Global labour markets, return, and onward migration

Abdurrahman Aydemir *Sabancy University*
Chris Robinson *Department of Economics, University of Western Ontario*

Abstract. There is increasing evidence that international migration is characterized by frequent return and onward migration. This has important consequences for the contribution of immigrants to the economy of the host country. Lack of longitudinal data has prevented much analysis of how frequently international migration involves a sequence of location decisions or how long the typical migrant stays in a host country. A newly available longitudinal data set covering all immigrants to Canada since 1980 provides the opportunity to address these issues. The results show that a large fraction of male immigrants who are of working age, especially among skilled workers and entrepreneurs, are highly internationally mobile. JEL classification: J61, J11, J68

Marchés globaux du travail, et flux migratoires de retour au pays d'origine ou de départ vers une nouvelle destination. Les flux migratoires internationaux sont caractérisés par des retours fréquents au pays d'origine mais aussi par des départs vers une nouvelle destination. Voilà qui a des conséquences importantes pour la contribution des immigrants à l'économie du pays d'accueil. Le manque de données longitudinales a empêché jusqu'ici le développement d'analyses de la fréquence de ces migrations internationales qui impliquent une séquence de décisions de relocalisation et de la longueur du séjour de l'immigrant typique dans un pays d'accueil. Une nouvelle source de données longitudinales couvrant tous les immigrants vers le Canada depuis 1980 permet d'analyser ces questions. Les résultats montrent qu'une forte fraction des hommes immigrants en âge de travailler (et particulièrement s'il s'agit de travailleurs spécialisés et d'entrepreneurs) sont très mobiles internationalement.

Aydemir is also affiliated with Family and Labour Studies, Statistics Canada. The authors wish to thank Audra Bowlus, Miles Corak, Garnett Picot, and participants at a workshop on immigration issues held at Family and Labour Division in March 2005 for their helpful discussions. Thanks are also due to referees on an earlier draft. This project is part of the research program of the Family and Labour Studies Division, Statistics Canada and the CIBC Project on Human Capital and Productivity. Email: aydeabd@statcan.ca; robinson@uwo.ca

Canadian Journal of Economics / Revue canadienne d'Economique, Vol. 41, No. 4
November / novembre 2008. Printed in Canada / Imprimé au Canada

0008-4085 / 08 / 1285–1311 /

1. Introduction

Immigration is an important issue in many developed countries. In recent policy debates, two issues are often prominent in the discussion. First is the role that immigration can play in avoiding population decline or stagnation which is implied by the low fertility rates in developed countries. The total fertility estimates for 2002 are well below replacement for many developed countries such as Australia (1.77), Canada (1.60), Germany (1.39), France (1.74), Italy (1.19), Japan (1.42), Sweden (1.54), and the United Kingdom (1.73) and approximately equal to it for the United States (2.07).[1] Immigration is a possible source of population increase to make up for the low domestic fertility rates both immediately in the form of the new immigrants themselves and in the future from the typically higher fertility rates among immigrant populations compared to native born in the developed countries. This role of immigration has received considerable attention (see, e.g., Beach, Green, and Reitz 2003).

The second issue is the role selective immigration can play in raising living standards in the host country by increasing the supply of highly skilled workers. It is often claimed that Canada faces a 'brain drain' of emigrants to the United States, and that skilled immigration can more than make up for this. It is generally recognized that immigrants are not randomly selected individuals from their countries of origin. They differ from non-migrants in terms of both observed and unobserved characteristics. These selection effects come from the behaviour of the migrants themselves and the behaviour of the host country in the selectivity implied by its immigration rules.

The contribution immigrants make to the host country depends on the numbers and skill levels of immigrants – an issue that has been studied extensively; on how long they stay – an issue that has received less attention; and on who stays – an issue that has received attention only recently. However, return or onward migration, particularly who stays, is increasingly recognized as an important issue requiring further study.[2] It is important because it can have a major impact on the net addition made to a host country's population by immigration.[3] It also affects, via the selective nature of the process, the quality of the immigrant stock, and ignoring it results in substantial biases in studies of immigrant assimilation (see Jasso and Rosenzweig 1982; Borjas and Bratsberg 1996).

In addition, evidence on out-migration is important for the design of immigration policy and has implications for the payoff to the costs incurred for settlement and assimilation. Canada and the United States, for example, are major host countries and incur settlement and assimilation costs for particular classes

1 Source: *The World Factbook* (Washington, DC: Central Intelligence Agency 2002); Bartleby.com (2002).

2 While the importance of assessing return migration is increasingly recognized, the phenomenon itself has a long history. Piore (1979), for example, provides estimates of large return migration flows from the United States in the early part of the 20th century.

3 Warren and Peck (1980) drew attention to the importance of the magnitude of return migration for an accurate picture of the net addition made to the United States population by immigrants.

of immigrants. To the extent that large numbers of immigrants return to their country of origin or use the initial host country as a stepping stone to another, the return to these costs will be reduced (see Devoretz and Ma 2002 for discussion of this phenomenon). If immigration policy is designed to attract permanent immigrants, it is important to understand the determinants of return or onward migration. Evidence on trends in out-migration is essential to keep policy up to date. The literature on return migration has raised awareness that migration is not necessarily a permanent move for many migrants. However, return migration itself has often been taken as permanent, if only because of the data limitations in treating it differently. In the increasingly global labour market it may be more appropriate to treat international migration more like internal migration. Individuals may move around from place to place for job-related or other reasons several times in a lifetime. Barriers to the international labour movement have been reduced in recent years. In North America, the NAFTA (North American Free Trade Agreement) provisions have made some type of movement much easier. There has been considerable debate about whether it has stimulated a brain drain of Canadians to the United States, though the literature contains no evidence on whether this is permanent, or part of an increased flow back and forth (see Devoretz and Coulombe 2005). An important new data set for Mexico (the Mexican Migration Project) has stimulated a literature that examines back-and-forth movement from a group of Mexican villages to the United States (Massey, Durand, and Malone (2002), Munshi (2003), Colussi (2004), and Angelucci (2003)).

The previous literature on return migration, briefly reviewed below, has already provided evidence of the total magnitude of return migration in several countries and a start has been made on modelling the process and testing hypotheses regarding the important determinants. There have been large changes in immigration patterns, however, particularly in the source country patterns for migration to developed countries such as Canada and the United States. Changes over time in the characteristics of immigrants and the speed of their assimilation have been the subject of much debate, but despite the strong connection, changes in the make-up of return or onward migration have not been investigated. The lack of data has also prevented much analysis of whether international migration is increasingly composed of a sequence of moves, not a once-for-all move with possibly a return should the move prove to have been a mistake. A newly available tax-based longitudinal data set, covering immigrants to Canada since 1980, provides the opportunity to address these issues.

The analysis in the paper uses two different measurement approaches to study return and onward migration. One uses landings records, which record all immigrant arrivals to Canada, together with Canadian Censuses that provide information on the number and characteristics of immigrants at a point in time after their arrival. Based on the information provided by these repeated cross-sections, out-migration rates, and variation in these rates by characteristics such as country of origin, the macroeconomic environment at arrival is investigated. The other approach uses the landings records and the longitudinal tax-filing information

and infers out-migration by long term absences from the tax files. The two methods provide very similar estimates of out-migration rates and the same qualitative results regarding variation in rates by country of origin and phase of the business cycle at arrival. This suggests that a substantial part of the absences of immigrants from the tax records is associated with not being in the country.

The annual frequency of tax-filing information as opposed to Census information that is available every five years also allows a finer analysis of the time path of these absences, and its association with important immigrant characteristics such as visa class and language ability at arrival, which are not available in Census files. The longitudinal approach using tax files provides evidence on the factors that determine how long immigrants remain in Canada in their first spell in the country and the extent to which there may be reappearances after a spell of absence. The length of the first spell is particularly important in assessing the total or life-cycle contribution of a new immigrant arrival to the population or labour force.[4]

The plan of the paper is as follows. In section 2, the previous research on return migration is briefly reviewed. It highlights the importance of return migration and the variation in return migration by source country documented in the previous literature. In section 3, a variety of evidence is presented on the factors affecting the length of the first spell of residence in the host country or, conversely, the extent of return or onward migration. This evidence shows a large amount of return or onward migration and substantial variation in magnitudes over time and by various characteristics, including class of immigrant and source country. For the male cohorts landing around the 1990–1 recession, a substantial fraction left the country within a relatively short period of time. The migrants from source countries such as Hong Kong or the United States had particularly short stays, as did those entering under the business-class or skilled-class category for source countries in general.

