LONGITUDINAL ANALYSIS OF EARNINGS ASSIMILATION AMONG IMMIGRANTS IN QUEBEC

Michael Beenstock

Department of Economics
Hebrew University of Jerusalem

Abstract
A cohort of a 1000 immigrants who entered Quebec in 1989 was followed-up during the subsequent 10 years. We find no evidence of immigrant wage assimilation in this cohort. Our results join a small but growing body of research that shows that there is no evidence of immigrant assimilation in longitudinal data. These negative findings stand in marked contrast to corroboration of assimilation theory in numerous synthetic cohort studies. Our results also show that for this cohort, the immigrant labor market is segmented from the native labor market.

March 9, 2006

We wish to thank Pavel Soifer for able and congenial research assistance.

LONGITUDINAL ANALYSIS OF EARNINGS ASSIMILATION AMONG IMMIGRANTS IN QUEBEC

1. Introduction

1.1 Synthetic Cohorts and Cross Section Data

Empirical research concerning the absorption of immigrants into the labor market of the host country has been hampered by lack of longitudinal data. Chiswick's (1978) pioneering study used cross-section rather than longitudinal data to show that immigrant earnings vary directly with time since migration (TSM). He interpreted this observation as being consistent with the hypothesis that immigrant human capital is not completely portable, and that it takes time for immigrants to adjust to their new milieu. Duration matters because it takes time to acquire language skills (Chiswick and Miller 1995) and to learn about labor market networks and institutions in the host country.

It has long been recognized (Chiswick 1980, Borjas 1985) that there are problems with using static, cross-section data to investigate dynamic phenomena such as immigrant assimilation. Cross-sectional analyses compare different individuals observed at the same point of time, under the assumption that apart from their different durations immigrants are otherwise similar. If, however, they are not similar, dynamic inferences made from static cross-sections will be misleading. Specifically, if more recent immigrant cohorts have less ability, then Chiswick's observation may be consistent with deteriorating cohort quality. More recent immigrants earn less not because they have had less time to assimilate, but because they belong to less able immigrant cohorts. Alternatively, if less successful immigrants re-emigrate (Beenstock 1996), Chiswick's observation may simply mean that more recent immigrants earn less because they have had less opportunity to re-emigrate.

In the absence of genuine longitudinal data, Borjas (1985) suggested the "synthetic cohort methodology", which uses two or more independent cross-sections, to identify the effect of duration on immigrant earnings. A comparison of two immigrants with identical durations, who immigrated at different times, identifies cohort effects and duration effects. The synthetic cohort methodology has been widely

applied in Canada[1] as well as in many other immigrant receiving countries. Indeed, there has been a deterioration in the initial earnings of immigrant in Canada, which Aydemir and Skuterud (2005) partially attribute to language ability and region of origin. While rates of assimilation may not have been strong enough for immigrant earnings to converge to the earnings of demographically similar Canadian natives, they provide clear evidence in favor of the immigrant assimilation hypothesis insofar as immigrant earnings grow more rapidly in the years following immigration and grow less rapidly thereafter. See also results reported by Green and Worswick (2002) using a different source of data on Canadian immigrants and the literature review by Hum and Simpson (2004a).

Synthetic cohorts cannot solve the problem of selective re-emigration, and they are sensitive to measurement error on the unobserved characteristics that define the cohort. Suppose, for example, that synthetic cohort data show that immigrant earnings display positive duration dependence. Suppose also that the immigrant labor market is Darwinian in the sense that fitter workers survive longer in better-paid jobs. The observed positive duration dependence may be just a statistical artifact induced by survivorship bias. The apparent assimilation effect may simply result from the survival in the labor force of fitter workers[2]. The same applies if remigrants are negatively selected. If fitter workers do not remigrate, synthetic cohort data will create the misleading impression that earnings vary directly with duration. In short, synthetic cohort data may fail to identify the assimilation effect.

1.2 Longitudinal Evidence on Immigrant Assimilation

The obvious solution to this identification problem is to use genuine cohort data, or longitudinal data, rather than synthetic cohort data. By observing the same immigrants over time belonging to the same immigrant cohort, there can be no disagreement about the source of the phenomena observed in the data. We are comparing the same immigrants at different points in time rather than different immigrants at different points in time. Here too sample attrition induced by selective re-emigration is a problem, but at least longitudinal data are informative about the attrition process.

[1] For example, Baker and Benjamin (1994), Bloom, Grenier and Gunderson (1995), Chiswick and Miller (1996, 2004), Grant (1999), Schaafsma and Sweetman (2001), Warman and Woswick (2004) and Aydemir and Skuterud (2005) who use census data for Canada for various years.

[2] Survivor bias may also apply to native workers, especially the younger ones. However, survivor bias is likely to be more pronounced among immigrants whose fitness has yet to be proved.

Unfortunately, longitudinal datasets on immigrants are scarce[3]. Borjas (1989) investigated the earnings dynamics of 1,166 immigrant scientists and engineers in the US over the period 1974-8. He found that there was no evidence of assimilation. It is possible that by 1974 Borjas' immigrants had already completed their assimilation process[4], in which case there would be little or nothing to be learned about assimilation by observing their earnings during 1974-8. In addition his results may suffer from selective movement into and out of scientific occupations since the earnings of Borjas' immigrants are only observed if they worked as scientists or engineers.

Duleep and Regets (1997) used longitudinal data for the United States to show that between 1987-8 the earnings growth of 351 immigrants identified in the rotating panels of the CPS was about 2% greater than for comparable natives, which seems to support the assimilation hypothesis. However, they did not have data on the duration of these immigrants in the United States, so they could not test whether the earnings growth of immigrants during 1987-8 varied inversely with duration as predicted by assimilation theory. Hu (2000) matched longitudinal earnings data from SSA (Social Security Administration) with the Health Retirement Study in the US over the period 1951 - 1991. He found that whereas the synthetic cohort methodology apparently corroborates assimilation theory, this result is overturned by the longitudinal data. On the other hand, Duleep and Dowham (2002) matched the same longitudinal data to the CPS (but over a shorter time period) to show that the immigrant assimilation hypothesis is corroborated by the longitudinal data. They attribute their difference with Hu to the fact that he did not allow the starting wage to vary by cohort.

