Paper presented at the John Deutsch Institute Conference, "Retirement Policy Issues in Canada", October 26-27, 2007 at Queen's University. The final version is published in a book entitled *Retirement Policy Issues in Canada*, edited by Michael G. Abbott, Charles M. Beach, Robin W. Boadway and James G. MacKinnon, 2009 (Kingston: John Deutsch Institute, Queen's University). Published in cooperation with McGill-Queen's University Press and available at: http://mqup.mcgill.ca .

**PRELIMINARY DRAFT ONLY**

**Oct. 17, 2007**

# The Retirement Prospects of Immigrants: Will it require a new Social Contract?

**Derek Hum**

**Wayne Simpson**

**Department of Economics**

**The University of Manitoba**

# The Retirement Prospects of Immigrants: Will it require a new Social Contract?

## Introduction

An aging population and low fertility rates are acknowledged features of Canada's demographic landscape.[1] Immigration is now seen as a key component influencing the growth of population and the labour force. However, demographical projections typically demonstrate that immigration as a policy tool cannot possibly offset the effects of low fertility and an aging population structure since the level of immigration required to do so is simply not feasible.[2] To a lesser extent, Canadian productivity growth and, concomitantly, our capacity to fund social benefits for all Canadians depend on our labour force and immigration. Canada's retirement programs, public and private, will come under strain. In the public sector, Canada's aging population together with continued reliance on pay as you go financing of public pensions will create pressures on our ability to maintain benefit levels, stable premiums and flexible retirement timing, or all three. In the private sector, increasing reliance on immigration from lesser-developed countries for population growth and productivity also has implications if immigrants do not easily and quickly integrate into Canada's labour market. For example, Picot et al (2007) suggest that the new face of poverty in Canada is increasingly "immigrant", and recent work on labour market assimilation of immigrants has established the deteriorating prospects in the wage labour market as well as a lack of training opportunities for

[1] See, for example, HRSDC (2007) "Older Workers: Challenges and Policy Issues", Background paper for an expert panel on older workers.
[2] See Denton and Spencer (2005).

immigrants vis a vis earlier generations (see Hum and Simpson 2004 a, b; 2003). If immigrants increasingly fall into poverty, or are increasingly unable to achieve economic integration in the labour markets, this bodes ill for their retirement prospects.

These trends have implications (macro and micro) for individual retirement plans and public policies, especially for immigrants. A lower lifetime earnings profile for immigrants implies permanent scarring discomfort in the retirement years. Past savings and asset accumulation, as well as the need for future income affect the choice of retirement date. Immigrants may have to postpone or forego retirement altogether in order to maintain certain target income levels. On the other hand, there is the suggestion that immigrants may have stronger desires to work, setting aside other factors, so that delay in retirement may be partly a matter of choice. It has been noted that Canada's public pensions have a major influence on work incentives (Baker et al 2003).

Canada's commitment to admitting and integrating new immigrants is unyielding and irreversible. Nonetheless, a failure to integrate immigrants into Canada's workforce will, over time, engender long run costs for Canada's social benefits, including its suite of retirement programs. A new social contract may be necessary if Canada continues to welcome large numbers of immigrants but fail to integrate them in the economy, all the while desiring generous retirement benefits despite sluggish productivity.

This essay compares the retirement prospects of immigrants (male and female) with their native-born Canadian counterparts (as a benchmark) using Census data in addition to SLID (Survey of Labour and Income Dynamics). Based upon estimated structural differences in life trajectories of immigrants

and native-born persons with respect to lifetime earnings, we simulate the potential "retirement gap" of both groups at various ages. This is useful for various policy investigations since one might examine, for example, the "gap" at age 60 (when early CPP is possible), or at age 65 (when OAS and GIS is possible, etc.) Since our objective is to provide a reasonable basis to inform policy, the main text merely sketches the econometrics of the simulation. Technical details are provided in Appendix A.

The structure of the essay is as follows. The next section (section 2) outlines how we construct profiles of immigrant earnings by employing various Censuses. The challenge is to incorporate data from all available Censuses, yet distinguish the earnings profile over time between immigrants and native–born workers After outlining our econometric approach, we summarize the main results of our estimates. Section 3 calculates what we term the "retirement gap", a measure that summarizes the difference in career earnings that an immigrant might expect at entry vis a vis a comparable native born worker. Section 4 then reexamines the pension gap with more direct evidence from SLID (Survey of Labour and Income Dynamic) on contributions and registered pension plans and private retirement income. Some concluding remarks are offered in section 5.

## 2 Immigrant Integration Earnings Profiles and the Pension Gap

Since we have data on earnings for immigrants and native-born workers in Canada for a variety of Censuses that span various periods of economic conditions, this poses a challenge. We wish to estimate earnings profiles for individuals over a lifetime, yet Census data provide information at a single specific point. Consequently, we use the now conventional "quasi-panel" approach.

The Census provides annual earnings at time t for immigrants who arrived in cohort i, denoted $y^1_{it}$, and for native born, denoted $y^0_{t.}$ We assume these Censuses occur k (=5) years apart. For any cross-section $t$ one can then estimate the predicted earnings difference between immigrant cohorts $i$ and $i+k$ relative to the native born, where the earlier cohort $i$ is associated with longer years since migration, as

$$\hat{y}^1_{i,t} - \hat{y}^1_{i+k,t} = \left[\left(\hat{y}^1_{i,t} - \hat{y}^1_{i,t-k}\right) - \left(\hat{y}^0_{t} - \hat{y}^0_{t-k}\right)\right] + \left[\left(\hat{y}^1_{i,t-k} - \hat{y}^1_{i+k,t}\right) - \left(\hat{y}^0_{t-k} - \hat{y}^0_{t}\right)\right] \qquad [1]^3$$

The first term on the right hand side of equation [1] then captures the difference in the growth of earnings for immigrant cohort $i$ and the native born from Census period $t-k$ to Census period $t$. This within-cohort growth measures the extent of immigrant integration of cohort $i$ relative to the native-born comparison group. The second term on the right hand side of equation [1] captures the difference in growth between cohort $i$ in period $t-k$ and cohort $i+k$ in period $t$, or across-cohort growth for given years since migration, relative to the native born counterfactual. The second term represents the bias associated with cross-sectional estimates of within-cohort earnings growth.