Section 3 also examines the time path of the exits, estimating hazard functions under a variety of specifications. There is a very clear pattern of particularly high hazard rates in the first year that subsequently fall rapidly to quite low levels. This occurs for all visa classes, indicating that most of the variation in length of stay is due to the differences in the hazard rates in the first year. For males entering at ages 25–45, the firs-year hazard rate for the business class, for example, is 0.311, compared with 0.167 for refugees and 0.206 for the family class. However, after this period, the hazards are very similar; the gaps are only 0.01 to 0.02 between classes. The hazard functions for the migrants from Hong Kong reflect the very strong influence of the handover to China. In particular, the hazard function for the 1980 to 1984 cohorts, who migrated largely before the handover discussions, is very similar to those for other source countries outside of North

4 While the length of the first spell is particularly important, the data show multiple moves among immigrants in Canada, confirming that neither initial nor return migration is permanent, but both are the kind of 'temporary' phenomenon observed in worker movement across job locations in internal migration.

America. However, for the 1990 to 1994 cohorts there is a dramatic increase in the first-period hazard relative to the earlier cohorts.

The results in section 3 are derived from tax-filing behaviour rather than direct evidence on residence. Tax-filing behaviour is of great interest in itself, in that it allows the assessment of the life-cycle contribution of newly landed immigrants to the labour market and tax payments. However, the relation to residence is also of interest, and section 4 presents a comparison with a Census-based approach that directly measures residence. The Census-based approach is much more limited in what can be done because of its repeated cross-sectional, rather than longitudinal, nature and the lack of information on characteristics such as visa status. However, using a synthetic cohort approach with the Census provides an alternative way of estimating a subset of the results obtained in section 3 using tax-filing behaviour. A comparison of these estimates provides substantial corroboration of the results reported in section 3. Some conclusions and an outline of future work are given in section 5.

2. Previous research on return migration

Empirical studies of out-migration of immigrants have been, until recently, hampered by the lack of longitudinal data on immigrants that would directly identify leavers. Many studies use repeated cross-sectional data, such as a national census, and focus on obtaining estimates of the amount of out-migration. Warren and Peck (1980) for example, use the U.S. censuses for 1960 and 1970, together with Immigration and Naturalization Service (INS) statistics on aliens admitted for permanent residence, to estimate total emigration in the period 1960 to 1970 and the fraction of immigrants admitted between 1960 and 1970 that had emigrated by 1970. Their estimates show that more than 1 million foreign-born persons left the United States during the decade. They conclude that the 'implications of substantial foreign-born emigration for United States population growth are obvious. Rather than 400,000 persons being added to the United States population each year (the level of net immigration currently used by the Census Bureau in its population projections), the real addition is probably closer to 250,000 each year' (Warren and Peck 1980, 79).

A related Census-based approach was used for Canada by Lam (1994). This method relies entirely on Census data, using a synthetic cohort approach. The estimates, based on the micro data files for the Censuses of 1971 and 1981, show a substantial amount of return or onward migration. In addition, by using the individual characteristics available in both censuses, the covariates associated with return migration were investigated. A result in common with the literature for the United States is the substantial variation by country of origin.

Jasso and Rosenzweig (1982) were able to use the U.S. Alien Address Report Program, which simulates a longitudinal research design. Combining this with

mortality records and survey data, Jasso and Rosensweig (1982) obtain estimates of cumulative net rates of emigration for the 1971 legal immigrant cohort at about eight years after entry. An important feature of these estimates is that they were obtained by country of origin, which permits consideration of some, possibly very important, selection effects in emigration. Like the authors of earlier literature, Jasso and Rosenzweig (1982) estimate large emigration rates: 'The emigration rate for the entire cohort could have been as high as 50 percent. Canadian emigration was probably between 51 percent and 55 percent. Emigration rates for legal immigrants from Central America, the Caribbean (excluding Cuba), and South America were at least as high as 50 percent and could have been as high as 70 percent. On the other hand, emigration rates for Koreans and Chinese could not have exceeded 22 percent' (Jasso and Rosenzweig 1982). Borjas and Bratsberg (1996) report a similar pattern of out-migration rates by country of origin.

The major disadvantage of the methods used in this literature to estimate return migration is that it cannot examine migration at the individual level because of the reliance on Census data to identify leavers. Since there is no individual link from the administrative data to the Census, individual characteristics of the leavers cannot be identified; only averages can be. Other disadvantages follow from the fact that the absence of this link requires a variety of adjustments to the Census figures to make sure that they are comparable to the administrative records cohort. These include Census enumeration problems, illegal immigrants, mortality issues, and Census respondent recall of their immigration date many years after the fact. The impossibility of an individual-level analysis from this method means that many important questions regarding the contribution of immigrants cannot be answered.

The Mexican Migration Project (MMP) provides longitudinal data to examine the patterns of movement of individuals between a group of Mexican villages and the United States. An early example of the use of this data set is Massey, Durand, and Malone (2002). This data set has provided a great deal of information on the temporary nature of much of this migration, and the information on potential wages in both locations over time has presented an opportunity to model this back-and-forth movement. Recent papers by Munshi (2003), Colussi (2004), and Angelucci (2003) are examples of this modelling effort. The MMP data are useful for understanding a particular example of back-and-forth international migration for the Mexican case, but are limited to a particular source country. In addition, they cannot provide a picture of return or onward migration in total for the host country.

Constant and Massey (2002) examine return migration using the German Socio-Economic Panel, which provides a source of longitudinal data beginning in 1984 on about 3,000 legal immigrants. This study focuses on examining selectivity in return migration and provides evidence on the nature of this selectivity. Like the Census-based studies in Canada and the United States, country of origin is very important. The largest immigrant population is from Turkey and these immigrants were much less likely to return than immigrants from the European Union. Studies based on national longitudinal panels can provide information on

return and repeat migration, which is not possible with the Census-based studies. They provide clear evidence that return and repeat migration are important phenomena. However, they are typically limited by relatively small sample sizes for immigrant populations, especially at a disaggregated level.

3. Data sources

The data sets that form the primary basis for the empirical analysis are the Landings Records (LIDS) and the Longitudinal Immigration Data Base (IMDB). The LIDS file is a rich source of immigration data, recording all landings in Canada from 1980 onward and containing a wide variety of personal, demographic, and program data, including the immigrant category. The IMDB matches the LIDS with information from the tax records, thereby providing a longitudinal earnings record for immigrants that remain in Canada after landing.[5] The longitudinal aspect of IMDB is especially valuable for a variety of important immigration-related questions.

The IMDB provides information on the tax behaviour of immigrants who landed since 1980.[6] However, whether they work or reside in the country in any subsequent year must be inferred from the tax records. The IMDB tax records show intermittent filing for many immigrants. The IMDB records include many immigrants who have landed and filed taxes who go on to have periods of non-filing for up to four years, and yet who subsequently recommence filing. From the IMDB it is not possible to know if these individuals left the country for a period and subsequently returned. They may have been permanently residing in the country and had intermittent periods of non-filing. This intermittent nature of tax filing is of interest in itself. If the lack of tax filing does indicate absence, as an increasingly global labour market might suggest, then it will allow the calculation of the contribution of a given cohort of immigrants to the workforce. If it indicates instead just the absence of paying taxes, it will allow the calculation of the contribution of the cohort to taxes.

The IMDB provides a unique opportunity to examine the life-cycle profile of recent immigrants regarding their residence or tax-filing behaviour in Canada, in the context of a new global labour market where mobility among immigrants is increasing. The analysis in this section examines the determinants of the interval between landing and the first consecutive four-year spell of non-filing. The analysis is based on the tax-filing behaviour; hence residence behaviour inferred from tax files may include some bias if, for example, some individuals do not file tax

5 To be included in the IMDB, an individual has to file at least one tax return after landing.

6 Landing refers to the process whereby immigrants arrive in Canada with their landings documentation that starts their permanent resident status. Temporary residents who are already residing in Canada and are accepted as immigrants have to leave the country and re-enter with their landing documents for their permanent resident status to take effect. Thus, for some individuals permanent residence may start following a period of temporary residence.

returns for four or more consecutive years, although they are in the country.[7] However, comparison with Census data that directly measures residence suggests that this bias is likely to be small.[8] Recent international migration is viewed as being sufficiently influenced by the global labour market to treat all 'spells' in a given country as analogous to job spells. In an internal migration setting, many individuals have intermittent spells in various jobs or occupations. All moves are potentially temporary and all jobs may be returned to. The focus of interest in this setting is an examination of the determinants of the spell lengths. In this section, the same approach is used with international migration.