Panel data covering immigrant earnings during the first three years in Australia (Chiswick, Lee and Miller 2005) and Israel (Beenstock and Ben Menahem 1997) corroborate the assimilation hypothesis over the short-term. Beenstock, Chiswick and Paltiel (2005), however, failed to find evidence of long-term (12 years) assimilation in a large sample of immigrants in Israel. Therefore, evidence of short-term assimilation may not corroborate assimilation theory over a longer time frame. In this paper, we use a longitudinal database that covering the first decade in the host country, and which sheds light on immigrant assimilation over the short and longer terms.

[3] Longitudinal data sources on immigrants have usefully been reviewed by Black et al (2003).
[4] Borjas reports that in 1974 immigrants' average sojourn in the US was as high as 19.2 years.

The Immigrant Database (IMDB) available at Citizenship and Immigration Canada (CIC) contains longitudinal data on tax returns of immigrants matched with immigrant landing files. However, the panel nature of these data has not been taken into consideration in the investigation of immigrant earnings[5]. Another source of panel data on Canadian immigrants is the Survey of Labour and Income Dynamics (SLID) which has been used by Hum and Simpson (2000, 2004b) to show that during the 1990s immigrant earnings did not grow faster than native earnings. It is noteworthy that these results, which reject the immigrant assimilation hypothesis, break the empirical consensus in Canada[6], which has been exclusively informed by the large body of evidence accumulated from applications of the synthetic cohort methodology to a number of independent data sources including census data, SLID and IMDB.

1.3 Panel Data for Quebec

In this paper we investigate the assimilation hypothesis using the ENI (Établissement des Nouveaux Immigrants) database, which comprises longitudinal data for a single immigrant cohort to Quebec. Since immigrants to Quebec are selected on a different basis to Canadian immigrants as a whole[7], the population study group is specific to Quebec.

These data provide a rare opportunity to investigate the assimilation hypothesis during the first decade in the host country, which happen to be the crucial years for investigating immigrant assimilation. Moreover, the data refer to a single immigrant cohort. Matters might have been more complicated had our longitudinal data referred to different immigrant cohorts since, as suggested e.g. by Aydemir and Skuterud (2005), assimilation may depend upon when immigration occurred, especially regarding the state of the business cycle. The main question we ask is: given everything else, including personal characteristics and the state of the labor market in Quebec, do ENI immigrant earnings vary directly with time since migration, and is this relationship concave as predicted by assimilation theory? By following the same

[5] For example, Green and Worswick (2002) estimate earnings functions using IMDB that ignore the panel dimension to the data. Note that IMDB is only available to a small number of authorized users at CIC, hence we do not use it here. Nor do we use LSIC (Longitudinal Survey of Immigrants to Canada) since it was only started in April 2001.

[6] Hum and Simpson (2000) do not control for duration in Canada in 1993, so they do not properly test the immigrant assimilation hypothesis. The wage growth of immigrants with shorter durations should have been greater than those with longer durations.

[7] Quebec is the only province that has an independent immigration policy and gives priority to francophone immigrants.

immigrants over time belonging to the same immigrant cohort, a positive and concave relationship between earnings and TSM would serve as direct empirical corroboration of assimilation theory.

2. Econometric Methodology

2.1 The Model

Assimilation theory predicts that immigrant earnings converge upon the earnings of demographically comparable natives. Let W_{it} denote the logarithm of real earnings of immigrant i (i = 1,2,..,N) in time period t (t= 1,2,…,T), let X_i denote a vector of personal characteristics (such as age, education etc) measured at the time of immigration (t = 0), and let V_{it} denote the log earnings of comparable natives, i.e. conditioned on X_i. Dropping subscript i and expressing time continuously, we write the assimilation model as:

$$\dot{W}_t = \dot{V}_t - \lambda(W_t - V_t) \qquad (1)$$

$$V_t = v + \beta X + \phi t t \qquad (2)$$

Equation (1) states that when the immigrant-native wage gap is zero, immigrant wages grow at the same rate as the wages of comparable native. If however, the wage gap is negative, immigrant earnings grow faster than native earnings. Equation (2) states that native earnings grow at a common rate ϕ, and the wage level depends upon X. For example, more educated workers earn more.

The general solution to equations (1) and (2) is:

$$W_t = Ae^{-\lambda t} + v + \beta X + \phi t \qquad (3)$$

where the initial condition $A = W_0 - V_0 < 0$ is equal to the log wage gap on immigration. Equation (3) states that immigrant wages converge on native wages from below. Empirical evidence suggests (Aydemir and Skuterud, 2005) that the initial wage gap may depend upon observed heterogeneity measured by X. We therefore let $A = A_0 + aX$. Substituting this into equation (3) and restoring subscript i, equation (3) may be reparameterized in terms of average native earnings:

$$W_{it} = (A_0 + aX_i)e^{-\lambda t} + \beta X_i + \overline{V_t} \qquad (4)$$

Equation (4) states that immigrant earnings vary directly with time since migration, they depend upon X in terms of levels and rates of change, and they are proportionate to average native earnings.

It may be shown that variants of equation (4) are obtained if the assimilation process in equation (1) is not first order. For example, if it is second order there will be two roots instead of one, and the convergence process may be quadratic, as originally suggested by Chiswick (1978). Also, equation (2) is unlikely to be static, and wages are likely to be influenced by the business cycle. Indeed, as suggested by Aydemir and Skuterud (2005), the business cycle might affect native and immigrant wages differentially, in which case it would be appropriate to specify the state of the business cycle in equation (4). In this context there are two quite different hypotheses. In the first, what matters is the state of the business cycle on entering the labor market. Aydemir and Skuterud suggest that the long term prospects of natives and immigrants are adversely affected if they happened to enter the labor market during a recession. The second hypothesis predicts that the business cycle affects the relative pay of immigrants as they assimilate, i.e. after they have entered the labor market. Since ENI immigrants belong to a single cohort and entered the labor market at about the same time, it is impossible to investigate the former effect. Therefore, we only investigate the latter effect.