While much of the focus in the literature is on five-year growth rates of the earnings of **entering immigrants** in the first five years after entry, longer segments of the immigrant integration profiles can be derived from equation [1] for a sequence of Census cross-sections. In particular, consider immigrant cohort $i$ that entered $r$ Census periods earlier. One can estimate the entry effect, the difference in earnings between the entering immigrant cohort and the native born as $\hat{y}^1_{t-rk} - \hat{y}^0_{t-rk}$, evaluated for the characteristics of immigrant cohort $i$. Then the within-cohort growth measures for immigrant

[3] This is equation [3] in the Appendix and is found in Baker and Benjamin (1994, equation [8], 381), Grant (1999, equation [3], 939), and Frenette and Morissette (2003, equation [4], 2).

cohort $i$ relative to the native born over Census periods $t, t-k, \ldots, t-rk$ provide fairly lengthy estimates of the integration profile (the earnings gap for years since migration) for immigrants who arrived a long time ago.

[Figures 1 and 2 about here]

Figures 1 and 2 portray the immigrant integration profile that incorporates the entry and within growth effects at 5-year intervals, estimated for a randomly drawn native born comparison group and a matched native-born comparison group, for the immigrant arrival cohorts from 1976-80 to 1991-95. The horizontal axis represents the gap between the mean earnings of an immigration cohort and its native born counterparts, using OLS regression (the conventional method) and propensity score matching (marked with an M) to determine the native-born comparison group. Figure 1 uses the sparser Baker and Benjamin (1994) specification while the Figure 2 uses the richer Frenette and Morissette (2003) specification that includes visible minority (rather than just black) and urban/regional variables (Montreal, Toronto, Vancouver, Quebec except Montreal, Ontario except Toronto, Manitoba, Saskatchewan, Alberta, B.C. except Vancouver). The F&M specification is only calculated for the Censuses from 1986 because visible minority was not defined before then.

We highlight the following results:

(1) **The results are similar for the B&B and F&M specifications**; we therefore refer to the results from the B&B specification below for (2) to (5).

(2) **The entry effects are increasing**; the estimates from the matched comparison group are slightly larger for all cohorts except 1976-80.

(3) **The assimilation (within growth) effects are substantial and do not necessarily suggest that later cohorts will not achieve parity**; for example, the largest entry effect for the 1991-95 cohort is combined with a substantial assimilation effect (about 15%) in years 5-10 which, if it continues, would permit parity within 20 years.

(4) **Projections based on particular specifications of the form of immigration integration profile are unreliable** (including our own in Hum and Simpson (2004)). It is difficult to project assimilation rates because they are not uniform; for example, cohorts IM76-80 and IM86-90 faltered in the first five years (esp. with the matched sample) and IM81-85 falters after doing well in the first five years. Contrary to Grant's (1999) projection, her IM81-85 may not achieve parity with the native born.

## 3. Calculation of Lifetime (Pensionable) Earnings Gap

The immigrant integration profiles depicted in the previous section reflect the percentage gap in mean earnings between an immigrant cohort and that of a comparable group of native-born workers. Despite our expressed reservations, we now use these profiles to calculate the cumulative lifetime difference in earnings between these two groups (suitably discounted). This provides an estimate of the "retirement gap" between the two groups and the proportionate difference in pensionable earnings, since absent inheritances, lottery winnings, or other unexpected windfalls; it is the pattern of lifetime earnings (and savings) that determine the economic resources available for retirement.

One useful measure might be the net present value of the earnings gap, which would then represent the lump-sum gap in career earnings that an

immigrant could expect at entry. This can also be expressed as a percentage of the earnings of a comparable native-born worker. If private pension income, and to a lesser extent CPP/QPP income, is closely related to earnings, the lump-sum earnings gap provides a measure of the pension gap between immigrants and the native born.

The logic of our calculation is as follows. Suppose we normalize native-born earnings to \$1 per year over a working career of $T$ years and suppose that $r$ is the real rate of interest and discount rate. Then initial native-born earnings will have a present value of \$1 and the value at retirement will be $\$1(1+r)^T$. Over T years, the stream of earnings will have a present value of $P_{nb} = \sum_{i=1}^{T} \$1/(1+r)^i$ and a value at retirement of $L_{nb} = \sum_{i=1}^{T} \$1(1+r)^i$.

Suppose now that immigrants initially earn a proportion $1-\gamma_0$ of native-born earnings, where $0 < \gamma_0 < 1$ is the entry gap that is eroded with time spent in Canada ($\partial\gamma_i / \partial i < 0$). Parity with native-born earnings ($\gamma_i = 1$) may be achieved at some year $i$ during the work career (or years since migration). Suppose further that a constant portion $s$ of earnings is saved for a private pension, such that an annuity is financed from a retirement earnings pool of $sL_{nb}$ with a present value of $sP_{nb}$. Then the corresponding present value of earnings for the foreign born will be $P_{fb} = \sum_{i=1}^{T} (1-\gamma_i)/(1+r)^i$ and the value at retirement will be $L_{fb} = \sum_{i=1}^{T} \$(1-\gamma_i)(1+r)^i$. Assuming a common savings rate for foreign and native born workers, the retirement

earnings pool will be $sL_{fb}$ with a present value of $sP_{fb}$ such that the pension gap will be $\left[sP_{nb} - sP_{fb}\right]/sP_{nb} = \sum_{i=1}^{T} \gamma_i /(1+r)^i \Big/ \sum_{i=1}^{T} 1/(1+r)^i$ , [1]

which corresponds to a pension gap at retirement of

$$\left[sL_{nb} - sL_{fb}\right]/sL_{nb} = \sum_{i=1}^{T} \gamma_i (1+r)^i \Big/ \sum_{i=1}^{T} (1+r)^i \quad .$$

Consider the 1976-80-immigrant cohort whose lifetime earnings pattern, relative to the native born, is captured by subsequent Censuses to 2001. We adopt the Baker and Benjamin (1994) specification, which is compatible with all previous Censuses to 1976. (The expanded Frenette and Morissette (2003) specification yielded similar results in the cases we estimated.) We use the estimates derived from the traditional OLS estimates first. For this cohort, the estimated immigrant integration profiles imply a pension gap of 11.4% using a discount rate (r) of 5% and a pension gap of 13.1% using a discount rate of 10%. A larger pension gap is to be expected with higher discount rates because the smaller differences between native and foreign-born earnings in the future (arising as immigrant integration proceeds) are more heavily discounted.