The IMDB contains a rich set of characteristics for all immigrants such as visa class, education level, and language ability. The IMDB makes it possible to examine how immigrant life-cycle profiles are related to the various characteristics that are used to shape immigration policy. In addition, it is possible to distinguish immigrants according to source region and the immigrant class under which they are admitted, such as skilled worker or refugee. The evidence provided below shows that immigrants admitted from different regions and under different visa classes have very different life-cycle profiles of residence in Canada.

4. A Duration model for immigrant retention

A duration analysis was conducted on the role of the covariates of interest in determining survival of immigrants landing in Canada to the end of the first spell of residence, as indicated by tax-filing behaviour. For the purpose of this analysis, a spell is referred to as spell of residence. As noted above, tax filing is not equivalent to residence, but a comparison with census methods reported in section 4 below suggests that a substantial portion of the absence from tax files for immigrants is absence from the country. From the point of view of an immigrant's contribution to a host country, the spells of tax filing are of interest in themselves. However, the interpretation of the coefficients of the discrete time duration model reported in this section as reflecting the determinants of residence spells is subject to the caveat that the data refer to tax filing. The data are such that, while the landing date is known precisely, that is, no left censoring, the date

7 Defining absences based on shorter spells, such as two years, shows that this type of shorter absence is more likely to refer to short-term absences from the workforce as more than half of the individuals who experience such spells reappear in the data. On the other hand, a spell definition longer than four years provides results very similar to four-year definition as people who are absent from tax files for four years are very likely to be absent an additional one or more years. Increasing spell-length definition also requires dropping later cohorts who were not in the country long enough – hence, loosing some information. Given very similar estimates between the Census method and tax filing method based on the four year definition, we prefer this definition.

8 A Census-based approach in section 5 that directly measures residence shows that non-filing behaviour studied in sections 3 and 4 is mostly associated with absence from the country rather than being in the country but not filing. As discussed later, for example, for the 1981 cohort 1-year and 20-year Census retention estimates are 76.2% and 67.6%. The estimates from tax-filing method for the same periods are 78.6% and 65.5%.

of the end of the spell is known only within an interval because of the annual nature of tax filing, or is not known at all of because of right censoring.[9]

This presents a problem for calculating empirical survival and hazard functions using the standard Kaplan-Meier method, where survival times are treated as observations on a continuous variable. Life table analysis can be used to produce empirical survival and hazard function estimates when the survival data have to be grouped into intervals. The procedure is as follows:

Let τ_i be the individual failure or censoring times aggregated into K time intervals,

$$\begin{aligned} I_k &= [t_k, t_{k+1}), \qquad \mathrm{k} = 1, 2, \ldots, \mathrm{K} \\ d_k &= \text{number of failures in interval } I_k \\ m_k &= \text{number of censored spell endings in interval } I_k \\ N_k &= \text{number of persons at risk of failure at start of } I_k. \end{aligned} \tag{1}$$

The product limit estimate of the survivor function is defined as

$$S_k = \prod_{k=1}^{K} \left(\frac{n_k - d_k}{n_k} \right) \tag{2}$$

where $n_k = N_k - m_k/2$ is the adjusted number at risk at the start of the interval.

While this procedure can deal with the fact that the survival data have to be grouped into intervals, an exact implementation requires the initial point to be the same. Using data on all immigrants arriving in December allows all the intervals to be one year. The sample size can be increased by pooling several months and, as an approximation, assuming a common arrival date. The empirical survivor function to the first absence for the December sample is given by figure 1, which estimates a survival rate of about 78% by the end of the first year after arrival and just above 63% by 20 years after arrival. Most of the first absences take place within the first year. This is also reflected in the empirical hazard rate presented in figure 2. The hazard rate declines sharply in the first year and more gradually thereafter. These figures help to determine the basic shape of the hazard function, facilitating the development of an appropriate specification for the proportional hazards regression model employed below.

Table 1 presents the estimates from a multivariate analysis of duration to the first absence using a discrete time (grouped data) proportional hazards regression framework. The included individual characteristics, all recorded at the time

9 The landings data record the day, the month, and the year of arrival. Therefore, it is possible, for example, to distinguish immigrants by month of arrival.

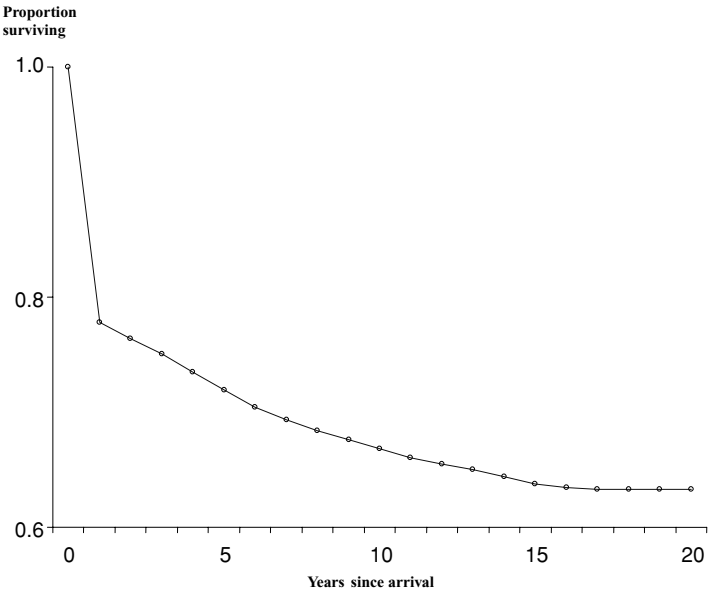

FIGURE 1 Survival functions using Life Tables – all immigrants
SOURCE: Authors' calculation based on life table analysis

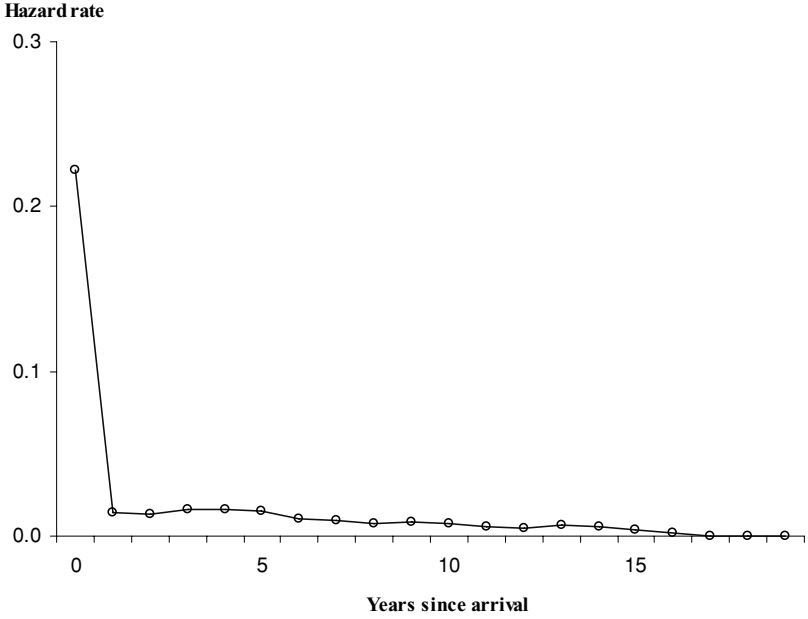

FIGURE 2 Estimated hazard rate from Life Tables – all immigrants
SOURCE: Authors' calculation based on life table analysis

TABLE 1
Discrete time duration model for landing cohorts 1980 to 1996: males, aged 25 to 45 at arrival