2.2 Panel Data Analysis

The assimilation hypothesis predicts that given X, V and the state of the business cycle (Z) there is a concave relationship between earnings and time since migration (TSM). In equation (5) we use Chiswick's original quadratic specification:

$$W_{it} = \alpha_i + X_i\beta + \theta Z_t + \phi V_t + \gamma TSM_{it} - \eta TSM_{it}^2 + \varepsilon_{it} \qquad (5)$$

Equation (5) is a one-way specific effects model where α_i denotes the individual specific effect. The α's capture immigrant heterogeneity; immigrants who assimilate more successfully have larger α. We do not specify time specific effects because the number of rounds in the panel is small, because TSM naturally varies directly with time, and because equation (5) includes time dependent controls (Z and V). Given everything else, we expect a positive relationship between immigrant earnings and native earnings. Indeed, the model in Section 2.1 predicts that $\phi = 1$. If the immigrant wage gap is anticyclical we expect $\theta < 0$. Following equation (4) TSM may be interacted with X.

We include standard demographic controls in X, including education, age and its square, and region of origin. The role of age here serves a double function because it not only reflects wages over the life-cycle, it also captures the affect of age at

immigration. Schaafsma and Sweetman (2001) point out that older immigrants find it more difficult to assimilate, so given everything else, older immigrants in ENI are likely to earn less. On the other hand, older immigrants are likely to earn more because they are more experienced. Therefore, the net effect of ageing is unclear.

We prefer specifying random effects rather than fixed effects on a priori and practical grounds. Fixed effects are more suitable when the data refer to populations, and random effects are more suitable when the data refer to random samples[8]. ENI is a sample of immigrants who were randomly observed on first entering Quebec in 1989, therefore their specific effects are a random sample of the specific effects in the population. In any case, Hsiao (2003) recommends using random effects estimators in short panels, since estimates of fixed effects are biased when T is small. Finally, ε is a residual error assumed to be normally distributed with constant variance. It is hard to think of reasons why either ε or α should be related to X, Z, V and TSM. Usually these random components capture unobserved cohort effects, but there is only one cohort in ENI. Since the explanatory variables in equation (5) are independent of ε and α, random effects estimation of equation (5) yields consistent parameter estimates. A Hausman test is used to investigate the matter.

An alternative formulation of equation (5) specifies the model in terms of wage differentials between immigrants and demographically similar natives. In this case the dependent variable is defined as $d_{it} = W_{it} - V_{it}$, where V_{it} is the predicted value of native wages conditional upon X_{it}, and d is hypothesized to depend upon TSM and TSM^2. This would be equivalent to equation (5) if the Mincer model used to predict native wages was also valid for immigrants. If, however, it happens to be invalid the estimated wage gap data will be incorrect and the test of assimilation theory will be invalid. By contrast, equation (5) does not impose the restriction that immigrant wages must equal native wages in the long run.

As explained in Section 3, not all immigrants reported positive earnings, and an even smaller number reported positive earnings in all 4 rounds. We decided to use a balanced sample[9], and therefore initially exclude immigrants reporting positive earnings in 3 rounds or less. Although we concentrate upon the balanced sample with $T = 4$, we also experiment with unbalanced samples, i.e. in which $T < 4$. Equation (5) with random effects is estimated by feasible GLS, and the estimates are consistent as

[8] See Greene (2003) p293-4 and Hsiao (1986) p43.
[9] In a balanced sample immigrants enter and exit the sample at the same time.

either T or N tends to infinity. In a short panel such as ENI, it is the latter that is important. Taylor (1980) shows that in short panels FGLS rapidly approaches the Cramer-Rao bound with N, so that N > 50 may be considered as a "large" sample when T = 4. While the sample size in ENI maybe small when compared with samples obtained from censuses and other large datasets, it is sufficiently large to carry out statistical inference since the balanced sample size is 80 and the unbalanced sample size is about 250.

The immigrant assimilation hypothesis predicts $\gamma > 0$ and $\eta > 0$, implying a concave relationship between earnings and duration in Quebec. The efficiency of the estimates of these parameters varies directly with the variance of TSM in the data. TSM naturally varies between rounds of ENI, but as explained in Section 3.2, it also varies within rounds because immigrants entered Quebec over a 6 month period and because they were interviewed over a 6 month period in each round. Had they all arrived on the same day and been interviewed on the same day, matters would have been different, and the variance in TSM would have been smaller.

The first 3 rounds in ENI are equally spaced (1990, 1991 and 1992), but the 4th round took place in 2000. This means that observations on TSM exist more or less continuously during the first 4 years in Quebec and during years 10 and 11, but they do not exist over the range of 4 – 9 years. This dead patch makes it impossible to identify the assimilation curve over this range. Nevertheless, the first 3 observations are informative about immigrant assimilation in the short term, while the last observation is informative about the longer term[10].

2.3 Attrition Bias

Our balanced sample may be self-selected since immigrants who re-emigrated cannot report in all 4 rounds, but especially in Round 4. There may also be self - selection in employment. Members of the balanced sample did not re-emigrate and were employed in all 4 rounds. Fortunately we know which immigrants belong to our balanced sample and which do not. Let $D_i = 1$ if immigrant i belongs to the balanced sample and $D_i = 0$ otherwise. $D_i = 1$ when latent variable $D_i^* \geq 0$ with:

$$D_i^* = \varphi + Y_i\psi + u_i \qquad (6)$$

[10] Retrospective earnings data for 1994 - 1999 were collected in the 4th round, but we have decided not to use them since they are less reliable than data collected in real time.

where Y denotes a vector of variables hypothesized to determine balanced sample membership. If u and ε have a bivariate normal distribution with correlation ρ we may correct equation (5) for sample selectivity by specifying in equation (5) the inverse Mills ratio as a supplementary regressor:

$$\lambda_i = \sigma_\varepsilon \frac{\phi(\varphi + Y_i\psi)}{1 - \Phi(\varphi + Y_i\psi)} \qquad (7)$$

where $\phi(\ )$ is the standard normal density, $\Phi(\ )$ is its cumulative counterpart, and λ has slope coefficient ρ. If $\rho > 0$, immigrants in the balanced sample are positively self-selected. If, however, $\rho = 0$ selectivity does not induce bias in the estimates of the parameters in equation (5). Identification ideally requires that X be a subset of Y, i.e. selection is influenced by instrumental variables that do not directly affect earnings. If such instrumental variables do not happen to be available, i.e. Y = X, then the effect of attrition bias is entirely identified parametrically, by assuming that u and ε have a bivariate normal distribution[11].