Our alternative estimates derived from propensity score matching produce a slightly more pessimistic picture of the immigrant integration profile and hence a slightly larger pension gap. For the 1976-80 cohort, we estimate a pension gap of 16.7% at a 5% discount rate and 17.4% at a 10% discount rate.

For other immigrant cohorts, the pension gap is more difficult to estimate because the immigrant integration profile is incomplete. Our

approach is simply to “eyeball” the trajectory of the immigrant integration profile for each cohort; this produces the results reported in Table 1 for each cohort from 1976-80 to 1991-95. More sophisticated approaches could be taken but are unlikely to produce very different pension gap estimates since the later earnings are discounted more heavily. As might be expected, the pension gap is growing for more recent cohorts since their initial earnings disadvantage (entry effect) rises. The OLS estimates suggest that the pension gap has doubled from 11% to 22% between the 1976-80 and 1991-95 cohorts, compared to the matching estimates that indicate the gap increasing from 17% to 28%, for a discount rate of 5%. With a discount rate of 10%, the OLS estimates again suggest a doubling of the pension gap from 13% to 26%, while the matching estimates suggest an increase from 17% to 33%. Our results in Table 1 quantify the growing prospective pension gap; this should not be much of a surprise since it is a mirror of the declining labour market fortunes of more recent immigrant cohorts.

Table 1. Estimated Pension Gaps as Percentage of Native Born

| | Discount rate | IM7680 | IM8185 | IM8690 | IM9195 |
|---|---|---|---|---|---|
| OLS estimates | 5.0% | 11.4% | 17.5% | 21.0% | 21.7% |
| | 10.0% | 13.1% | 20.4% | 23.5% | 26.3% |
| Matched est. | 5.0% | 16.7% | 26.4% | 26.7% | 27.9% |
| | 10.0% | 17.4% | 29.2% | 29.3% | 32.6% |

Source: Estimates of immigrant earnings profiles from the Canadian Censuses of 1976, 1981, 1986, 1991, 1996 and 2001 plus imputed (“eyeballed”) estimates of the profile over a working career of 25 years.

## 4 Evidence on the Pension Gap from SLID

Previous sections have been guided by Census data. We now examine more direct evidence on the pension gap with data from the Survey of Labour and Income Dynamics (SLID) 2002 Public File. SLID is designed as an overlapping 6-year panel to capture labour market activity and financial income information for two panels of individual respondents in each survey period. In particular, SLID provides what amounts to tax record information for Registered Pension Plan (RPP) contributions as well as private pension income for respondents identified by immigration status, age and sex. Consequently, SLID is a valuable data source in addition to the Census information. We restrict our analysis to males to be consistent with our earlier evidence.

We ask two questions. First, is there a difference in RPP contributions for immigrant and native-born men, **who are not retired** (have no pension income), by age? Second, is there a difference in private pension income between immigrant and native-born men **who have retired** (are drawing pension income) by age? We restrict our retirement group to those over the age of 55.

Figure 3 provides nonparametric estimates of RPP contributions by age for immigrant and native-born men.[4] It is clear that there is a pension contribution gap at almost all ages, beginning at very early ages and increasing to age 50, then declining. A small RPP contribution advantage

[4] Those who responded, "don't know" to the question on immigrant status are deleted from our analysis. This group is, however, quite large in SLID. The nonparametric estimates are derived from locally weighted regressions that use the tricube weighting function in STATA8.0 LOWESS.

for immigrants occurs after age 70. This does not appear to reflect a difference in rates of labour force participation between immigrant and native born men after age 65, but rather a higher rate of contribution by immigrants who remain in the labour force --- perhaps motivated by a desire to "catch up" for low contributions earlier in life.

Figure 4 provides similar evidence for private pension plan income. Immigrants declare less private pension income at almost all ages, consistent with the lower RPP contribution rates shown in Figure 3. Of course, immigrants drawing pensions likely arrived much earlier than those who are now working and making RPP contributions in Figure 4. Nonetheless, the private pension income gap is considerable: At age 60 (when early withdrawal of CPP is permitted), the gap is about $2500 or 10% of the average native born private pensions at that age. At ages 65 (the traditional benchmark age) and 69 (when withdrawal of pension monies is often mandatory), the pension gap is about $5,000 or 21-22% of the average native born private pension at these ages. This gap is consistent with our estimates using the Censuses of the differences in lifetime earnings between immigrant and native-born men. To rephrase the matter slightly, at certain conventional benchmark ages when retirement decisions must be considered, immigrants will have approximately 10% less income if they attempt "early retirement" at age 60, and 21-22 % less if they choose to delay retirement.

It must be remembered in all this discussion that earlier cohorts had more success with economic integration than the present generation, and therefore our estimate of the integration profiles contain immigrant experiences that reflect this better integration performance. Since the current cohorts of immigrant who are still working have greater earnings gaps compared to the native born, our results predict that this pension gap

will only continue to grow over time. This will pose a tremendous challenge for public policy in Canada, not only respecting the design and funding of our public pension programs, but also respecting our enduring legacy of efforts to assist and integrate immigrant workers.

## 5 Concluding Remarks

Canada's commitment to admitting and integrating new immigrants is irreversible. Unlike the cohorts that entered in the 1960s, immigrants to Canada within the last three decades have not fared as successfully. Over time, a continuing failure to integrate immigrants into the workforce will incur long run costs for Canada's social benefits, including its suite of retirement programs.