	Coefficient	Standard error	Probability of absence
1981 cohort	−0.106	0.092	0.543
1982 cohort	−0.176	0.106	0.518
1983 cohort	−0.149	0.114	0.528
1984 cohort	−0.005	0.111	0.579
1985 cohort	−0.252	0.113	0.491
1986 cohort	−0.455	0.102	0.424
1987 cohort	−0.147	0.093	0.528
1988 cohort	−0.059	0.088	0.560
1989 cohort	0.162	0.084	0.641
1990 cohort	0.175	0.082	0.646
1991 cohort	0.084	0.080	0.612
1992 cohort	−0.328	0.086	0.466
1993 cohort	−0.452	0.094	0.425
1994 cohort	−0.378	0.094	0.449
1995 cohort	−0.495	0.096	0.412
1996 cohort	−0.201	0.088	0.509
Non-university post-secondary	−0.046	0.039	0.564
BA degree or above	0.073	0.037	0.608
French	0.110	0.068	0.622
English & French	0.175	0.059	0.646
Neither English nor French	−0.154	0.039	0.526
Married	−0.193	0.033	0.512
Widowed, separated, divorced	−0.017	0.093	0.575
Business class	0.275	0.065	0.682
Skilled class	0.194	0.043	0.653
Assisted relative class	0.044	0.055	0.598
Refugee class	−0.144	0.050	0.529
Other admission category	−0.009	0.061	0.578
Age at arrival 30 to 34 years	0.086	0.037	0.613
Age at arrival 35 to 39 years	0.122	0.043	0.626
Age at arrival 40 to 45 years	0.099	0.049	0.618
Europe	−0.619	0.071	0.374
Asia, excluding Hong Kong	−0.471	0.068	0.419
Hong Kong	0.302	0.072	0.692
Middle East	−0.537	0.087	0.399
Africa	−0.520	0.083	0.404
Caribbean and Guyana	−0.828	0.092	0.316
South & Central America	0.112	0.083	0.623
Oceania and Australia	−0.401	0.144	0.442
Constant (1980)	−3.880	0.100	0.581
Number of observations		111,822	
Log likelihood		−14907.03	

NOTES
Pooled October-December sample. The model includes a piecewise linear baseline hazard for 1–12 months, 13–24 months, 25–36 months, 37–60 months, and after 60 months.
SOURCE: Calculations by authors based on the Landing Records (LIDS) and the Longitudinal Immigration Database (IMDB) data

of landing, are education level, language ability, marital status, admission category, age at arrival, and source country. All characteristics are entered as dummy variable sets to minimize functional form issues. The omitted category is a single individual with no post-secondary education, fluent in English, admitted under the family class with age at landing 25 to 29, and arriving from North America in the 1980 landing cohort.

The model used is based on Prentice-Gloeckler (1978) and Meyer (1990). This model, called a 'complementary log-log' (cloglog) model, can be interpreted as the discrete time model corresponding to an underlying continuous time proportional hazards model. The underlying process is assumed to be continuous, but the survival time data are recorded in bands (groups). Suppose that there are N individuals ($i = 1, 2, ..N$) each entering a state (landing in Canada) at time $t = 0$. The instantaneous hazard rate function (corresponding to the first absence of four consecutive years from the tax files) for person i at time $t > 0$ is assumed to take the proportional hazards form:

$$\theta_i(t, X) = \theta_0(t) \exp(X_{it}\beta), \tag{3}$$

where $\theta_0(t)$ is the baseline hazard function, X_{it} is a vector of covariates summarizing observed differences between individuals at time t, and β is a vector of parameters. For simplicity, assume that all intervals are of unit length (e.g., a month), so for each person i the recorded duration corresponds to the interval $[t_i - 1, t_i)$.

The individuals who are recorded as having left the state (i.e., disappeared) are identified by the censoring variable $z_i = 1$, while those who are still remaining in the state contribute to the right-censored spell data and are indicated by $z_i = 0$. The likelihood function for this problem can be written in terms of hazard functions as

$$\log L = \sum_{i=1}^{n} \left\{ z_i \log \left\{ h_{t_i}\left(X_{it_i}\right) \prod_{s=1}^{t_i-1} [1 - h_s(X_{is})] \right\} + (1 - z_i) \log \left\{ \prod_{s=1}^{t_i} [1 - h_s(X_{is})] \right\} \right\}, \tag{4}$$

where the discrete time hazard in the jth interval is given by

$$h(t_j, X_{ij}) = 1 - \exp[-\exp(x_{ij}\beta + \gamma_j)], \tag{5}$$

and where γ_j refers to the baseline hazard.[10]

10 The interval specific parameter may differ in each interval, allowing for a non-parametric duration dependence. If several intervals are assumed to have same hazard rather than a different hazard for each interval, a piece-wise constant baseline hazard is obtained. Baseline

Let $y_{it} = 1$ if person i exits the state during the interval $[t - 1, t)$; $y_{it} = 0$, otherwise. Then the log likelihood can be rewritten as

$$\log L = \sum_{i=1}^{n} \sum_{j=1}^{t_i} \left\{ y_{ij} \log h_j(X_{ij}) + (1 - y_{ij}) \log[1 - h_j(X_{ij})] \right\}. \tag{6}$$

The interpretation of the coefficients, β_k, is the proportionate change in the hazard θ, given a one-unit change in X_k. The exponentiated coefficients, $\exp(\beta_k)$, give hazard ratios, allowing a comparison of hazard rates with the reference group. Given the life table estimates of high hazard rates in the first year followed by a sharp decline, a piece-wise constant baseline hazard specification was adopted. The model uses a piecewise linear baseline hazard for 1–12 months, 13–24 months, 25–36 months, and after 60 months.

5. Results from the duration analysis

The predicted survival function estimated from the cloglog model with the piecewise constant baseline hazard specification is very similar to the empirical survival function reported in figure 1, suggesting a good overall fit for the duration model. The first two columns of table 1 present the estimates and standard errors of the coefficients in the cloglog model. These measure the effect on the hazard of ending the first spell. Each row in the last column reports the probability that the first spell ended within 20 years for an individual with the indicated row characteristic and the reference group values for all other characteristics.

5.1. Visa class, source country, and other immigrant characteristics

There are marked differences for visa class and region of origin that are clearly apparent from the estimated coefficients in the duration model. Canada's immigration system admits individuals on the basis of family ties; a refugee process; or a points system in a variety of immigrant classes, each with its own criteria for admission: business class, skilled class, and assisted relative class. These classes have substantially different implications for the length of stay in Canada. The coefficients on the business and skilled classes are significantly higher than for the family class; the coefficient for refugees is significantly smaller. The length of stay implied by these coefficients can be quite short, especially for those entering under the business or skilled-worker class. Examination of the shape of the hazard function shows that this is primarily due to particularly high hazard rates in

hazard can also be specified parametrically, allowing a Weibull model or an nth-order polynomial. Note that the specification of the hazard rate implies

$$\log(-\log[1 - h_j(X)] = X\beta + \gamma_j, \tag{9}$$

hence the name cloglog model.

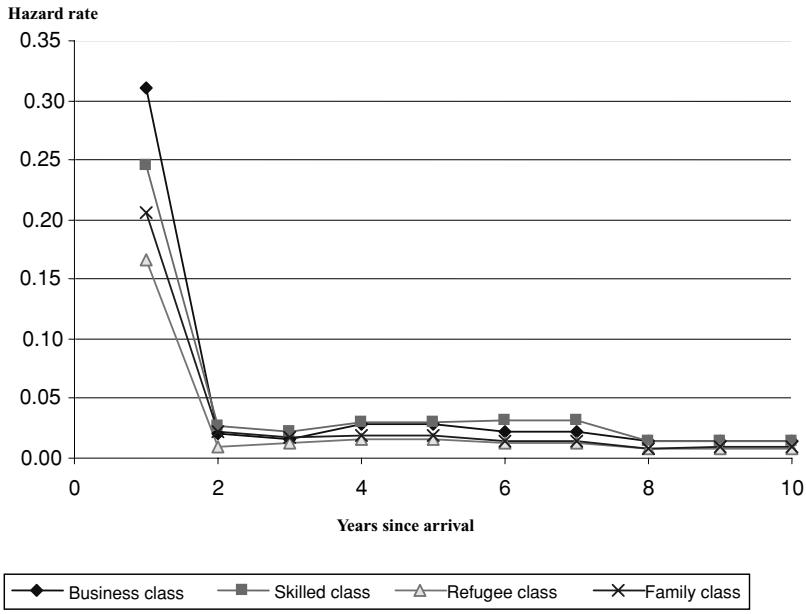

FIGURE 3 Discrete hazard rates by visa class: separate visa class samples
SOURCE: Authors' calculation based on duration model estimates by visa class

the first year after landing. Figure 3 plots the estimated discrete hazard by visa class using the cloglog variant of the proportional hazards model with a piecewise linear hazard.[11] It is clear that, for all visa classes, the hazard rates are high in the first year, but subsequently fall rapidly to quite low levels. This indicates that most of the variation in stay length is due to the differences in the hazard rates in the first year.[12] The first-period hazard rate for the business class, for example, is 0.311 compared with 0.167 for refugees and 0.206 for the family class. After this period, the hazards are all very similar; the gaps are typically only about 0.01 to 0.02 between classes.