3 The Data

3.1 The Sampling Frame for ENI

The population under study consists of immigrants in Quebec who obtained their immigrant visas for Quebec prior to immigration, and who entered Quebec directly via Montreal's two airports (Dorval and Mirabel) or via the Blackpool land crossing with the United States. This population excludes immigrants in Quebec who obtained visas for the rest of Canada and subsequently chose to reside in Quebec, and holders of immigrant visas for Quebec who entered Quebec indirectly[12]. It also excludes holders of immigrant visas for Quebec who chose to live in the rest of Canada. ENI therefore recognizes that immigrants in Quebec with prior experience in the rest of Canada might be different to immigrants who entered Quebec directly.

In the absence of reliable databases from which to sample immigrants[13], the data were collected as follows. 9645 bearers of Quebec immigrant visas, aged 18+ years, entered Quebec via Dorval and Mirabel airports or via Blackpool during mid

[11] In the absence of convincing instruments, we followed Horowitz and Manski (1998) by calculating non-parametric bounds on attrition bias. However, these bounds turn out to be very wide, and we therefore do not report them.

[12] For example, via airports in the rest of Canada or via other land crossings. This means that Asian immigrants are under-represented since many of them enter Canada via Vancouver.

[13] For example the database of Regie de l'Assurance Maladie du Quebec (RAMQ) excludes many immigrants since they do not have regular addresses.

June to November 1989. These 9,645 immigrants constitute the cohort population. All of these immigrants were invited on their arrival to participate in ENI. During this period there were 12,591 immigrants who entered Canada with Quebec visas, but this figure includes immigrants who entered Quebec indirectly.

1,880 of the 9,645 immigrants accepted the invitation to participate in ENI, but only 1000 of the latter were interviewed in Round 1 some 6 - 15 months later. The main causes of this attrition were problems with telephone numbers (488 cases) and people who had left Quebec (169 cases). 64 refused to be interviewed. These data refer therefore to a single cohort of 1000 immigrants who arrived in Quebec in 1989, and who were followed up in 4 face-to-face interviews.

The ENI sample as of Round 1 therefore constitutes 10.4% of the population in the cohort. In contrast to other databases, ENI contains true information on the population. Therefore, whereas with other databases it may be difficult to check their fidelity because accurate data on the population are unknown, we are able to check the fidelity of ENI. To these ends a logit model was estimated for sample participation in which the covariates include variables on the visa application form. According to the logit model[14] women, married people, francophones and bilinguals are more likely to be sampled. However, the sampling error for these parameters never exceeds 10%. Since representative samples of immigrants are particularly hard to construct and validate, ENI constitutes a satisfactory if not perfect database for our present purposes.

Table 1 presents demographic characteristics based on data collected in Round 1. Not surprisingly the immigrants are relatively young and educated. The most important source region is North Africa and the Middle East. Indeed, a quarter of the immigrants came from Libya. For 39% of immigrants the mother tongue was Arabic. The next most frequent mother tongue was French (10%). However, 50.6% of visa applicants and 47.5% of the sample knew French, while 42.8% of visa applicants and 40.2% of the sample knew English. These data naturally underscore the importance of French in the visa selection process. Despite the representation of Arab countries in the data, only 12.6% of immigrants in the sample declared themselves to be Muslims.

3.2 The 4 Interview Rounds of ENI

[14]Details of this model are available on request.

First Round interviews were carried out between June and October 1990 by which time the immigrants had been in Quebec between 6 and 15 months. Round 2 interviews took place between July and November 1991, in which 729 immigrants were interviewed. Round 3 interviews took place between May and August 1993, in which 508 were interviewed. Note that immigrants interviewed in Round 3 are a subset of immigrants interviewed in Round 2. This means that immigrants who were not interviewed in Round 2 were also not interviewed in Round 3. Finally, Round 4 took place during January to April 2000, when 429 were interviewed. In this last round an effort was made to trace respondents lost since Round 1 using records at RAMQ (Régie de l'Assurance-Maladie du Quebec).

Table 2 reports that 25.8% of immigrants were interviewed in all 4 rounds, 33.3% were interviewed in 3 rounds (25% in Rounds 1,2 and 3, and 8.3% in Rounds 1,2 and 4), 22.6% in 2 rounds and 18.3% in Round 1 only. Note that the cell for Rounds 1 and 3 is empty because, as mentioned, ENI did not trace immigrants in Round 3 who were lost in Round 2. Table 1 shows that the sample proportions for the balanced panel (interviewed in all 4 rounds) are quite similar to their counterparts in Round 1. This means that sample selection on observables is not a major problem in the balanced panel. Similar results apply to the non-balanced panels.

Sample attrition at the provincial level is bound to be larger than at the national level. Immigrants who left Quebec to live in another province will constitute attrition in ENI, but not in a national sample (such as SLID or LSIC). Data from IMDB show that 11.3% of immigrants who arrived in Quebec in 1989 were living in other provinces in 1995[15].

Table 3, which reports the number of employed immigrants who reported positive earnings, shows that many immigrants were not continuously employed. In the balanced panel of 258 immigrants only 81 immigrants reported positive earnings in all 4 rounds. 168 reported positive earnings in 3 rounds, of which 148 reported earnings in Round 1, 158 in Round 2, 127 in Round 3 and only 71 in Round 4. 196 reported positive earnings in 2 rounds, of which 160 were in Round 1, 124 in Round 2, 43 in Round 3 and 65 in Round 4. Finally, 214 reported earnings in only one round, of which 137 were in Round 1, 26 in Round 2, 16 in Round 3 and 35 in Round 4. Table 3 also reports the rates of employment in each round. In Round 1 the column sum is

[15]IMDB data show that in 1995 there were 760 immigrants in Quebec who had immigrated to the rest of Canada in 1989.

