tend to be higher then those from the LAD-IMDB cross-sectional data. This is consistent with Frenette, Green and Picot (2006) who find that income values from tax data tend to be lower than the Census.[15]

[15] Frenette et al. (2006) also documents that the difference between the Census and the tax records is more noticeable at the bottom of the income distribution. This explains why the immigrant earnings gap is smaller in the Census as compared to the LAD-IMDB, for all cohorts. If the differences between the two data sources were uniform across the earnings distribution, we would only see higher earnings levels in the Census, but no discrepancies in the gap.

Table 2
Average immigrant earnings: longitudinal vs. cross-sectional data

		Year earnings are measured			
		1990	1995	2000	2005
Cohort		Longitudinal data in LAD-IMDB			
	N	12,087	11,577	12,177	14,082
1985-89	Log earnings	10.12	10.27	10.51	10.46
	Earnings gap	-.3893	-.2721	-.1384	-.0705
	N		19,839	21,502	24,940
1990-94	Log earnings		9.90	10.32	10.31
	Earnings gap		-.5645	-.3079	-.2687
	N			21,294	24,925
1995-99	Log earnings			10.15	10.31
	Earnings gap			-.4035	-.2638
		Repeated Cross Section in LAD-IMDB			
	N	19,494	15,576	14,689	14,082
1985-89	Log earnings	10.04	10.18	10.40	10.46
	Earnings gap	-.3891	-.2892	-.1507	-.0705
	N		29,049	26,703	24,940
1990-94	Log earnings		9.80	10.20	10.31
	Earnings gap		-.5749	-.3403	-.2687
	N			28,116	24,925
1995-99	Log earnings			10.04	10.31
	Earnings gap			-.4228	-.2638
		CENSUS			
	N	20,746	17,625	16,645	16,493
1985-89	Log earnings	10.09	10.21	10.40	10.47
	Earnings gap	-.3232	-.2230	-.1228	-.0401
	N		28,483	26,326	25,354
1990-94	Log earnings		9.86	10.27	10.37
	Earnings gap		-.4841	-.2459	-.1848
	N			27,819	25,643
1995-99	Log earnings			10.13	10.38
	Earnings gap			-.3123	-.1719

Notes: Authors' calculations from LAD-IMDB and Census. The sample size N refers to immigrants only. In each year, the population consists of males 25-64 years of age with positive earnings. Immigrants migrated between 25-44 years of age.

4.1 Econometric estimates

The patterns above are based on the raw data, but most of the reported results in the literature stem from some form of regression model. We use a standard econometric framework to examine the absolute and relative economic performance of immigrants in longitudinal and repeated cross-sectional data. Rather than pool the data across cohorts, we prefer to study the earnings

trajectories of successive immigrant cohorts separately, because we do not want to impose a constant earnings growth across all cohorts. As previously noted, the evidence from the U.S. suggests that the wage progression of immigrants in the longitudinal data was not consistent across all entering cohorts. Much of the estimated 'out-migration' bias in Lubotsky (2007) seems to derive only from the 1970-79 cohort, and not from the other two cohorts examined. The results from the raw data above suggest there may be some cross-cohort differences in the Canadian data as well.

While most of the empirical literature focuses on the relative (to natives) earnings growth of immigrants, it is useful for a number of reasons to describe the trajectories in earnings levels of immigrants, in addition to the earnings gap, and assess how they differ in cross-sectional and longitudinal data.[16] Hence, we start with estimating the absolute earnings trajectories of entering immigrants, running a regression for each of the three cohorts separately. A simple way to capture these trends is to estimate the following regression for the entering cohorts:

$$w_{it} = \alpha + \beta_1 Age_{it} + \theta ysm_{it} + \varepsilon_{it} \quad (1)$$

where w_{it} is the log of annual earnings for individual i in year t; Age_{it} is a polynomial in the individual's age and ysm_{it} is the number of years in the host country since arrival, which is specified as a categorical variable: 0 to 5, 6 to 10, 11 to 15, and 16 to 20 years since migration. In this immigrant-only regression, collinearity does not allow us to estimate period effects. Therefore, calendar time controls are not included. Table 3 reports the estimated coefficients for years since migration in model (1) for the three immigrant cohorts under analysis. Estimates are provided separately for the two LAD-IMDB samples, cross-sectional and longitudinal, as well as for the Census sample.