The return migration literature for the United States shows strong differences by source country. This is clearly apparent in table 1 for Canada, holding constant other important covariates. The omitted group is North America. Differences from North America in the estimated coefficients for the other cource countries are generally highly statistically significant and often of large magnitude. All,

11 The plots in this section are for an individual with average values for the other characteristics used in the duration model. The model is estimated separately by visa class. Pooling across visa class samples produces an almost identical plot.

12 The estimated hazard rates are discrete, owing to the nature of the data. The rates presented here are for 12-month intervals. It would be interesting to have more detailed information on the path of the hazard within the first 12 months. Unfortunately, the nature of the data makes it difficult to get useful hazard information below the level of the 12-month interval. See the appendix for a discussion of these problems.

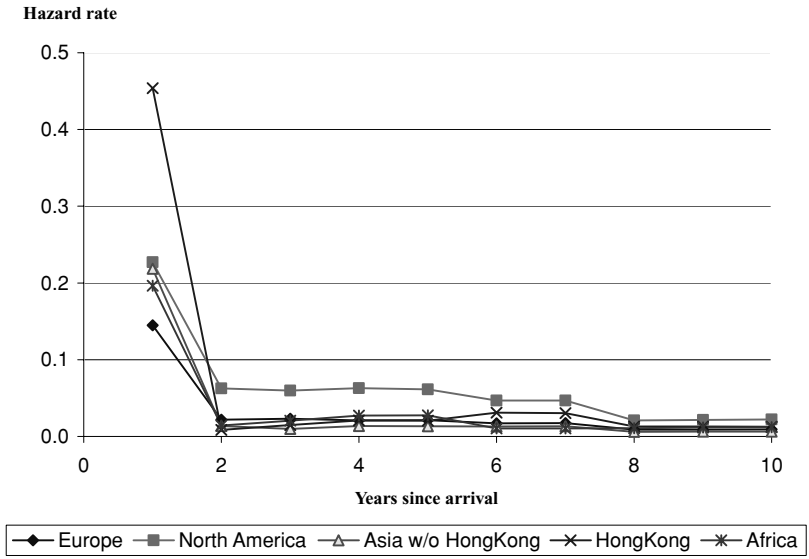

FIGURE 4 Discrete hazard rates by source region: separate source region samples
SOURCE: Authors' calculation based on duration model estimates by source country

except for the special cases of those from Hong Kong and those from South and Central America, show coefficients that are significantly smaller than for North America, implying stays that are much longer for these groups.[13] These major differences by source country suggest that any variation in the relative weights of source countries over time will have important implications for the permanence of the migration and the overall contribution of immigration to the labour market.

Differences in the shape of the hazard function by source country are shown in figure 4, which plots the discrete hazard functions estimated separately by source country. It is clear from these estimates that the shape of the hazard is different for North America and that Hong Kong is an outlier in terms of the magnitude of the difference in the first-year hazard values. The immigrants from the United States show a clear tendency for the hazard rates to remain quite high for many years after entry. The other source countries show a drop to very low levels after the first year. This is consistent with immigration from the United States being particularly similar to internal migration. That is, the pattern is consistent with a relatively low cost of migration and a relatively high probability of continuing movement in response to movement in wage differences or career demands.

Hong Kong is a special case because of the effects on migration to Canada, owing to the impending handover to China that finally took place in 1997. The right of residence in Canada became particularly attractive and many business

13 The special case of Hong Kong is examined below in more detail.

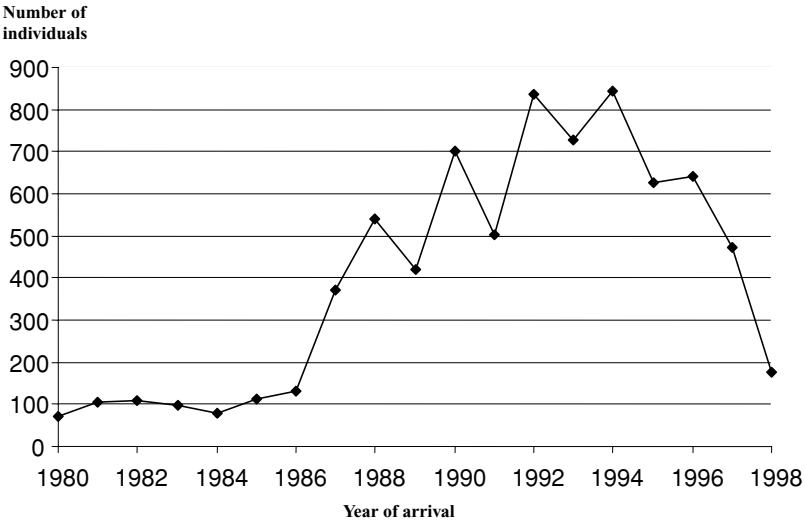

FIGURE 5 Landings of males aged 25 to 45 from Hong Kong, 1980 to 1996
SOURCE: Authors' calculation based on landings data

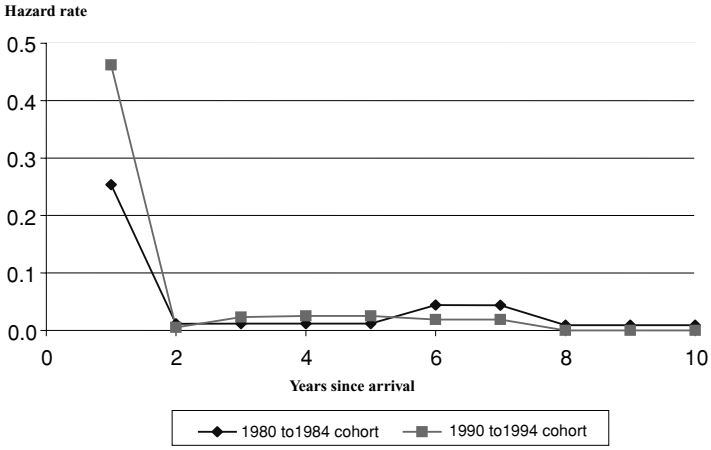

FIGURE 6 Discrete hazard rates for Hong Kong by cohort
SOURCE: Authors' calculation based on duration model estimates for Hong Kong

people acquired this right through immigrating under the business-class provisions. Figure 5 shows landings in Canada from Hong Kong from 1980 to 1996. It is clear that there were large increases after 1984, when the Joint Declaration was signed between Britain and China. Figure 6 compares the estimated hazard rates for the 1980–4 and the 1990–4 landings from Hong Kong. After the first year, the hazard rates are the same, but the first-period hazard rates are dramatically

different. For the 1980–4 cohorts, the hazard function is very similar to the function for the rest of Asia (figure 4). However, for the 1990–4 cohorts, while the hazards look the same after the first-period, the first-period hazard rate is much higher than for the rest of Asia and the earlier Hong Kong landings. This suggests that a significant fraction of these landings may have been stimulated by the attractiveness of establishing citizenship, rather than the prospect of a long-term stay in Canada.[14]

The other immigrant characteristics recorded in the landing records are education, marital status, and age at landing. There is evidence of shorter stays for those who are single, bilingual, and with a BA degree or above. The coefficients for bilingual and a BA degree and above are both statistically larger than zero, while the coefficient for married is significantly negative. The implied variation in length of stay is less than the case for visa class and country of origin. However, the effects are economically significant. Age at landing shows the smallest hazard for those landing at the youngest age. There is a statistically significant effect, but the magnitude is modest.