526 so the rate of employment in 1990 was 52.6%. In Round 2 the rate of employment is 53.4%, in Round 4 it is 52.6% and in Round 4 it is 58.7%. Therefore, the rate of employment was stable during the first 3 years after immigration, bur rose subsequently.

3.3 Balanced Sample Selectivity

We define $D1_i = 1$ if immigrant i was interviewed in all 4 rounds, and therefore is a member of the balanced sample, otherwise $D1_i = 0$. We define $D2_i = 1$ if immigrant i who was interviewed in all 4 rounds was also employed in all 4 rounds, otherwise $D2_i = 0$. In Section 2.3 we defined $D_i = 1$ if immigrant i was employed in all 4 rounds, and zero otherwise. Therefore, $D_i = D1_iD2_i$. Since D2 is naturally nested in D1 (first decide whether to remain in Quebec, then decide whether to be employed) a multinomial selection model is inappropriate.

In Table 4 we report probit models for D1 and D2 estimated separately. The covariates that feature in Table 4 result from applying the general-to-specific methodology[16] applied to the variables in Table 2. The probit model for D1 shows that survivorship in the sample depends mainly on region of origin. Immigrants from the US and Western Europe were more likely to be interviewed in all four rounds, whereas immigrants from Sub-Saharan Africa were least likely to be interviewed in all four rounds. The probit model for D2 shows that employment rates among immigrants interviewed in all four rounds depends mainly on education, with university graduates being the most likely to be employed in all four rounds. Also, immigrants from the US and Western Europe were more likely to be employed in all four rounds.

3.4 Time Series

Monthly time series data from *Statistics Canada* were obtained for the following variables. The consumer price index for Montreal was used to deflate immigrant earnings. Since the vast majority of ENI immigrants chose to reside in greater Montreal, we chose to deflate their wages by the CPI for Montreal rather than the CPI for Quebec as a whole. Native wages are represented by average hourly wages of employees in Quebec. Unfortunately, time series data on native wages for Montreal are not available. Nevertheless, we deflate native wages by the CPI for Montreal. Nor

[16] See e.g. Thomas (1997) for a description of the general-to-specific methodology.

are time series data available on the rate of unemployment in Montreal, therefore we use the rate of unemployment for Quebec.

3.5 The Economic Context

During the early 1990s the economy of Quebec experienced a deep recession together with the rest of Canada. Real wages in Quebec barely grew during 1990 – 1993 and unemployment increased from about 9% in 1989 to almost 15% in 1993-4. The national rate of unemployment increased from 6.2% to 10.5% over the same period. In 2000 real wages in Quebec were only 16% higher than in 1990, recording an average rate of growth of only 1.5% per year, of which the majority was achieved in the second half of the decade. Also, by 2000 the rate of unemployment in Quebec had fallen to about 7%, which was less than half of its peak in the early 1990s. Therefore ENI immigrants arrived at a particularly difficult time. However, matters eased during the second half of the 1990s.

In Figure 1 we plot real wages in Quebec and the real wages of the immigrants in the sample. Note that these sample means are based on all immigrants reporting in each round. We have centered these sample means in July 1990, July 1991, June 1992, and February 2000 when most of the immigrants were sampled. Figure 1 shows that immigrants earned less than natives in 1990 and 2000 and more than natives in 1992. This comparison is obviously unfair because immigrants were more educated than natives. More important than absolute earnings of immigrants and natives is relative earnings between these groups. The assimilation hypothesis predicts that relative immigrant earnings should rise. In terms of Figure 1 this means that immigrant earnings should grow more rapidly than native earnings. This appears to happen during 1990 – 1992, but matters are quite different when we compare 1992 and 2000. This issue is examined econometrically in the next section.

4. Results

4.1 Estimating Equation (5)

We begin by estimating an unrestricted version of equation (5) in which the X vector includes dummy variables for age group, gender, schooling, source region, and language skill. We also include in the specification the log of real wages and the rate of unemployment (V and Z in equation 5) in Quebec[17]. Finally, we specify a polynomial in TSM, which is measured in weeks, although the estimates for γ and η have been multiplied by 52 and 52^2 respectively to annualize them. We use a prioritized general-to-specific specification search methodology in which we give priority to TSM, which is the main variable of interest, and also to native wages (W).

In Table 5 we report estimates of the restricted models in which statistically insignificant variables have been eliminated using the general-to-specific specification search procedure. Model A in Table 5 is estimated using data on a balanced sample of 81 immigrants who reported positive earnings in each of the 4 rounds. Model B is estimated using an unbalanced sample in which 249 individuals have participated in at least 3 rounds.

The main result in both models is that there is no evidence of assimilation. The coefficient on TSM should be positive and the coefficient on TSM^2 should be negative. The opposite tends to be the case; the coefficient on TSM is negative, but for the most part is not statistically significant, while the coefficient on TSM^2 is positive. Both models imply that effect of TSM on earnings during 1990 – 2000 is negative. We tested for higher order effects in TSM, but they were not statistically significant.

Other results in Table 5 are as follows. Men earned about 30% more than women according to both models, and earnings vary directly with schooling. According to both models older immigrants (aged 41+ years) earn more than younger immigrants (26 – 40 years). As mentioned in Section 3.3, this captures two effects, age at immigration and life-cycle effects. According to models A and B francophones earned substantially less than anglophones. Indeed, there is no evidence that knowledge of French or bilingualism is advantageous[18]. Immigrants from US and Western Europe earned substantially more than immigrants from elsewhere. Another surprising result

[17] We also controlled for the rate of unemployment (see Fig 1), but it had no significant effect on immigrant earnings.

[18] Nor do Chiswick and Miller (2003) find that immigrants who speak neither English or French are disadvantaged.

is that immigrant earnings are not, on the whole, affected by native earnings, suggesting that the immigrant labor market is segmented from its native counterpart. Nor are they affected by the business cycle as measured by the rate of unemployment.

While the parameter estimates of models A and B are qualitatively similar, and in several respects quantitatively similar, there are some differences too. For example, francophones earn 45% less than Anglophones according to model A but 25% less according to model B.