This essay compares the retirement prospects of immigrants with their native-born Canadian counterparts employing data from the Census and SLID. Census data cover the entire population but represent a snapshot at a single moment. Because our interest is in the lifetime income profiles of individuals, we employ a quasi-panel approach by combining several Census data sets to estimate an economic integration time path. We also examine data from SLID. SLID is a useful data source because of its wide variable set (including government transfers) and panel nature. We employ matching methods to determine an appropriate comparison group (based on demographic characteristics, etc.) in order to compare immigrant earnings with native-born earnings. Based upon structural differences in life trajectories of immigrants and native-born persons (as a benchmark), we calculate what we term a "retirement gap", defined as the net present value of the earnings gap between immigrant and comparable native born

individuals. This retirement gap represents the lump-sum gap in career earnings that an immigrant could expect at entry, expressed as a percentage of the earnings of a comparable native-born worker. We believe the retirement gap measure provides a reasonable basis to inform retirement policy in Canada respecting both immigrants and native-born individuals.

**References**


Baker, M. and D. Benjamin (1994) "The Performance of Immigrants in the Canadian Labor Market." *Journal of Labor Economics* 12(3), 369-405

Baker, Michael, Jonathan Gruber & Kevin Milligan (2003)"The Retirement Incentive Effects of Canada's Income Security Programs." *Canadian Journal of Economics* 36(2): 261-290.

Denton, Frank & Byron Spencer (2005). *Population Aging and the Macroeconomy: Explorations in the Use of Immigration as an Instrument of Control.* Quantitative Studies in Economics and Population (QSEP) Report No. 398. McMaster University, Canada.

Frenette, M. and R. Morissette (2003) "Will They Ever Converge? Earnings of Immigrant and Canadian born Workers Over the Last Two Decades," Analytical Studies Branch research paper series no. 215, Statistics Canada

Grant, M. (1999) "Evidence of new immigrant integration in Canada." *Canadian Journal of Economics* 32(4): 930-955

Hum, Derek & Wayne Simpson (2003) "Job-Related Training Activity by Immigrants to Canada." *Canadian Public Policy/Analyse de Politiques.* Vol. 29, No. 4, 1-22.

Hum, Derek & Wayne Simpson (2004a**)** "Economic Integration of Immigrants to Canada: A Short Survey." *Canadian Journal of Urban Research.* Vol. 13, 1, 46 - 61.

Hum, Derek &Wayne Simpson (2004b) "Reinterpreting the Performance of Immigrant Wages from Panel Data." 2004 AD. *Empirical Economics.* Vol. 29, No.1, 129-47**.**

Picot, Garnett (2004) "The Deteriorating Economic Welfare of Canadian Immigrants". *Canadian Journal of Urban Research.* 13, 1, 25-45.