5.2. Business cycle effects

An important issue in immigration policy is whether fewer immigrants should be admitted during recessions. One concern is that a poor labour market on entry may lead to high unemployment for new immigrants, which may stimulate return migration. A full discussion of this issue is beyond the scope of the present paper. However, the pattern of coefficients on the landing dates in table 1 are consistent with some business-cycle effects. The lowest probability of leaving for the earlier cohorts is the 1986 cohort. This cohort entered at a time when the economy had largely recovered from the 1982 recession and when the double-digit unemployment rates from 1982 to 1985 had finally given way to single digits; this cohort's entry was also sufficiently far from the business cycle trough in the early 1990s to be largely unaffected by this recession. As the cohort entry date approaches this recession, the probability of leaving rises rapidly, peaking at a level that is 50% higher than the 1986 level. However, there is a surprisingly rapid decline in the probability of an absence after the 1991 landing cohort at a time when the unemployment rate had not begun to fall. In fact the clearest summary description of the relation between the probability of absence and the business cycle is given by a simple regression of the probability of leaving (P) on the change in the unemployment rate in the two years immediately following landing (ΔU).

14 The specification for table 1 imposes common visa class effects across source countries. Given the importance of business-class migrants from Hong Kong, the model was re-estimated without Hong Kong to test the extent to which the large-business and skilled-class effects may have been due in part to very large effects from the Hong Kong sample. In fact, when the Hong Kong observations are dropped, the positive coefficients for business and skilled classes are even larger than those reported in table 1.

The resulting estimates are as follows (standard errors in parentheses):

$$P = \underset{(.014)}{0.521} + \underset{(.006)}{0.021}\Delta U \qquad R^2 = 0.424.$$

The 24-month unemployment changes are about 2 to 3 percentage points at their most rapid during recessions and recoveries. For the reference group in table 1, the estimated coefficient from the above regression indicates that entry during the recession ($\Delta U = -3$) is associated with probability of leaving of 0.581 compared with 0.461 for entry at recovery ($\Delta U = +3$).[15]

While these results suggest some possible business-cycle effects, they should be interpreted with caution for a number of reasons. The annual time series permits only a small number of degrees of freedom. This constrains the analysis of other factors that are changing over time. These include the size of the cohort and the immigration policy. There is substantial variation in cohort size and in visa-class shares over this period. Moreover, immigration policy, especially in the early years, discouraged entry during recessions and may have resulted in selection effects in terms of the type of immigrant admitted.

5.3. Implications of selective survival

The large differences in survival by characteristics such as visa class and source country indicate that survival is selective. Since many of these characteristics are related to income, the selective nature of survival has implications for skill levels and earning capacity of the immigrants who remain. While a full analysis of this issue is beyond the scope of this paper, in this section we describe the differences in income for immigrants by the characteristics that are the most important source of variation in survival. The analysis is limited by the fact that no direct income measure is available for those leaving in the first year when many of the departures take place. There may be important selection effects that are missed because of this data limitation.

Table 2 reports mean log income, in constant 1992 dollars by visa class, source country, language, and education.[16] The variation in survival by visa class reported in table 1 was very large. There was much less survival among the business-class and skilled-class entrants compared with family, assisted relative, and refugee. Table 4 shows that the income measure for the skilled class is approximately 30% higher than for the assisted-relative class, 40% higher than for the family class, and 50% higher than for refugees and others. The low survival of the skilled-class entrants therefore results in a lower average income for the

15 Whether there is a downward period of recession or upward period of recovery at entry is more clearly related to the probability of leaving than to the unemployment level. In fact, in a simple regression the actual entry unemployment rate has a negative sign.

16 Income is the sum of wages and salaries and self-employment income. Table 2 reports the income measure taken from the first income report after the landing year, given positive earnings. Other measures using averages over time show the same pattern.

TABLE 2
Income by visa class, source country, language, and education: males, aged 25 to 45 at arrival

	Log Income	Count
No post-secondary	9.497	5023
Non-university post-secondary	9.676	2723
BA degree or above	9.918	3466
English	9.766	6605
French	9.422	601
English & French	9.722	711
Neither English nor French	9.514	3290
Family class	9.613	2958
Business class	9.102	484
Skilled class	10.018	3030
Assisted relative class	9.722	1201
Refugee class	9.470	2516
Other admission category	9.507	1017
North America	10.107	434
Europe	9.987	3247
Asia, excluding Hong Kong	9.557	3583
Hong Kong	9.178	692
Middle East	9.300	961
Africa	9.490	879
Caribbean and Guyana	9.708	541
South & Central America	9.480	142
Oceania and Australia	10.061	733

NOTES
Log income is the mean of log income, measured in 1992 dollars, in the first earnings observation for those with positive income. Income is the sum of wages and salaries and self-employment income.
SOURCE: Calculations by authors based on the Landing Records (LIDS) and the Longitudinal Immigration Database (IMDB) data

remaining immigrants. Survival is also low for the business-class entrants. The income measure shows a particularly low value for this group. It is unclear how this should be interpreted. On the one hand, this could indicate difficulties of setting up highly successful businesses in Canada; alternatively, it could reflect the standard difficulties of adequately capturing income for self-employed. In addition, relative to the skilled-worker and other visa classes, this is a small group.

Source country shows considerable survival variation in table 1. There were particularly large outflows for North, South, and Central America and for Hong Kong. North, South, and Central America are, in fact, not very important source countries for Canada. It is clear from table 2 that the highest income is reported for North America. However, Oceania and Australia is a more important source and has almost the same income, and Europe is a much more important source and also has a relatively high income. The differential out-migration from American source countries thus has little effect on the average income of the surviving immigrants. The Hong Kong income level is also low, reflecting the correlation with business class.

The selectivity in survival has important implications for assimilation studies. Analysis of immigrant earnings using multiple cross-sectional data may provide biased estimates of immigrant outcomes if there is selectivity in the return migration of immigrants. For example, if immigrants with the least favourable outcomes have a greater propensity to leave the host country, then estimates of earnings assimilation, under the assumption of no selection among the leavers, will be positively biased. Conversely, a large portion of the early leavers from Canada may have been the more mobile immigrants who could have done very well in Canada but had better outside opportunities. The evidence presented here for the Canadian data suggests that, since the return migration is both large in magnitude and selective on observed characteristics, potential biases could be quite large.[17]

6. Immigrant retention: comparison with Census methods

The IMDB provides information on tax filing behaviour rather than residence behaviour directly. In this section, the results are compared with an analysis using census data that measures residence directly. In the previous literature on return migration, a major focus of interest is estimating the retention rate of immigrants to answer the question: given that return migration takes place, what fraction stay in the host country? Census data are used in this calculation. In this section, estimates of the retention of immigrants in Canada are presented, based on a conceptual framework analogous to that of Borjas and Bratsberg (1996) for the United States. They work with a generic out-migration rate defined as

$$q(t, t') = \frac{I(t) - R(t')}{R(t')}, \tag{7}$$

where $I(t)$ is the number of persons who immigrate in year t and $R(t')$ is the number of those immigrants who remain as of t'. The source of $I(t)$ in Borjas and Bratsberg (1996) is INS microdata that recorded every legal immigrant admitted into the U.S. between 1 July 1971 and 30 September 1986. The source of $R(t')$ is the 1980 Census, so that t' is 1 April 1980.[18] The analogous sources for Canada are LIDS for $I(t)$ and the relevant Canadian census for $R(t')$. The retention of

17 Using Swedish data, Edin, LaLonde, and Aslund (2000) find that failure to adjust for emigration leads to an overestimation of the rate of economic assimilation for Nordic and OECD immigrants by around 90%. Similarly, Hu (2000) finds that results from longitudinal data suggest Census data estimates overstate the rate of growth of immigrant earnings and understate the worsening of immigrant entry earnings. Jasso and Rosenzweig (1988) find that estimates obtained from Census data significantly understate the returns to the U.S. experience of permanent resident aliens with respect to occupational attainment, owing to selective emigration of higher-occupation immigrants. On the other hand, Lindstrom and Massey (1994), using U.S. data, find that selective emigration does not significantly affect cross-sectional models.