While both models are statistically significant, R^2 tends to be small indicating a large degree of unobserved heterogeneity in immigrant earnings, which, however, is broadly in line with results obtained by others. In both models the Hausman test statistic is far from significant, so that, as expected, the estimated random effects are independent of the regressors. Therefore the estimates reported in Table are consistent.

4.2 Testing for Selectivity and Attrition Bias

We use the two models in Table 4 to calculate the inverse Mills ratios using equation (7), λ_{1i} and λ_{2i}, which we add as regressors to model A in Table 5. Their t-values are respectively 0.6 and -1.05, and the Wald test statistic for excluding these selection regressors from model A is F = 0.23, which is far below its critical value. These results suggest that there is no selection bias in model A in Table 5. Although Table 4 establishes that selection is not random at each selection hurdle, the Wald test indicates that the random component of earnings in model A is not significantly correlated with the random component of selection in both hurdles.

4.3 Has the Assimilation Curve Shifted?

In equation (5) the coefficients γ and η, which determine the shape of the assimilation curve, are assumed to be time invariant. However, empirical studies of immigrant assimilation indicate that the shape of the assimilation curve may change. It may become steeper or flatter even if it retains its same overall asymptotic form. The assimilation curve may differ between economies at the same point in time, and it may differ within economies at different points in time. If immigrants face a greater incentive to invest in host country skills they will accumulate these skills more rapidly and the assimilation curve will become steeper. The same will happen if the return to host country skills happens to increase. Indeed, these two effects will tend to reinforce each other.

Dropping subscript i from equation (5), but permitting the parameters of the assimilation curve to vary over time gives:

$$W_t = \alpha + \gamma(t)TSM_t - \eta(t)TSM_t^2 \qquad (8)$$

The assimilation curve steepens either when γ increases or when η decreases. If these parameters are time invariant the rate of change in immigrant earnings is equal to $\gamma - 2\eta TSM_t > 0$. If, however, they are time varying the rate of change of immigrant earnings is equal to:

$$\frac{dW_t}{dt} = (\gamma_t - 2\eta_t TSM_t) + (\gamma_t TSM_t \frac{d\gamma}{dt} - \eta_t TSM_t^2 \frac{d\eta}{dt}) \qquad (9)$$

Even if the first term in brackets happens to be positive, the sign of equation (9) is indeterminate. If equation (9) happens to be zero, it will appear as if assimilation is not taking place, when in fact it is being concealed by the changing shape of the assimilation curve. Or, equation (9) may be negative instead of positive if γ is decreasing and η increasing. In this case it will look as if dissimilation is taking place. Finally, if γ is increasing and η is decreasing immigrant earnings may grow more rapidly over time, in which event it will look as if super–assimilation is taking place.

Longitudinal data are incapable of handling structural change in the assimilation curve, unless auxiliary hypotheses are made about the specific nature of structural change. Ideally, we need to hold calendar time constant, while duration in the host country varies. But this is obviously not possible. Therefore, even longitudinal data cannot unambiguously identify the assimilation effect in the presence of structural change. It may be shown that the identification of structural change is made easier if there are data on successive immigrant cohorts. In our case, however, we have data on only one immigrant cohort, which makes it very difficult to disentangle structural change in assimilation from our longitudinal data.

5. Conclusions

Econometric testing of the immigrant assimilation hypothesis has been hampered by the absence of appropriate longitudinal data on immigrants. This absence has prevented follow-up of individual immigrants during their first years in the host country. In Canada, as elsewhere, researchers have resorted instead to the synthetic cohort methodology as the work-horse to test the immigrant assimilation hypothesis.

Results obtained using the synthetic cohort methodology typically support the immigrant assimilation hypothesis in the sense that during their first years in the host country immigrants narrow the earnings gap with respect to natives, even if they do not manage to close it entirely. The synthetic cohort methodology also typically indicates, in Canada as elsewhere, that more recent immigrant cohorts are less positively selected on unobserved characteristics.

In this paper we have used longitudinal data on a single immigrant cohort in Quebec to test the immigrant assimilation hypothesis. This is most probably the first time that longitudinal data are used to test the immigrant assimilation hypothesis during the first decade in the host country. The immigrants in the sample arrived in Quebec in 1989, and were followed-up four times during their first decade in the host country. Two striking conclusions arise out of the results. First, there is no evidence at all in favor of the immigrant assimilation hypothesis. Immigrant earnings do not initially grow more rapidly than native earnings, and then grow less rapidly as duration in the host country increases. Secondly, immigrant earnings seem to be independent of labor market conditions in the host country as measured by native earnings and the native rate of unemployment.

The former result suggests that putative evidence in favor of the immigrant assimilation hypothesis obtained using the synthetic cohort methodology may be spurious. The latter result suggests that immigrant labor markets are segmented from the native labor market.

Signally, our results are consistent with Hum and Simpson (2000, 2004b), who used longitudinal for Canada as a whole data over a shorter horizon, and who reject the immigrant assimilation hypothesis. While two similar findings hardly constitute a trend, it may be more than just a coincidence that whereas almost every synthetic cohort study corroborates the immigrant assimilation hypothesis, the handful of longitudinal studies rejects it. More longitudinal studies are obviously required before any stronger conclusions can be reached. If, indeed, further longitudinal studies are consistent with our findings and the findings of Hum and Simpson, this would imply that the synthetic cohort methodology is flawed. The flaw lies in the fact that the synthetic cohort methodology compares different immigrants over time rather than the same immigrants at different points in time. If survivors in the labor market are positively selected, the synthetic cohort methodology may simply be picking up survivorship bias rather than genuine immigrant assimilation. Immigrant earnings

appear to grow faster initially, not because immigrants are assimilating into the labor market, but simply because the immigrants whose earnings are observed say 10 years after immigration happen to be more positively selected than the immigrants whose earnings are observed after say only 5 years after immigration. In short, the assimilation effect in synthetic cohort data may be a fiction of the data.

We have shown that structural change in the assimilation curve may conceal underlying assimilation behavior. It is conceivable that in our case structural change during the 1990s made it appear as if assimilation has failed to take place when the converse is true. We cannot rule out this possibility. However, it seems to us that it would be too much of a coincidence if all the effects of assimilation just happened to be totally eclipsed by structural change.