**Figure 1: Immigrant Integration Profiles, B&B Specification**

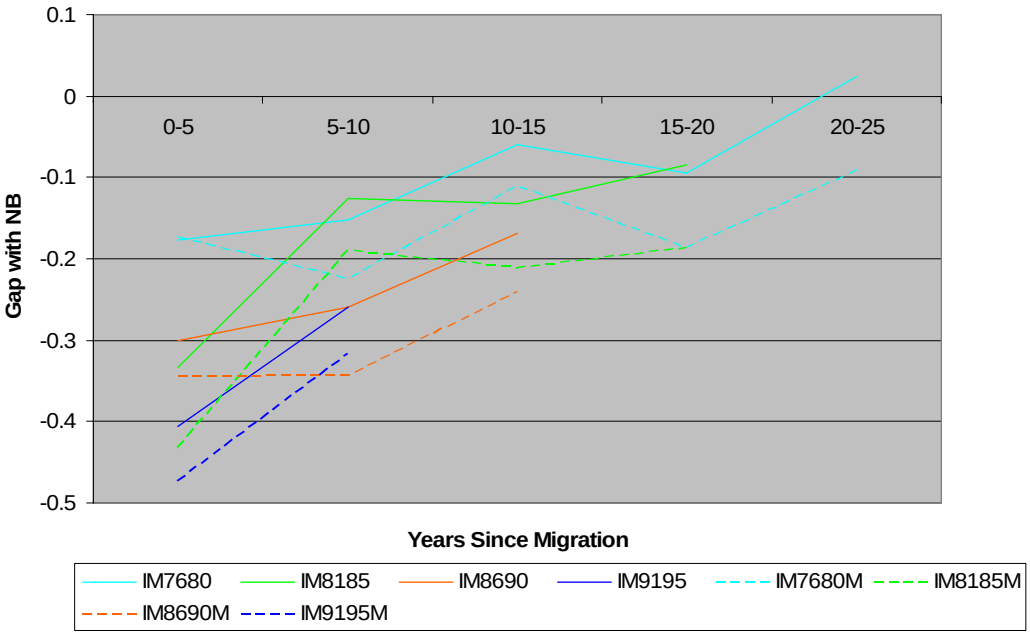



**Figure 2: Immigrant Integration Profiles, F&M Specification**

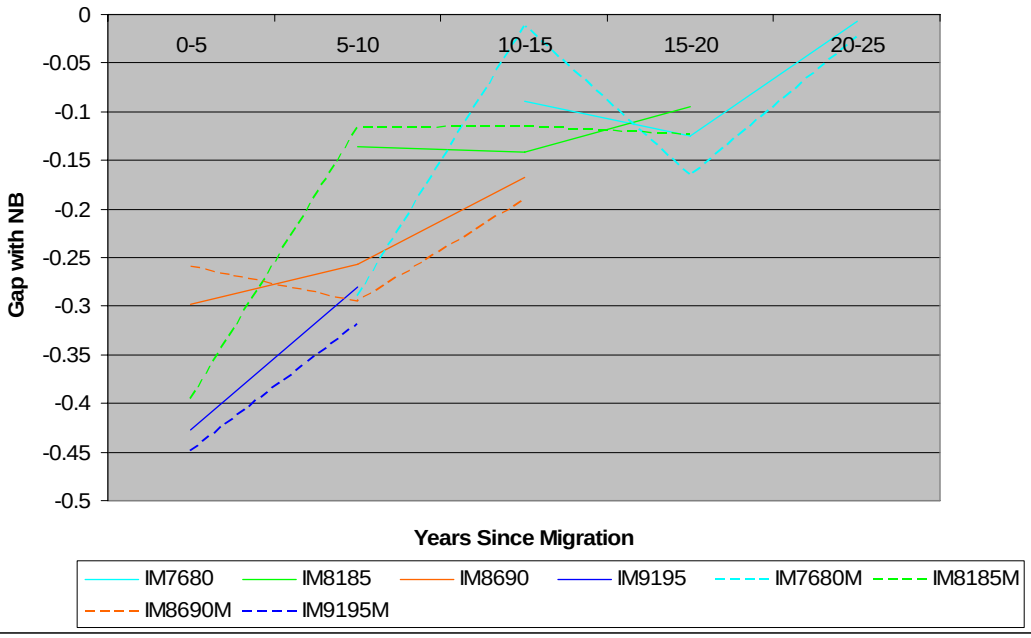


**Figure 3. Registered Pension Plan Contributions by Age, Immigrant vs. Native Born Men**

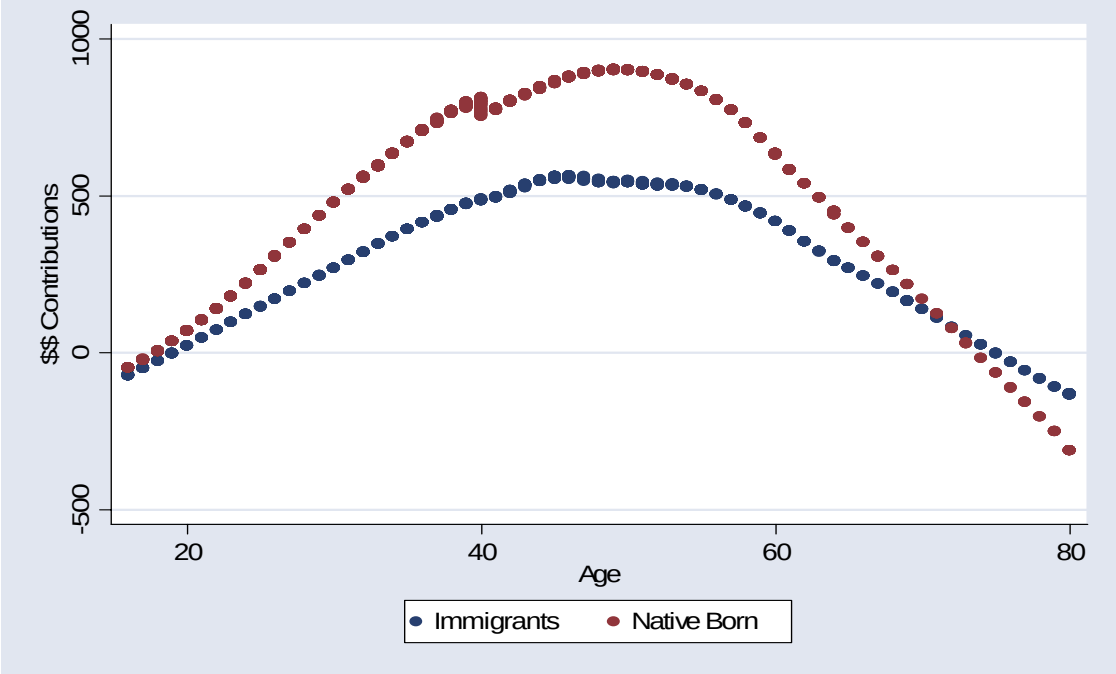
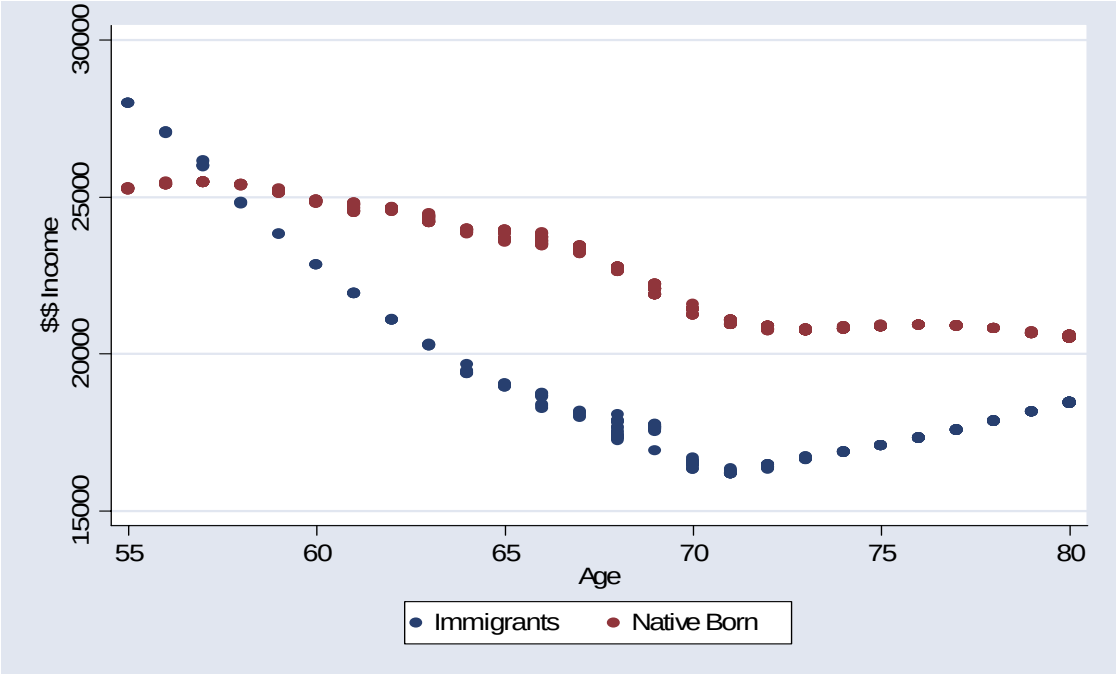



**Figure 4. Private Pension Income by Age, Immigrant vs. Native Born Men 55 Years and Older**




Source: SLID 2002 Public File

## Appendix A

## Toward Robust Economic Modelling of Immigrant Integration

### 1. Introduction

Key to our discussion of "retirement gap" between immigrants and native-born workers is the estimated lifetime career earnings of these two groups. This appendix discusses in greater detail the econometric complexities involved in this exercise, and our approach to the issues.