For all three cohorts, there is evidence that the earnings trajectory is overestimated in the cross-sectional as compared to the longitudinal data. For the 1985-89 cohort, Table 3 shows that the earnings growth between 0 to 5 and 16 to 20 years in Canada was about 27% in the longitudinal data, and 33% in the cross-sectioanl LAD-IMDB. That is, immigrant earnings growth after 16-20 years in Canada is about 6% less in the longitudinal sample, suggesting an upward bias in the cross sectional results. For the 1990-94 cohort, the earnings growth after 11 to 15 years in Canada is about 49% in the cross-sectional data, and only 39% in the longitudinal sample, a 10% difference. A bias is also observed for the latest cohort (21% vs. 27%). Note also that the estimates from the cross-sectional LAD-IMDB

[16] First, there are more covariates available for immigrants than the Canadian-born in the LAD-IMDB data, notably education level. Second, with an "immigrants only" sample we do not have the problem of the "comparison group" including some immigrants, and finally, knowledge of the absolute earnings growth of immigrants is in itself interesting.

sample are very much in line with those based on the Census. This confirms that our administrative records provide a reliable source of information on the earnings trajectory of immigrants in Canada.

Table 3
Immigrant's earnings growth in Canada: longitudinal vs. cross-sectional data*

	LAD-IMDB		
	Longitudinal (1)	Cross-sectional (2)	Census (3)
	1985-89 Cohort		
Years since migration			
6-10	.117	.065	.057
11-15	.271	.266	.237
16-20	.271	.333	.314
	1990-94 Cohort		
Years since migration			
6-10	.321	.371	.374
11-15	.391	.494	.498
	1995-99 Cohort		
Years since migration			
6-10	.216	.275	.267

Notes: Data from Census and LAD-IMDB files. Reference category is immigrants with 1 to 5 years since migration. All coefficients are statistically significant at 1% level.
* the table reports the coefficients on the "years since migration" dummy variables in model 1

Many standard regression models incorporate educational attainment, and hence we add educational attainment at time of entry to the model above, and report the results in Table 4. By controlling for education, we would eliminate part of the bias between the two data sources, if that bias is driven by the higher probability of exit by less educated immigrants. But this is not what we find: in terms of differences between the two data sources, the results reported in Table 4 are similar to those obtained from the unconditional regression. There is evidence of faster earnings growth in the cross sectional as compared to the longitudinal data, especially for the 1990-94 cohort.[17] This provides indirect evidence of a higher probability of exit by low-earning immigrants *within* education groups.

[17] We do not run the regression with controls for education on the Census sample as the education categories used in the LAD-IMDB do not match those reported in the Census.

Table 4

Immigrant's earnings growth in Canada with controls for education: longitudinal vs. cross-sectional data*

	LAD-IMDB	
	1. Longitudinal	2. Cross-sectional
	1985-89 Cohort	
Years since migration		
6-10	.143	.096
11-15	.320	.325
16-20	.335	.408
	1990-94 Cohort	
Years since migration		
6-10	.345	.399
11-15	.435	.536
	1995-99 Cohort	
Years since migration		
6-10	.243	.287

Notes: Data from LAD-IMDB files. Reference category is immigrants with 1 to 5 years since migration. All coefficients are statistically significant at 5% level.

*the table reports the coefficients on the "years since migration" dummy variables

From the estimates in Tables 3-4, we can infer that immigrants exiting the sample, among the three cohorts analysed, are more likely to have poorer labour market outcomes than those who stay. That is, changes over time in the composition of the repeated cross-sections due to selective exits between the lower and higher earners introduce a bias in the *absolute* earnings trajectories estimated on cross-sectional data.

Most of the earlier research on immigrant earnings growth, however, focuses on the change with time spent in the host country in the earnings *gap* between immigrants and native born, not on the earnings growth among immigrants alone. To test for a bias in those studies based on cross-sectional data, we introduce a comparison group, as described in the data section. In the Census data, the comparison group consists of Canadian-born males aged 25 to 64 with positive earnings. In the LAD-IMDB, the comparison group includes the same population, plus some small number of immigrants who have been in Canada for a number of years. To evaluate the immigrants' progress in earnings (with years since migration) relative to the native-born, we apply the standard empirical framework in this type of analyses (Chiswick, 1978; Borjas, 1999).