18 Some adjustment was necessary to make the Census comparable with the INS. In particular, an estimate of illegal immigrants in the Census was necessary, since the INS covers only legal immigrants. In addition, the immigrant cohorts were 'aged' to 1 April 1980 using age/sex-specific mortality rates to estimate survival. For a detailed description of the adjustments, see Borjas and Bratsberg (1996, 168–70).

TABLE 3
Census based retention rates at 5, 10, 15, and 20 years after landing: males

Year	Landings	Retention rates at various years after landing			
		5 years	10 years	15 years	20 years
		All males			
1981	63,470	80.9	77.5	71.6	68.8
1986	49,380	90.22	86.6	82.7	..
1991	116,720	72.6	64.0	..	..
1996	111,290	76.3	..	..	..
		Males aged 25 to 35 at landing			
1981	18,040	76.2	76.9	70.8	67.6
1986	15,580	88.6	86.3	78.5	..
1991	40,860	63.7	57.2	..	..
1996	32,920	65.8	..	..	..

NOTES
The number of landings are from the Landings Records and are for the calendar year. The retention rates are based on the census counts in the relevant census years of individuals recording their year of migration.
.. – not available for a specific reference period.
SOURCE: Landing Records (LIDS) and Census data

immigrants in Canada can then be measured by fraction of immigrants arriving at time t who are still retained at time t':

$$r(t', t) = \frac{t'}{I(t)}. \tag{8}$$

The time path of the retention percentages for males, $r(t', t)$, for the one-year landing cohorts that match up with the Canadian census periods are given in table 3.[19] The time pattern is very similar to that documented in the duration analysis using the IMDB. The census years 1981 and 1991 were recession years; the 1986 and 1996 years were not. Comparison of each pair of years shows the five-year survival rates to be substantially lower for the more recent cohorts. For all males, the 1981 cohort percentage retained after five years is 80.9, compared with 72.6 for the 1991 cohort – a decline of 10%. The 10-year survival rates fall even more: the retained percentage for the 1981 cohort after 10 years is 77.5, compared with 64.0 for the 1991 cohort – a decline of 17%. Similarly, across the 1986 and 1996 cohorts, there is a fall in the percentage retained after five years from 90.2 to 76.3 – a decline of 15%.

The population of male immigrants includes both workers and non-workers whose emigration rates are likely to be influenced by various factors in different ways. In the lower half of table 3, the survival percentages are presented for males with age at landing between 25 and 35 to capture a young working-age population.

19 This abstracts from problems of mortality and illegal immigrants dealt with in Borjas and Bratsberg (1996), focusing on trends rather than absolute rates. Implicitly, it is assumed that mortality and illegal immigration rates are stable over the period.

TABLE 4
Tax-filing retention rates at 5, 10, 15, and 20 years after landing: males aged 25 to 35 at landing

Year	Landings	Retention rates at various years after landing			
		5 years	10 years	15 years	20 years
1981	18,040	78.6	70.0	67.1	65.5
1986	15,580	84.5	81.5	78.5	..
1991	40,860	77.7	73.4	..	..
1996	32,920	77.0	..	..	..

NOTES
The number of landings is from the Landings Records. The retention rates are based on the assumption that non-tax-filing in four consecutive years in the IMDB constitutes emigration.
.. – not available for a specific reference period.
SOURCE: Calculations by authors based on Landing Records (LIDS) and the Longitudinal Immigration Data Base (IMDB) data

The patterns observed in the total population of males show yet higher mobility in this population. The five-year survival percentages for the 1981 and 1991 cohorts are 76.2 and 63.7, respectively – a decline of 16%; after 10 years the decline is 26%. For the 1986 and 1996 cohorts the decline in the five-year survival percentage is 25%. These are large declines in landing cohorts separated by only a decade. There is also the same business-cycle pattern, as observed in the previous section, with low retention rates for cohorts arriving in the trough census years of 1981 and 1991 on either side of the high retention rates in the peak census year of 1986.

Table 4 reports a comparison of the tax-filing-based residence definition and the Census-based method. When the retention rates are calculated if four years of consecutive non-filing were considered equivalent to emigration, table 4 shows that rates are close to the estimates in table 3. The 20-year retention rate for males with landing age 25 to 35 in table 3 is 67.6%, while the same rate in table 4 is 65.5%. The 15-year retention rates are identical for the 1986 cohort, while the rate for 1981 cohort is about 4 percentage points lower in table 4. For 1991 and 1996 the estimates from both methods show a decline from 1986, but less so in the tax-filing method. The Census data are affected by a change in the question in 1991. In 1991 the questions ask specifically about the year landed immigrant status was obtained, whereas prior to that the question was more vague, asking in what year the person immigrated to Canada, which may not always correspond to the year of landing. This may be part of the difference. However, tax-filing behaviour could also change over time. During the late 1980s and early 1990s, in order to get a number of government transfers such as child tax benefits, individuals were required to file tax returns even if they had no taxes payable. There was no such requirement before this period and this might have affected tax-filing behaviour. The census estimates of retention rates abstract from issues of disappearance and reappearance. They measure what fraction of a cohort is

in the country at a particular point, whether they stayed there all the time, or left and re-entered. The tax-filing estimates presented in table 4 neglect re-entry. The data show evidence of re-entry, especially for recent cohorts, so that the tax-filing-based retention estimates in table 3 may be underestimates. The magnitude of the re-entry, however, is modest, so that the degree of under-estimation will be small.

Overall, the results in table 4 suggest that tax-filing behaviour is closely related to residence behaviour as reflected in estimates of retention rates. Of additional interest is how close the relationship is in terms of the shape of the hazard rate.

The hazard rate estimates presented in section 3 (figure 2) show a very sharp drop in the first year. This results in the survival rate falling to about 78% by the end of the first year. After this, the drop is much more gradual. When we use the 1991 and 1996 Censuses, the same basic pattern appears with a very similar magnitude for the initial drop. For those landing in 1990, the 1991 Census records a survival rate of 79.4%; for those landing in 1995, the 1996 Census shows a survival rate of 82.5%.[20] Subsequent survival rate drops are much smaller. These magnitudes imply very similar hazards to those obtained from the tax-filing-based estimates. Thus, even for the first year after landing, where many of the disappearances take place, the tax-filing behaviour appears to closely mirror the residence behaviour, as measured in the Census.[21]

However, it remains necessary to interpret the results with some caution, since the landing date does not necessarily coincide with initial residence in the country. The data set provides evidence on landings, but not on residence prior to landing. Thus, the landings in a particular year could include individuals who came to Canada some years earlier. The hazards reported here are for 'departure since landing.' These must be higher than the first-year hazard of 'departure since entry' for a cohort of individuals who either immediately or subsequently 'land,' since some individuals who leave, say, two years after entry are classified as leaving one year after landing. Conversely, the hazard for departure since landing for subsequent years will on average be lower than the hazard for departure since entry (with either coincident or subsequent landing).[22]

20 An alternative method to estimate emigration rates using the tax filing behaviour would be to use a sample of tax filers only, that is, those who ever appear in the IMDB. About 10% to 15% of immigrants never appear in the tax files. If we restrict our sample to those who ever appear in the IMDB, then the estimated emigration rates would be half of those implied by the Census method. This bias in estimates highlights the importance of starting with a full count of immigrants rather than using tax filers only for the analysis.

21 In addition, the disaggregated patterns obtained using Census-based estimates support the patterns found using the tax-filing data in the IMBD. See Aydemir and Robinson (2006) for more detailed results on the Census comparison. To the extent that the patterns in the Census and tax-filing record-approaches are very similar where they overlap, this lends credibility to the use of the tax-record approach in analyzing the other characteristics available in the IMDB, as well as to the longitudinal analysis at the individual level, which cannot be carried out by the Census approach.

22 The incidence of living in Canada prior to 'landing' is investigated using the Longitudinal Survey of Immigrants to Canada (LSIC) that captures a cohort of immigrants that landed

7. Conclusions

International migration is not a permanent move and many immigrants either return to the source country, perhaps many times, or move on to another country. There is increasing evidence that skilled workers in particular are becoming more internationally mobile in the new global market. The IMDB presents an opportunity to study this phenomenon as it affects a large group of immigrants from a wide variety of source countries over more than a 20-year period. Since the IMDB contains information on all the characteristics used to implement the points system used in Canada to determine eligibility for admittance, this evidence is particularly relevant for providing background evidence for informed discussion of amendments to immigration policy based on changes to the points system.