In the absence of experimental evidence, researchers regularly use nonexperimental data. The *modus operandi* involves specification of a linear regression model with a set of critical explanatory variables that vary across agents and which, in the case of panel data, vary over time as well. Any misspecification involving the omission of relevant variables will introduce bias when they are correlated with included explanatory variables, as is normally the case in nonexperimental data. Misspecification may include errors in the functional form.

Recognition of these potential specification pitfalls typically leads to specification searches in which a variety of models are estimated. The scientific hope is that the impact of the critical variables is robust, that it varies little as the specification of additional explanatory variables is altered. This hope is often not realized, leaving considerable doubt about the issue under study, the validity of published research, and the prospect for further analysis of nonexperimental evidence to resolve the question. Ho, Imai, King and Stuart (2006) refer to this problem as "model dependence."

Recent research in program evaluation has provided new insights to the search for robust models, or models which eliminate the problem of model dependence. The basic

evaluation problem is to estimate the impact of a program by comparing the behaviour of the program group with an appropriate comparison group. This literature demonstrates that important sources of bias arise but can be eliminated by careful statistical matching of the program and non-recipient groups, based on their observed characteristics, and assessment of what valid comparisons are supported by the data.

The argument for statistical matching to select appropriate comparators and produce robust model estimates is potentially more general. In this appendix we review the arguments for applying this approach to estimating the performance gap between immigrants and the native born, or immigrant integration. We review the relevant empirical literature and methods used to estimate immigrant integration. Our re-analysis of this evidence using matching estimators is the forms the basis of our results in the main text.

**2. Model Dependence and Matching**

The problem of model dependence in the context of regression models applied to nonexperimental data arises from the literature on program evaluation. Suppose that you have a nonexperimental data set with an identifier of program participation. Evaluators wish to answer the question: What would have been the effect on those who took the program if they had not taken it? This measure of program impact is the "average treatment effect on the treated." The problem with this question is, of course, that we cannot observe someone as both a recipient and a non-recipient of the program at the same time. The standard solution is to use the non-recipients to estimate what the outcomes would have been for the recipients had they not taken the program. The

identifying assumption is that the mean outcome of nonrecipients is identical to that of recipients had they not taken the program, conditional on whatever characteristics we can observe about these groups that affect the outcome. But this may not hold, leading to bias in the estimation of the average treatment effect.

Heckman, Ichimura, Smith and Todd (1998) show that there are three sources of selection bias with this approach. One bias arises from lack of "common support" when either the distribution of observable characteristics for the program group does not overlap the distribution of observable characteristics for the non-recipient group or *vice versa*. A second bias arises from "differential weighting" of the observable characteristics in the program group and non-recipient group samples where there is common support. The third bias is "true selection bias" which arises from unobservables and remains even when common values of the regressors for both the program group and an appropriate comparison group are used. True selection bias cannot be eliminated, but the sources of bias arising from lack of common support and differential weighting can be eliminated by careful statistical or propensity score matching of the program and non-recipient groups to eliminate differential weighting and to assess what valid comparisons are supported by the data available. Essentially, this involves relatively straightforward estimation of a model of program participation for the entire sample of program participants and non-recipients, based on their observed characteristics. This model yields "propensity to participate" scores which can be used to match program participants with one or a weighted average of non-recipients that have comparable predicted participation probabilities. The adequacy of the match can be assessed by balancing tests, which examine the similarity of the properties of the samples of participants and

matched non-recipients. Heckman et al (1998) and other have found that matching estimators, which then estimate program impacts nonparametrically from the refined samples of participants and non-participants, can provide more accurate estimates of program impacts in certain circumstances.

**3. Application to Immigrant Integration**

We apply the above approach to analyze the labour market performance of immigrants over time relative to their native born counterparts, or what we term immigrant integration. Our "program impact" is simply "immigrant status" with a specified number of years in Canada.. Our counterfactual or comparison group is the native born; that is, we use those born in Canada to estimate what the outcomes would have been for the immigrants had they been born in Canada.

To be specific, let $y$ represent earnings. Then $y$ will depend on a set of observable characteristics, $x$, and, for immigrants, an immigrant status function, $\gamma(h)$, where $h$ represents years since migration. Let $D$ be a dummy variable identifying our samples of immigrants ($D=1$) and native born ($D=0$). Then we can write our linear regression model to estimate immigrant integration effects in the form:

$$y^i = x^i\beta^i + \gamma(h)D + \varepsilon^i, \qquad i = 0(D=0), 1(D=1) \qquad [1]$$

That is, we can estimate immigrant integration effects by estimating $\gamma(h)$ from regression analysis of equation [1] if the identifying assumption that the mean labour market outcome of the native born is identical to that of immigrants had they been born in Canada, conditional on observable characteristics is valid.

A body of empirical research has emerged in Canada following Borjas' (1985) pioneering quasi-panel approach to the analysis of immigrant integration in the U.S.

Borjas showed that immigrant integration tends to be overestimated in cross-sectional studies when unobserved differences among immigrant cohorts, or what Borjas terms cohort quality, is declining. This bias from cross-sectional evidence can be corrected by using a series of cross-sections of individual workers over time to separate immigrant integration within cohorts from differences across immigrant cohorts.