Consider the following regression model of log annual earnings:

$$w_{it} = \alpha + \beta_1 Age_{it} + \beta_2 Year_{it} + \lambda I_i + \gamma M_i \cdot I_i + \theta ysm_{it} \cdot I_i + \varepsilon_{it} \quad (2)$$

where the additional variables beyond those in equation (1) are a vector of calendar time dummies, $Year_{it}$, the immigrant's age at arrival in the host country, M_t, to proxy for foreign labour market experience; and a dummy variable identifying immigrant and native born status, I_t. Note that in model (2) the variables M_t (immigrant's age at arrival in the host country) and ysm_{it} (years since migrations) are now interacted with the immigrant status dummy, allowing the earnings trajectory over time to differ between immigrants and the native born. As before, the model is run separately on the three different cohorts.

The coefficient on years since migration, θ, is our parameter of interest and measures the change in the earnings gap with years spent in the host country, or put another way, the rate of earnings convergence over time between immigrants and native-born. The immigrant's earnings gap at the time of entry and the effect of foreign experience on earnings in the host country are captured by λ and γ, respectively. In order to separately identify the coefficients on the variables *Age*, *Year*, *M*, and *ysm*, we must impose the restriction that the age and period effects, β_1 and β_2, are the same for immigrants and native-born. As explained in Borjas (1999) and Lubotsky (2007), this assumption is not trouble-free. However, most existing estimates of immigrants' earnings growth are based on this standard assumption, and we choose to keep this restriction to focus on evaluating the difference in measured earnings growth between longitudinal and cross-sectional data. Moreover, this is the same model used in the U.S. study to which our estimates for Canada can be compared (Lubotsky, 2007).[18]

The first two columns in Table 5 report the results from the estimation of equation (2) based on the LAD-IMDB data, both on the longitudinal (column 1) and cross-sectional (column 2) samples. Column 3 reports the results based on the cross-sectional Census sample. We start by comparing the results based on the two LAD-IMDB samples, thereby eliminating any differences due to data sources (survey vs. administrative). For the 1985-89 cohort, there is no major differences to speak of, either in the entry earnings gap, or the change in the gap over time. In the two samples the gap at entry is around 33 to 35 percent, and after 16 to 20 years in Canada, it has been reduced by about 24 percentage points. For the 1990-

[18] As in Lubotsky (2007), individuals' educational attainment is not controlled for, since the objective is to test for a bias in the unconditional earnings trajectories of immigrants and natives. The only difference with Lubotsky's specification is in the age variable, where he uses instead a potential experience variable (age minus years since completion of schooling). We do not have education information for native-born, and cannot differentiate potential labour market experiences.

94 cohort, the gap at entry is marginally larger among the cross-sectional sample (as one might expect if there were a bias), but the difference is statistically insignificant and there are no significant differences between the samples in the change in the gap, which is our main interest. The same applies to the 1995-99 cohort. Overall, the results from our administrative data do not suggest the existence of a bias in repeated cross-sections estimates of earnings assimilation—contrary to what we observed for absolute earnings. A similar conclusion is reached when we compare the longitudinal results from the LAD-IMDB data to the cross-sectional Census results. Earnings convergence in the census sample slightly differs from that estimated on the longitudinal LAD-IMDB sample, but the differences are not great, nor are they in one particular direction.

Table 5.
Relative immigrants' earnings growth in Canada: longitudinal vs. cross-sectional data

	LAD-IMDB		
	1.Longitudinal	2. Cross-sectional	3. Census
	Change in the earnings gap		
	1985-89 Cohort		
Years since migration			
6-10	.121	.081	.094
11-15	.216	.192	.181
16-20	.247	.242	.239
Entry gap	-.348	-.328	-.334
	1990-94 Cohort		
Years since migration			
6-10	.212	.221	.236
11-15	.252	.265	.279
entry gap	-.447	-.465	-.444
	1995-99 Cohort		
Years since migration			
6-10	.151	.153	.141
entry gap	.267	.274	-.224

Notes: Data from Census and LAD-IMDB files. Reference category is immigrants with 1 to 5 years since migration. All coefficients are statistically significant at 5% level.

We performed a number of robustness checks on the results presented in this section. Changes in the age restriction rules do not alter our main findings

(see appendix Tables A2-A3). We also tested the sensitivity of the results to the earnings restriction threshold (0, 1, 100, 1000 dollars) and to trimming individuals at the top/bottom 1% of the distribution, but found no significant impact on the coefficients.