It is clear that a substantial part of migration to Canada is temporary. The estimated out-migration rate 20 years after arrival is around 35% among young, working-age, male immigrants. About 6 out of 10 of those who leave do so within the first year of arrival, which suggests that many immigrants make their decisions within a relatively short period of time after arrival. When we control for other characteristics, the out-migration rates are higher among immigrants admitted under the skilled worker or business class visa. About 4 in 10 left within 10 years after arrival. For the assisted-relative class and the refugees the corresponding rates were around 3 in 10 and 2 in 10. Finally, the out-migration rates are higher for those who arrive during recessionary periods. Immigrants who arrived in 1990, for example, were about 50% more likely to leave than those who arrived in 1986, when we control for other characteristics.

In view of the potentially temporary nature of all migration, calculation of the contribution that can be expected from a new immigrant to the population, the labour force, or the human capital stock of a country has to take into account the probability of the immigrant being in the country at each point over the immigrant's remaining lifespan. The duration analysis in section 3 presents a start in providing the information needed for this calculation.

Appendix

The primary data used in the analysis are landing dates and subsequent tax-filing information. The focus is on the determinants of the length of the first spell

during 2000–1 period. Based on the interview that took place six months after landing, less than 6% of working-age male family-class immigrants stated that they lived in Canada prior to landing. Corresponding figures were 12% and 12.3% for skilled-worker and business-class immigrants. As discussed in the appendix, achieving landed status could be an incentive for some immigrants already in Canada who have received good alternative options to remain until landing and then to leave shortly afterward. These figures of relatively large numbers of the business- and skilled-class groups having prior residence in Canada could account for the relatively large outflow from these groups soon after landing.

in Canada, starting from the landing date. While the landing date is measured precisely from the landing records, the end of the first spell has to be inferred from the observed tax-filing behaviour.

A.1. Definition of the end of the first spell

The definition is motivated by the notion of opportunities to file and the benefits of filing.

- If tax-filing behaviour shows no sequence of four consecutive years of non-filing *after* the landing year, the first spell is assumed to be continuing.
- If tax-filing behaviour shows a sequence of four consecutive years of non-filing *after* the landing year, the end date for the first spell is assumed to fall between January and December of the first year of the first sequence of four consecutive years of non-filing.

This definition works best for a sample of immigrants all landing in the same month, since opportunities and benefits are likely to depend on the month of arrival, in terms of both proximity to the normal filing date in April and the likelihood of having earnings to declare and/or benefits to claim.

A.2. The first-year hazard: landing date and the start of residence

The start date of the first spell is simply defined as the exact landing date. About two-thirds of immigrants file a tax return in the same year as landing, which is consistent with some period of residence prior to landing.[23] The status change at landing is significant. For example, there may be two processes determining the leaving patterns after the 'pre-landing' entry to Canada: first, there is a process that determines the relative attractiveness of actual residence in Canada and the alternatives that are always present; second, there is a value to acquiring landed immigrant status in Canada, independent of residing in Canada.[24] In the pre-landing period there is a hazard for leaving based on the first process, which is partially offset by the second process – the investment in landed immigrant status. In the post-landing period there is only the first process. At the time of landing this results in a stock of individuals who will wish to leave as soon as the landed status is received. The existence of this stock means that the estimated hazard rates, especially in the years closest to landing, should be interpreted with caution. They are likely to be mixtures of the stock depletion and the hazard due to the first process.

Since there is monthly variation in the landing date, in principle this could be used to estimate a monthly hazard rate for the first year, given an assumption on the interval for the leaving date. The yearly interval of possible tax filing

23 A return filed in a given year refers to the previous tax year.

24 The results in section 3 for Hong Kong suggest that this is important.

suggests that the assumption of an interval corresponding to a tax year may be a reasonable approximation for estimating yearly hazards. This may break down if it is used to identify the finer gradations of a monthly hazard. The results from estimating empirical monthly hazard rates by using the monthly variation in arrival times together with a common 12-month interval for the departure date show a uniformly high hazard for the first 12 months, followed by a discrete jump at the 13th month. One interpretation is that the second process is one in which all leave in the first month after landing or stay, which, by construction, adds a relatively large and equal amount to all the first 12-month hazards, which dominates any underlying pattern from the first process, which continues for all other years. That is, the beginning of the common interval is the first time anyone can leave, whatever month they arrived in. The detailed assumptions that are made at the monthly level regarding possible filing dates and propensities will determine how the stock of leavers from the first process observed at the end of a particular tax year is spread over the first 12 months. Assumptions that spread them evenly result in a jump at month 12. Other assumptions could result in an alternative distribution, but how to choose among competing possible assumptions is unclear. In essence, it is asking too much of the data in this tax year form to provide good monthly detail.

References

Angelucci, Manuela (2005) 'US border enforcement and the inflow-outflow of Mexican illegal migration,' Institute for the Study of Labour (IZA) Discussion Paper Series No. 1642

Aydemir, Abdurrahman, and Chris Robinson (2006) 'Global labour markets, return and onward migration,' CIBC Centre for Human Capital and Productivity Working Paper 2006-1, University of Western Ontario

Beach, Charles M., Alan G. Green, and Jeffrey G. Reitz, eds (2003) *Canadian Immigration Policy for the 21st Century* (Montreal and Kingston: John Deutsch Institute for the Study of Economic Policy and McGill-Queen's University Press)

Borjas, George J., and Bernt Bratsberg (1996) 'Who leaves? The outmigration of the foreign born,' *Review of Economics and Statistics* 78, 165–76

Colussi, Aldo (2006) 'Migrants' networks: an estimable model of illegal Mexican immigration,' Working Paper

Constant, Amelie, and Douglas S. Massey (2003) 'Self-selection, earnings, and outmigration: a longitudinal study of immigrants to germany,' *Journal of Population Economics* 16, 631–53

Devoretz, Don J., and Diane Coulombe (2005) 'Labour market mobility between Canada and the United States: quo vadis 2003?' in *Social and Labour Market Aspects of North American Linkages,* ed. T. Lemieux and R. Harris (Calgary: University of Calgary Press)

Devoretz, Don J., and John Ma (2002) 'Triangular human capital flows between sending, entrepot and the rest of the world regions,' *Canadian Population Studies* 29, 53–69

Edin, Per-Anders, Robert J. LaLonde, and Olof Åslund (2001) 'Emigration of immigrants and measures of immigrant assimilation: evidence from Sweden,' *Swedish Economic Policy Review* 7, 163–204

Hu, Wei-Yin (2000) 'Immigrant earning assimilation: estimates from longitudinal data,' *AEA Papers and Proceedings, American Economic Review* 90, 368–72
Jasso, Guillermina, and Mark R. Rosenzweig (1982) 'Estimating the emigration rates of legal immigrants using administrative and survey data: the 1971 cohort of immigrants to the United States,' *Demography* 19, 279–90
— (1998) 'How well do US immigrants do? Vintage effects, emigration selectivity, and occupational Mobility,' *Research in Population Economics* 6, 229–53
Lam, Kit-Chun (1994) 'Outmigration of foreign-born members in Canada,' *Canadian Journal of Economics* 27, 352–70
Lindstrom, David P., and Douglas S. Massey (1994) 'Selective emigration, cohort quality, and models of immigrant assimilation,' *Social Science Research* 23, 315–49
Massey, Douglas S., Jorge Durand, and Nolan J. Malone (2002) *Beyond Smoke and Mirrors: Mexican Immigration in an Era of Free Trade* (New York: Russell Sage Foundation)
Meyer, B.D. (1990) 'Unemployment insurance and unemployment spells,' *Econometrica* 58, 757–82
Munshi, Kaivan (2003) 'Networks in the modern economy: Mexican migrants in the U.S. labor market,' *Quarterly Journal of Economics* 118, 549–99
Piore, Michael J. (1979) *Birds of Passage: Migrant Labor and Industrial Societies* (New York: Cambridge University Press)
Prentice, Ross L., and Lynn A. Gloeckler (1978) 'Regression analysis of grouped survival data with application to breast cancer data,' *Biometrics* 34, 57–67
Warren, Robert, and Jennifer Marks Peck (1980) 'Foreign-born emigration from the United States: 1960 to 1970,' *Demography* 17, 71–84