There are three comparable studies of immigrant integration in Canada that adopt Borjas' quasi-panel analysis. Baker and Benjamin (1994), analyzing the 1971, 1981 and 1986 Censuses, estimate very low, and even negative, rates of growth of log earnings within immigrant cohorts. Adding the 1991 Census, however, Grant (1999) estimates a rapid convergence for the 1980s immigrant cohort that implies parity within ten years. Using all Censuses from 1981 to 2001, Frenette and Morissette (2003) find little evidence to suggest that the earnings growth of immigrants who landed in the 1990s will be sufficient to ever achieve parity with their native born counterparts. These results have left a succession of analysts to wonder what might explain the rapid changes in immigrant fortunes through the last three decades of the 20$^{th}$ century.[5]

Public use microdata files from the Census provide a series of cross-sections which classify immigrant cohorts by period of immigration. Thus, researchers rewrite equation [1] to replace the specific immigrant integration profile $\gamma(h)$ with a set of cohort-specific intercepts, $\delta_{i,t}^{1}$, which identify immigrant cohort $i$ in Census time period $t$:

[5] Attempts to explain the changes in immigrant integration across cohorts include McDonald and Worswick (1998) and Green and Worswick (2003), using non-Census data, and Abdurrahman and Skuterud (2003) and Frenette and Morissette (2003) using Census data. At this point, we are not concerned with the analysis of what is causing changing immigrant fortunes but rather with the robustness of the estimates of immigrant integration on which this literature is predicated.

$$D = 1(immigrants) : y_t^1 = x_t^1 \beta_t^1 + \sum_i \delta_{i,t}^1 + \varepsilon_t^1$$
$$D = 0(native\,born) : y_t^0 = x_t^0 \beta_t^0 + \delta_t^0 + \varepsilon_t^0 \qquad [2]$$

The intercept term $\delta_t^0$ is common to all native born. The equations in [2] can be estimated for each Census cross-section. From these equations, researchers can extract estimates of the predicted (log) earnings of immigrants by cohort and time period and estimates of the predicted earnings of the native born by time period. For any cross-section $t$ one can estimate the predicted earnings difference between immigrant cohorts $i$ and $i+k$ relative to the native born, where the earlier cohort $i$ is associated with longer years since migration, as

$$\hat{y}_{i,t}^1 - \hat{y}_{i+k,t}^1 = \left[\left(\hat{y}_{i,t}^1 - \hat{y}_{i,t-k}^1\right) - \left(\hat{y}_t^0 - \hat{y}_{t-k}^0\right)\right] + \left[\left(\hat{y}_{i,t-k}^1 - \hat{y}_{i+k,t}^1\right) - \left(\hat{y}_{t-k}^0 - \hat{y}_t^0\right)\right] \qquad [3]^6$$

Then the first term on the right hand side of equation [3] captures the difference in the growth of earnings for immigrant cohort $i$ and the native born from Census period $t-k$ to Census period $t$. This within-cohort growth measures the extent of immigrant integration of cohort $i$ relative to the native born comparison group. The second term on the right hand side of equation [3] captures the difference in growth between cohort $i$ in period $t-k$ and cohort $i+k$ in period $t$, or across-cohort growth for given years since migration, relative to the native born counterfactual. The second term represents the bias associated with cross-sectional estimates of within-cohort earnings growth. Borjas showed that, since across-cohort growth is positive when cohort quality is declining,[7] cross-sectional estimates of immigrant earnings growth will overestimate within-cohort

[6] This equation is found in Baker and Benjamin (1994, equation [8], 381), Grant (1999, equation [3], 939), and Frenette and Morissette (2003, equation [4], 2).

[7] That is, earlier cohort $i$ does better than later cohort $i+k$ for given years since migration in relation to their native born counterparts. This declining cohort quality is broadly consistent with a shift in region-of-origin immigration patterns in North America from Europe to South Asia over the last four decades, if South Asian immigrants bring linguistic, work and social skills that are less valuable to the North American labour market.

growth, or true immigrant integration.

## 4 Limitations and the Case for Matching

The focus, then, will be on estimates of within-cohort growth represented by

$$\left[\left(\hat{y}^1_{i,t} - \hat{y}^1_{i,t-k}\right) - \left(\hat{y}^0_{t} - \hat{y}^0_{t-k}\right)\right] \qquad [4]$$

Predicted earnings in equation [3] or [4] are based on the estimates of equation [2] evaluated for some common bundle of characteristics. The convention is to evaluate predicted earnings at the mean sample characteristics for immigrant cohort $i$ in period $t$, $\bar{x}^1_{i,t}$. In that case, given the estimates of $\hat{\beta}$ and $\hat{\delta}$ from equation [2], equation [3] reduces to:

$$\begin{aligned}\hat{\delta}^1_{i,t} - \hat{\delta}^1_{i+k,t} &= [\bar{x}_{i,t}\{(\hat{\beta}^1_t - \hat{\beta}^1_{t-k}) - (\hat{\beta}^0_t - \hat{\beta}^0_{t-k})\} + \{(\hat{\delta}^1_{i,t} - \hat{\delta}^1_{i,t-k}) - (\hat{\delta}^0_t - \hat{\delta}^0_{t-k})\}] \\ &\quad + [\bar{x}_{i,t}\{(\hat{\beta}^1_{t-k} - \hat{\beta}^1_t) - (\hat{\beta}^0_{t-k} - \hat{\beta}^0_t)\} + \{(\hat{\delta}^1_{i,t-k} - \hat{\delta}^1_{i+k,t}) - (\hat{\delta}^0_{t-k} - \hat{\delta}^0_t)\}] \qquad [5] \\ &= [\hat{a}_{i,t,t-k} + (\hat{\delta}^1_{i,t} - \hat{\delta}^1_{i,t-k})] + [-\hat{a}_{i,t,t-k} + (\hat{\delta}^1_{i,t-k} - \hat{\delta}^1_{i+k,t})]\end{aligned}$$

where $\hat{a}_{i,t,t-k} = \bar{x}_{i,t}\{(\hat{\beta}^1_t - \hat{\beta}^1_{t-k}) - (\hat{\beta}^0_t - \hat{\beta}^0_{t-k})\} - (\hat{\delta}^0_t - \hat{\delta}^0_{t-k})$. It is well known that decomposition analyses may be sensitive to the choice of base characteristics, $\bar{x}_{it}$ [Horrace and Oaxaca, 2001] which in this case directly affects the computation of $\hat{a}_{i,t,t-k}$. This will no longer be an issue if the native born comparison sample is chosen to have characteristics identical on average to the immigrant cohort sample.