Taken together, our empirical results suggest that in Canada there is no evidence that estimates of immigrants' earnings *convergence* (with years since migration) to the earnings of the native born based on repeated cross-sections are significantly biased by selective attrition of immigrants over time, at least for the three immigrant cohorts we analyzed. On the other hand, there is evidence that estimates of the growth in the *absolute* earnings of immigrants, rather than the gap, based on repeated census cross-sections are in fact slightly biased upwards. This is likely due to higher probabilities of exit from the sample by lower earnings immigrants. To better understand why there is a bias in the absolute earnings trajectory, but not in the earnings gap trajectory, we take a closer look at the earnings characteristics of sample "exiters" and "stayers".

5. Who is leaving the sample?

We want to determine if people who were in the cross-sectional administrative sample but not in the longitudinal sample tend to be disproportionately low earners, rather than uniformly distributed throughout the distribution. We refer to these people as "exiters". Put another way, is the probability of exiting from the cross-sectional sample greater for low than high earners? If so, such an exit pattern would lead to an upward bias in the earnings trajectory based on cross-sectional data, of the kind we in fact observed above for absolute earnings growth, but did not observe for the earnings assimilation.

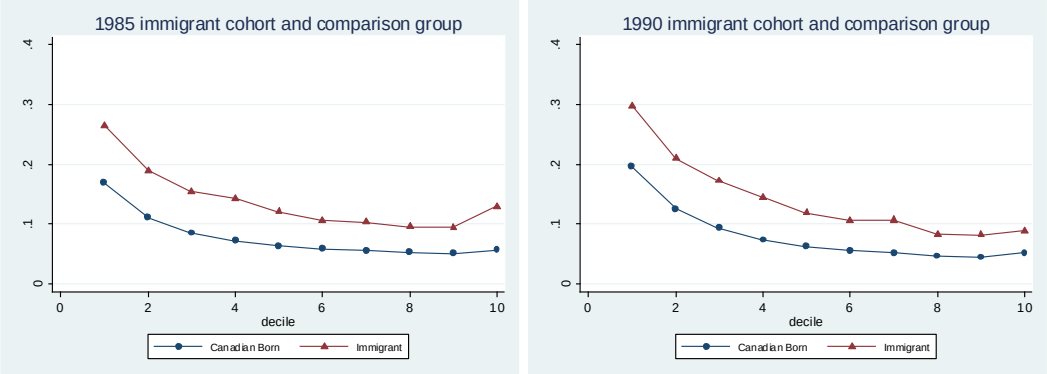
In order to answer this question, it is necessary to define the earnings distribution and where each exiter is located in the distribution in the appropriate manner. The main issue to be confronted is how to obtain the appropriate distribution for the entire immigrant cohort, given that some observations exit employment early in the study period, some later, and some are found in employment at both the beginning and the end of the period—i.e. the "stayers".

One could choose the earnings distribution that existed during the first few years in Canada. However, an immigrant's position in the distribution will likely change in later years with time spent in Canada, and such an approach could be misleading. A second option would be to compute mean earnings over the entire study period for each observation, and base the distribution on the individual means. However, earnings rise with years spent in Canada, and hence the earnings of early exiters will be significantly under-estimated relative to stayers and those who exit later, and their positioning the earnings distribution incorrectly

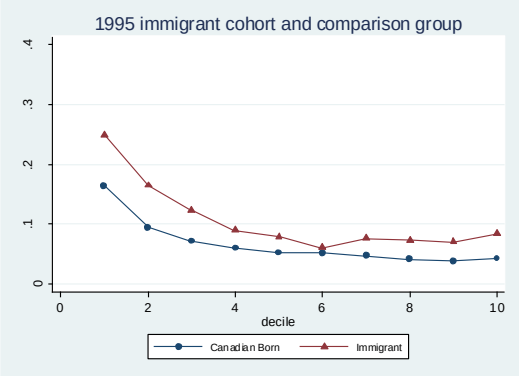
determined. What we want to know is how well the exiters did, economically, during the years they worked in Canada, relative to all other immigrants of the same cohort during *exactly* the same years in Canada. To achieve this, in any given year, T, we create a dummy variable 'exit' which equals one for all observations leaving the sample in T but who were in the sample in T – 1. We then calculate for each individual his *cumulative* earnings in year T as the sum of his earnings up to the year before (T – 1). This allows us to define a different earnings distribution for each year (e.g. for all years between 1986 and 2005 for the 1985 cohort) based on the cumulative earnings between 1986 and any given year, T. We can then assess the earnings level of each exiter relative to all other immigrants in his cohort over the exactly the same years—i.e. those that the exiter spent in the sample up to the year T, the year of exit. We apply this approach to the year of first exit. Some exiters may have exited, re-entered and then exited the sample, but we consider only the first exit.