Results for Quebec do not necessarily apply to Canada as a whole because Quebec has its own selection criteria for immigration. Also, the ENI cohort arrived at a particularly difficult time. Perhaps matters might have been different had they arrived when the economy was less recessed. On the other hand, we found no evidence of assimilation after the economy began to recover. Even if immigrant assimilation prospects are permanently affected by the state of the business cycle on their arrival, it would be unreasonable to maintain that there is nothing to be learnt from ENI. Their prospects might have been harmed but not completely destroyed. Finally, despite our efforts to check for sample selectivity, we cannot rule out the possibility that remigration between Quebec and other provinces is concealing evidence of assimilation. If the successful immigrants quit Quebec for other provinces, this would create the misleading impression that assimilation was not taking place in Quebec.

Table 1 Sample Frequencies (%)

	Round 1	Balanced Panel
Females	44.1	43.8
Males	55.9	56.2
Aged 18-25	21.6	20.5
Aged 26-40	53.8	55.7
Aged 41+	24.6	23.8
Married	56.0	
Christians	58.4	
Muslims	12.6	
Buddhists	7.6	
No schooling	0.6	0.1
Primary school	4.1	5.4
Secondary school	15.6	20.3
Post secondary	12.2	12.6
University	25.4	
English	15.8	17.1
English and French	24.4	24.5
French	23.1	50.8
No English or French	30.5	32.9
Latin America	12.4	13.0
USA & W. Europe	14.5	14.5
E. Europe	5.4	4.2
N. Africa & M. East	41.9	50.0
Other Africa	4.5	2.3
Asia & Pacific	21.0	16.5

Table 2 Panel Completion Rates

Round Number	Percent
1	18.3
1 and 2	13.8
1,2 and 3	25.0
1,2,3 and 4	25.8
1 and 4	8.8
1,2 and 4	8.3

Table 3 Employment

Number of Rounds	Round 1	Round 2	Round 3	Round 4	Total
1	137	26	16	35	214
2	160	124	43	65	196
3	148	158	127	71	168
4	81	81	81	81	81
Employed	526	389	267	252	
Employment %	52.6	53.4	52.6	58.7	

Table 4 Probit Models for Balanced Sample Selectivity

	D1	D2
Intercept	-1.0333 (9.921)	-1.7750 (8.008)
Age 26-40	0.15219 (1.733)	
Male		0.3771 (1.941)
Secondary	0.2662 (2.278)	1.4079 (5.491)
College		1.0183 (3.227)
University		1.4760 (6.186)
US and W. Europe	0.3391 (2.483)	0.7970 (3.422)
N.Africa & M.East	-0.3661 (3.613)	
Sub-Saharan Africa	-0.4995 (1.826)	
No French & English	-0.1972 (1.98)	
Log L	-555.6	-122.8
N D = 1	1000 258	258 81

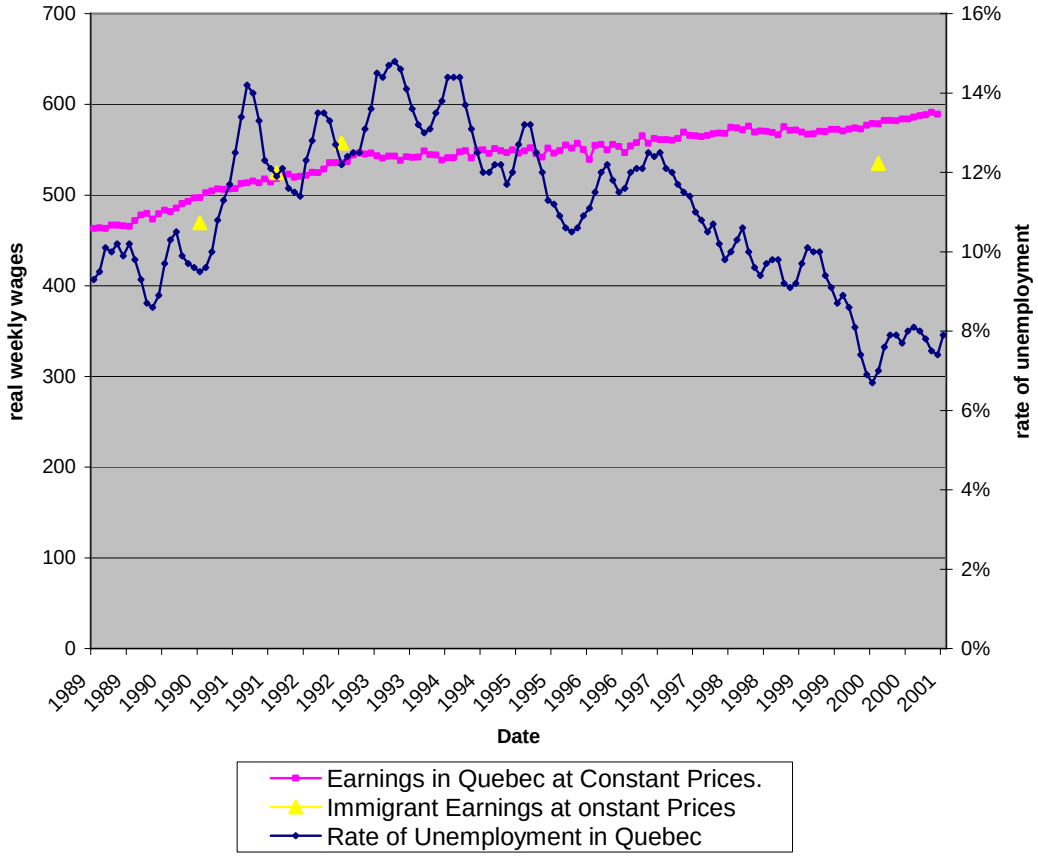
Absolute t-values in parentheses.