In a rarely cited paper, Yuengert (1994) finds that this approach to estimating immigrant earnings is sensitive to both the choice of comparison point and the particular specification of earnings used. The comparisons used by Borjas (1985) tend to understate U.S. immigrant earnings relative to the native born while standard Mincerian earnings specifications, linear in education and quadratic in experience, overstate relative earnings for immigrants at the extremes of the education spectrum. These concerns would likely

apply to immigrant integration studies for Canada, although perhaps in different ways since immigration patterns differ between the two countries.

Previous Canadian studies have used a random sample of the native born to construct a comparison group of comparable size to the immigrant sample for each Census. Baker and Benjamin (1994) use a one-sixth random sample, Grant (1999) uses a comparable (but unspecified) random sample of the native born except for a full sample of blacks, and Frenette and Morissette (2003) use a 20% random sample of the native born. From this perspective, we can ask whether a random sample of the native born provides an appropriate comparison group. In this case, the comparison group is intended to represent what the outcomes for immigrants would have been had they not been immigrants; i.e., had they been born and raised in Canada. This requires the identifying assumption that the mean outcome of the native born is identical to that of immigrants had they been born and raised in Canada, conditional on observable characteristics. Intuitively, this suggests that the native born sample should look like the immigrant sample to provide an appropriate counterfactual. As discussed above, recent literature suggests that a native born comparison sample matched to the immigrant sample can reduce model dependence and give more reliable estimates of immigrant integration.

This issue might not be important if the characteristics of the immigrant and native born samples were similar. Authors have consistently observed otherwise, however. Baker and Benjamin (1994, Table 1) show that immigrants tend to be better educated than the native born, although the immigrant advantage is declining over time. Grant (1999, Table 1) finds a reversal of this trend in the 1980s. She also observes that

immigrants are older on average with more potential work experience (age minus schooling minus 5), are regionally concentrated in Ontario and British Columbia, and are more ethnically diverse than the native born. Frenette and Morissette (2003, Table 1) find a widening gap in mean education between immigrants and the native born through the 1990s. More immigrants are members of visible minorities, although this information has only been collected since 1986. Moreover, while researchers typically focus on the mean characteristics of their samples, this is not adequate to ensure that the samples "match up" well. There could be dramatic differences in the distribution of characteristics between immigrant and native born samples whose means are the same. Smith (2006), for example, finds that the mean schooling of immigrants is lower by 1.3 years in the U.S. as of 2002 but, perhaps more importantly, that immigrants are more highly represented in both the lowest and highest education categories; i.e., their years of schooling are more dispersed.

This idea of matching the characteristics of the immigrant and native born samples is not entirely new. One approach to explain the rapid changes in immigrant fortunes in Canada in the latter part of the 20$^{th}$ century matches recent immigrants with native born labour market entrants. Frenette and Morissette (2003), using Census data, and McDonald and Worswick (1998) and Green and Worswick (2003), using non-Census data, find that much of the change in the prospects of entering immigrants can be explained by similarly dismal prospects for their native born counterparts entering the labour market, suggesting that at least the estimates of immigrant integration shortly after landing (the entry effect) are sensitive to the choice of comparison group. Frenette and Morissette, for example, find that the immigrant earnings disadvantage at entry has risen

only modestly from 7% to 12% over the two decades when arriving immigrants are compared only to native born entrants, compared to entry effects that rise from 17% to 40% during this period when all native-born workers are used as the benchmark.

The estimates of immigrant integration used in section 2 of the main text of this essay is based upon our re analysis using the quasi-panel approach to the Census public microdata files from 1971, 1981, 1986, 1991, 1996, and 2001.

**References**


Abdurrahman, A. and M. Skuterud (2003) "Explaining the Deteriorating Entry Earnings of Canada's Immigration Cohorts: 1966-2000," Family and Labour Studies Division, Statistics Canada, October

Baker, M. and D. Benjamin (1994) "The Performance of Immigrants in the Canadian Labor Market." *Journal of Labor Economics* 12(3), 369-405

Borjas, G. (1985) "Integration, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3(4), 463-489

Frenette, M. and R. Morissette (2003) "Will They Ever Converge? Earnings of Immigrant and Canadian born Workers Over the Last Two Decades," Analytical Studies Branch research paper series no. 215, Statistics Canada

Grant, M. (1999) "Evidence of new immigrant integration in Canada." *Canadian Journal of Economics* 32(4): 930-955

Green, D. and C. Worswick (2003) "Immigrant Earnings Profiles in the Presence of Human Capital Investment: Measuring Cohort and Macro Effects," paper presented to the John Deutsch – CIC conference on immigration, Queen's University, revised September

Heckman, J., H. Ichimura, J. Smith, and P. Todd (1998) "Characterizing selection bias using experimental data," *Econometrica* 66, 1017-98

Ho, D., K. Imai, G. King, and E. Stuart (2006) "Matching as Nonparametric Preprocessing for Reducing Model Dependence in Parametric Causal Inference," mimeo, Harvard University, May 5

Horrace, W., and R. Oaxaca (2001) "Inter-Industry Wage Differentials and the Gender Wage Gap: An Identification Problem," *Industrial and Labor Relations Review* 54(3), 611-18

McDonald, J. and C. Worswick (1998) "The Earnings of Immigrant Men in Canada: Job Tenure, Cohort, and Macroeconomic Conditions." *Industrial and Labor Relations Review* 51(3): 465-482

Smith, J. (2006) “Immigrants and the Labor Market,” *Journal of Labor Economics* 24(2), 203-33.

Yuengert, A. (1994) “Immigrant Earnings, Relative to What? The Importance of Earnings Function Specification and Comparison Points,” *Journal of Applied Econometrics* 9, 71-90