We estimate a simple model from which we obtain the probability of exit, Pr(exit = 1), for each earnings decile. This is done by using a probit model where the dependent variable is the exit dummy as defined above. The independent variables include dummy variables for the 10 earnings deciles and a second degree polynomial in age. The age controls are designed to eliminate any effect of the age-earnings profile on the likelihood of exiters being in the lower earnings deciles.[19] This exercise is carried out using the LAD-IMDB data separately for three specific immigrant cohorts (1985, 1990, and 1995), as well as for their respective "comparison groups" (largely the Canadian born population of the same age group).[20]

Figure 1.
Probability of exit by earnings deciles



[19] Young people tend to have both lower earnings, and to be more mobile than older workers.
[20] We focus only on single-year cohorts in this section because it simplifies determining the cumulative earnings deciles. The findings generalize, however, to the other single-year cohorts.



Notes: Author's estimation from LAD-IMDB files.

Figure 1 shows the results from the estimation of the probit model. Each panel plots the probability of exit from the cross-sectional LAD-IMDB sample by earnings decile. We first notice that the probability of exit is always higher for immigrants than for native born. Some (or most) of this difference could be due to a higher probability of out-migration from Canada for immigrants, but our data do not allow us to differentiate the reason for exiting employment. When focusing on the immigrants only, we note a much greater probability of exiting the sample for immigrants in the bottom three or four deciles than for those in the top deciles. For the 1985 cohort, immigrants in the bottom decile had almost a 30% chance of exiting, compared to around 10% for those in the top half of the distribution. A similar pattern is observed for the other two cohorts. This is the pattern leading to the upward bias in the immigrant absolute earning trajectory we observed on repeated cross-sectional data.[21]

The degree of the bias for any cohort would depend upon the difference in the probability of exit between the low and high earners, and this difference varies across cohorts. For the 1985 cohort, the probability of exit in the bottom three deciles (at around 20%) is 1.9 times that for the top three deciles (at around 11%). In the 1995 cohort, the same ratio is at 2.2. But the differences are greater for the 1990 cohort, where the probability of exit was 2.7 times greater among low (bottom three deciles) than high earners. This is consistent with the greater upward bias in the *absolute* earnings trajectories based on cross-sectional data in the 1995 cohort, as compared to the other cohorts, that was observed in the previous section of the paper.

[21] An interesting pattern to note is the increase in the probability of exit at the very top of the earnings distribution for the immigrants, particularly so in the 1985 cohort data. This could be related to the phenomenon of high-skilled immigrants using Canada as a step-in stone to the North-American labour market.

But what of the fact that little bias was observed in the *relative* earnings trajectories in the cross-sectional data? Focusing on the "comparison" group, we observe virtually the same pattern: the probability of exit is significantly higher among workers at the bottom of the earnings distribution than among the higher paid. This explains why we observe little bias in the *relative* earnings trajectory. As we mentioned above, the probability of exit is lower among the comparison group than among the immigrant cohort at all points in the distribution. This is not surprising, as immigrants are more likely to separate from earnings than the Canadian-born, particularly during their early years in Canada when jobs are less stable. However, our primary concern is with the relative probability of low and high earners exiting. Here, we observe the same pattern among the comparison group as among immigrants.[22] The earnings growth of both the immigrant cohort and the "comparison group" estimated from cross-sectional data is biased upwards due to the higher likelihood of lower earnings individuals exiting the sample. The gap, which is the difference between these two trajectories, is then only marginally affected, as the bias appears to apply to both immigrants and the "comparison" group similarly.