Table 5 Panel Regression Estimates of Equation (5)

	A: 4 Rounds		B: 3 – 4 Rounds	
	β	t-stat	β	t-stat
Intercept	-27.95	-1.35	-14.68	-0.99
lnW	5.402	1.62	3.274	1.36
Age 26-40	0.422	3.21	0.418	4.54
Age 41+	0.644	3.83	0.451	3.91
Male	0.261	2.32	0.238	3.05
Secondary school	0.231	1.25		
College	0.524	2.21	0.135	1.27
University	0.556	2.85	0.304	3.48
English & French	-0.323	-2.13	-0.198	-1.84
French	-0.416	-2.67	-0.216	-1.93
No English French	-0.328	-1.91	-0.271	-2.45
US and W. Europe	0.366	3.20	0.348	3.82
TSM	-0.263	1.71	-0.171	-1.56
TSM^2	0.0168	1.88	0.0116	1.81
R^2	0.2085		0.1188	
RMSE	0.2780		0.2873	
N	81		249	
Hausman Test	0.22		1.46	

Dependent variable: logarithm of real earnings. Estimated with random effects. Base: Women, primary school, aged 18-25, English speaker.

Figure 1: Earnings and Unemployment in Quebec



References

Adeymir A. and M. Skuterud (2005) Explaining the deteriorating entry earnings of Canada's immigrant cohorts, 1966 – 2000. *Canadian Journal of Economics*, 38: 641-72.

Baltagi B.H. (2005) *Econometric Analysis of Panel Data*, 3rd edition, John Wiley, New York.

Baker M. and D. Benjamin (1994) The performance of immigrants in the Candian labor market. *Journal of Labor Economics*, 12: 369-405.

Beenstock M. (1996) Failure to absorb: return migration by immigrants to Israel. *International Migration Review*, 34: 950-78.

Beenstock M. and Y. Ben Menahem (1997) The labour market absorption of CIS immigrants to Israel: 1989 – 1994. *International Migration*, 35: 187-224.

Beenstock M., B.R. Chiswick and A. Paltiel (2005) Endogenous assimilation and immigrant adjustment in longitudinal data, mimeo.

Black R., T. Fielding, R. King, R. Skeldon and R. Tiemoko (2003) Longitudinal Studies: An Insight into Current Studies and the Social and Economic Outcomes for Migrants. Working Paper 14, Sussex Centre for Migration Research.

Bloom D.E., G. Grenier and M. Gunderson (1995) The changing labour market position of Canadian Immigrants. *Canadian Journal of Economics*, 28: 987-1005.

Borjas G.S. (1985) Assimilation, changes in cohort quality, and the earnings of immigrants. *Journal of Labor Economics*, 3: 463-89.

Borjas G.S. (1989) Immigrant and emigrant earnings: a longitudinal study. *Economic Inquiry*, 27: 21-37.

Chiswick B.R. (1978) The effect of Americanization on the earnings of foreign-born men. *Journal of Political Economy*, 85: 897-921.

Chiswick B.R.,Y.L.Lee and P.W. Miller (2005) Immigrant earnings: a longitudinal analysis. *Review of Income and Wealth*, 51: 485-.

Chiswick B.R. and P.W. Miller (1995) The endogeneity between language and earnings: international analyses. *Journal of Labor Economics*, 13: 248-87.

Chiswick B.R. and P.W. Miller (1996) Language and earnings among immigrants in Canada: a survey. H. Duleep and P. Wunnava (eds) *Immigrants and*

Immigration Policy: Individual Skills, Family Ties and Group Identities. JAI Press, Greenwich, CT.

Chiswick B.R. and P.W. Miller (2003) The complementarity of language and other human capital: immigrant earnings in Canada. *Economics of Education Review*, 22: 469-80.

Chiswick B.R. and P.W. Miller (2005) *Review of Income and Wealth*

Duleep H.O. and M.C. Regets (1997) Measuring immigrant wage growth using matched CPS files. *Demography*, 34: 239-49.

Duleep H.O. and D.J. Dowhan (2002) Insights from longitudinal data on the earnings growth of US foreign born men. *Demography*, 39: 485–506.

Frenette M. and R. Morissette (2003) Will they ever converge? Earnings of immigrants and Canadian-born workers over the last two decades. Statistics Canada, Ottawa.

Grant M.L. (1999) Evidence of new immigrant assimilation in Canada. *Canadian Journal of Economics*, 32: 930-55.

Green D.A. and C. Worswick (2002) Earnings of immigrant men in Canada: the roles of labor market entry effects and returns to foreign experience. Citizenship and Immigration Canada, Ottawa.

Greene W.H. (2003) *Econometric Analysis*, 5th edition, Prentice Hall, New Jersey.

Horowitz J.L and C.F. Manski (1998) Censoring of outcomes and regressors due to survey nonresponse: identification and estimation using weights and imputations, *Journal of Econometrics*, 84: 37-58.

Hsiao C. (1986) *Analysis of Panel Data*. Cambridge University Press.

Hsiao C. (2003) *Analysis of Panel Data*, 2nd edition. Cambridge University Press.

Hu W-Y. (2000) Immigrant earnings assimilation: estimates from longitudinal data. *American Economic Review*, 90: 368-72.

Hum D. and W. Simpson (2000) Closing the wage gap: economic assimilation of Canadian immigrants reconsidered. *Journal of International Migration and Integration*, 1: 427-441.

Hum D. and W. Simpson (2004a) Economic integration of immigrants to Canada: a short survey. *Canadian Journal of Urban Research*, 13: 46-61.

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

Labeaga J. (1999) A double-hurdle rational addiction model with heterogeneity: estimating the demand for tobacco. *Journal of Econometrics*, 93: 49-72.

Renaud J. (2003) *What a Difference 10 Years Can Make!: The Settlement Experience of Immigrants Admitted to Quebec in 1989*, Les Publications du Québec.

Schaafsma J. and A. Sweetman (2001) Immigrant earnings: age at immigration matters. *Canadian Journal of Economics*, 34: 1066-99.

Taylor W.E. (1980) Small sample considerations in estimation from panel data, *Journal of Econometrics*, 13: 203-33.

Thomas R.L. (1997) *Modern Econometrics*, Addison-Wesley, New York.

Warman C.R.,and C. Worswick (2004) Immigrant earnings performance in Canadian cities: 1981 through 2001. *Canadian Journal of Urban Research*, 13: 62-84.