6. Conclusions

The majority of studies of immigrants' earnings assimilation have been based on census data, using quasi-longitudinal cohorts constructed from repeated cross sections. But the population in these types of data changes over time, as some members of the immigrant cohort leave the country, or at least exit from the employed state. If there is a selection effect introduced, by the less successful and lower paid immigrants disproportionately exiting (as compared to higher earnings immigrants), then the earnings trajectories based on repeated cross-sectional data will be biased upwards. Some of the upward earnings growth would not be "real", but rather would be due to the changing composition of the immigrant cohort with years since migration.

We use a novel longitudinal administrative data base, as well as the census data, to determine whether such a bias exists. We focus on both the *absolute* earnings trajectories of entering immigrant cohorts, as well as the trajectories of the immigrant-Canadian born earnings *gap* over their first 20 years in Canada. Regarding the growth in absolute earnings of entering immigrant cohorts, we conclude that the results based on repeated cross-sectional do differ significantly

[22] Our results do not vary when using a logit or a linear probabillty model instead of the probit. The findings are also robust to an alternative definition of the cumulative earnings that does not count the earnings from the year prior to exit (to avoid potential issues deriving from part-year earnings data).

from the result based on true longitudinal data. In the three immigrant cohorts examined, the earnings growth (with years spent in Canada) based on a standard regression model was between 20 and 27 percent higher in the repeated cross-sectional sample as in the true longitudinal data. This bias was introduced because immigrants in these cohorts who found themselves at the bottom third of the earnings distribution were roughly 2.5 times as likely to exit the sample (i.e. exit employment and possibly the country) as those who were at the top third of the distribution.

However the change in the relative (to the native-born) earnings of immigrants with years since migration, which is the solely focus of most of this literature, is very similar in the cross-section and longitudinal data. We find little evidence of a significant upward bias in the earnings *gap* trajectory computed from repeated cross-sections as compared to true longitudinal data. Although the less successful and lower paid immigrants in the various cohorts are more likely to exit the sample, the same is true for the Canadian-born. That is, the earnings growth of both the immigrant and Canadian born samples is over-estimated in cross-sectional data, by roughly the same extent. Hence, the "gap" trajectory obtained by estimating the standard assimilation model on longitudinal data points to little bias in previous studies of earnings assimilation of immigrants to Canada.

A comparison of the results offered in this paper to the existing evidence from the United States (Lubotsky, 2007) suggests that immigrants to Canada display labour market participation patterns that are more similar to the native born than appears to be the case in the U.S.

7. References

Aydemir A. and C. Robinson (2008). Global Labour Markets, Return, and Onward Migration. *Canadian Journal of Economics*. 41: 1285-1311

Borjas, G. J., (1999). "The Economic Analysis of Immigration." In (D. Card and O. Ashenfelter, eds.), *Handbook of Labor Economics*, Vol. 3, Amsterdam: North-Holland.

Chiswick, B. R. (1978). "The effect of Americanization on the earnings of foreign-born men." *Journal of Political Economy*. 86: 897-921.

Duleep, Harriet Orcutt and Mark Regets (1997), Measuring immigrant wage growth using matched CPS ¯les," *Demography* 34(2): 239-49.

Duleep, Harriet Orcutt and Daniel Dowhan (2002), Insights from longitudinal data on the earnings growth of U.S. foreign-born men," *Demography* 39(3): 485-506.

Edin, Per-Anders, Robert Lalonde and Olof Aslund (2000), Emigration of immigrants and measures of immigrant assimilation: Evidence from Sweden," *Swedish Economic Policy Review* 7: 163-204.

Frenette, M., D. A. Green and G. Picot (2006). Rising Income Inequality in the 1990s: An Exploration of Three Data Sources. In (D. A. Green and J. R. Kesselman, eds.), *Dimensions of Inequality in Canada*, UBC press.

Frenette, Marc and René Morissette. 2003. "Will they ever converge? Earnings of immigrant and Canadian-born workers over the last two decades." Analytical Studies Research Paper Series. Catalogue No. 11F0019MIE2003215.

Green, D. A. and C. Worswick (2004). Immigrant Earnings Profiles in the Presence of Human Capital Investment: Measuring Cohort and Macro Effects. *mimeo*.

Hu, Wei-Yin (2000), Immigrant Earnings Assimilation: Estimates from Longitudinal Data," *American Economic Review Papers and Proceedings* 90(2): 368-72.

Hum, Derek and Wayne Simpson (2000), Closing the wage gap: Economic assimilation of Canadian immigrants reconsidered, *Journal of International Migrations and Integration* 1: 427-41.

Hum, Derek and Wayne Simpson (2004), Reinterpreting the performance of immigrant wages from panel data," *Empirical Economics* 29: 129-47.

Lubotsky D. (2007). Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings. *Journal of Political Economy*. 115: 820-867.

Meng, R. (1987). "The earnings of Canadian immigrant and native-born males." *Applied Economics*. 19: 1107-1119.

Picot, G. and F. Hou. (2003). The Rise in Low income Rates Among Immigrants in Canada. Statistics Canada Catalogue no. 11F0019MIE – No. 198. Analytical Studies Branch Research Paper Series.

Picot, G. and F. Hou (2009). Immigrant Characteristics, the IT Bust, and Their Effect on Entry Earnings of Immigrants. Statistics Canada Catalogue no. 11F0019M – No. 315. Analytical Studies Branch Research Paper Series.

Skuterud, M. and M. Su (2009). Immigrant Wage Assimilation and the Return to Foreign and Host-Country Sources of Human Capital. CLSRN Working Paper No. 30.

Statistics Canada (2009). Longitudinal Administrative Data Dictionary, 2006. Catalogue No 12-585-X.

8. Appendix

Table A1. Incidence of immigrants over time in the Census and LAD-IMDB

	Proportion of immigrants in the sample						Proportion of immigrants landed *before* 1980 in the native-born group[a]
	LAD-IMDB			Census			Census
	cohort: 1985-89						
Year	N. of obs	N. of immigrants	%	N. of obs	N. of immigrants	%	
1990	801,772	19,494	2.43	863,459	20,746	2.40	5.98
1995	748,578	15,576	2.08	803,699	17,625	2.19	5.72
2000	710,157	14,689	2.07	749,061	16,645	2.22	5.72
2005	663,986	14,082	2.12	697,100	16,493	2.37	5.56
	cohort: 1990-94						
Year	N. of obs	N. of immigrants	%	N. of obs	N. of immigrants	%	
1995	804,403	29,049	3.61	856,416	28,483	3.33	3.27
2000	776,789	26,703	3.44	815,563	26,326	3.23	3.31
2005	753,034	24,940	3.31	782,863	25,354	3.24	3.33
	cohort: 1995-99						
Year	N. of obs	N. of immigrants	%	N. of obs	N. of immigrants	%	
2000	780,890	28,116	3.60	813,354	27,819	3.42	1.28
2005	770,008	24,925	3.24	795,307	25,643	3.22	1.34

Notes: Authors' calculations from LAD-IMDB and Census. In each year, observations are restricted to 25-64 years of age and positive earnings. Immigrants migrated between 25-44 years of age.

[a] Native-born refers to the sample we use as control group in the LAD-IMDB files (which includes immigrants landed before 1980).

Table A2. Immigrant's earnings growth in Canada: longitudinal vs. cross-sectional data

	LAD-IMDB		
	Longitudinal (1)	Cross-sectional (2)	Census (3)
	1985-89 Cohort		
Years since migration			
6-10	.103	.056	.036
11-15	.251	.255	.236
16-20	.248	.330	.311
	1990-94 Cohort		
Years since migration			
6-10	.323	.360	.359
11-15	.399	.491	.481
	1995-99 Cohort		
Years since migration			
6-10	.206	.258	.233

Notes: Same specification as Table 3 in the text but different age at migration restriction (20 to 54). All coefficients are statistically significant at 1% level.

Table A3. Relative immigrants' earnings growth in Canada: longitudinal vs. cross-sectional data

	LAD-IMDB		
	1.Longitudinal	2. Cross-sectional	3. Census
	Change in the earnings gap		
	1985-89 Cohort		
Years since migration			
6-10	.111	.084	.082
11-15	.195	.187	.181
16-20	.215	.222	.223
Entry gap	-.333	-.323	-.327
	1990-94 Cohort		
Years since migration			
6-10	.201	.216	.218
11-15	.247	.265	.276
entry gap	-.447	-.469	-.461
	1995-99 Cohort		
Years since migration			
6-10	.148	.151	.144
entry gap	-.264	-.266	-.211

Notes: Same specification as Table 4 in the text but different age at migration restriction (20 to 54). All coefficients are statistically significant at 5% level.