

CHAPTER 10

Selective Out-Migration and the Estimation of Immigrants' Earnings Profiles

Christian Dustmann, Joseph-Simon Görlach

Centre for Research and Analysis of Migration (CReAM) and Department of Economics, University College London, Drayton House, 30 Gordon Street, London WC1H 0AX, UK

Contents

1. Introduction	489
2. Evidence on Temporary Migration and Selective Out-Migration	491
2.1 Selective out-migration by country of origin	492
2.2 Selection on education	496
2.3 Selection on earnings	498
2.4 Other characteristics and out-migration	500
3. Estimating Immigrants' Career Profiles	501
3.1 Key research questions and immigrants' career profiles	502
3.2 Estimation and identification of immigrant career profiles	504
3.2.1 Stock sampled data	506
3.2.2 Complete longitudinal data	510
3.3 Numerical example	513
3.4 Interpretation: A simple model of return migration	518
4. Existent Studies on the Estimation of Earnings Equations when Out-Migration is Nonrandom	522
4.1 Studies using stock sampled longitudinal data	523
4.2 Studies using longitudinal data	524
5. Conclusions	528
Acknowledgments	529
References	529

1. INTRODUCTION

A major analytical focus in migration research, itself a widely studied area in modern applied economics, is the career paths of immigrants after their arrival in the destination country (see the seminal work by [Chiswick, 1978](#); and papers by [Carliner, 1980](#); [Long, 1980](#); [Borjas, 1985](#); [LaLonde and Topel, 1992](#); [Dustmann, 1993](#); [Hu, 2000](#); [Cortes, 2004](#); [Eckstein and Weiss, 2004](#); [Lubotsky, 2007](#); and [Abramitzky et al., 2013](#)). This focus is hardly surprising given its importance for understanding how immigrants contribute to the labor market, fiscal system, and the economy at large. Yet empirically

identifying immigrants' career profiles and their evolution over time presents serious challenges. For example, as pointed out by Borjas (1985), not only may failing to account for changes in the quality of immigrant inflows lead to biased estimates of earnings profiles but the temporariness of many migrations poses its own serious problems for assessing how immigrants' economic outcomes evolve over time.

One primary reason that migration temporariness engenders serious identification issues in estimates of immigrant earnings profiles is that the non-randomness of out-migration may lead to endogenous selection of the resident immigrant population, so that if earnings profile estimations ignore the possibly selective character of out-migration, this omission may lead to biased estimates of the immigrants' career progressions in the destination country. In fact, migration temporariness may itself transform the dynamic optimization problem of the individual immigrant, leading to human capital investment and job search behavior that is set in conjunction with out-migration decisions and determined by—usually unobserved—expectations about future economic conditions in the immigrants' home countries.

In this chapter, we focus on the first issue, the identification of immigrants' career profiles in the destination countries in the presence of out-migration that is selective.¹

We can illustrate the problem as follows. Suppose that the log earnings of a particular entry cohort of immigrants c are given by $w_{it}^c = \mu_t^c + \varepsilon_{it}^c$, where i and t are index individuals and time respectively, μ_t^c is the mean earnings of entry cohort c at time t , and ε is the deviation of individual earnings from the cohort mean, which collects individual characteristics and follows a certain distribution. The identification problem in estimating the career profile of a particular immigrant entry cohort is then as shown in Figure 10.1, in which the distribution of earnings in period t_1 of all immigrants who arrived in period c has the mean μ_1^c . Assuming a random sample of immigrants interviewed in t_1 and again in period t_2 , a simple (linear) earnings regression can be used to identify the wage progression of immigrants in the host country as $(\mu_2^c - \mu_1^c)/(t_2 - t_1)$, where μ_t^c is the mean of the log earnings of the immigrants still residing in the country at time t .

This outcome, however, although an unbiased estimate of the growth in the mean earnings of the migrant populations who arrived in period c and residing in the destination country in each of the periods t_1 and t_2 , it is not an estimate of the wage growth of individuals from the original arrival cohort, $(\mu_2^c - \mu_1^c)/(t_2 - t_1)$, had out-migration not occurred. That is, an OLS estimator using the two waves of cross-sectional data only produces an unbiased estimate of $(\mu_2^c - \mu_1^c)/(t_2 - t_1)$ if the entire cohort of immigrants that entered in period c still resides in the country in periods t_1 and t_2 or out-migration is random. In the case of non-randomness—for example, only the least successful leaving the country—the distribution of immigrants residing in the host country would be truncated

¹ We furthermore focus on selection on earnings levels rather than selection on earnings growth, as does the vast majority of the literature.

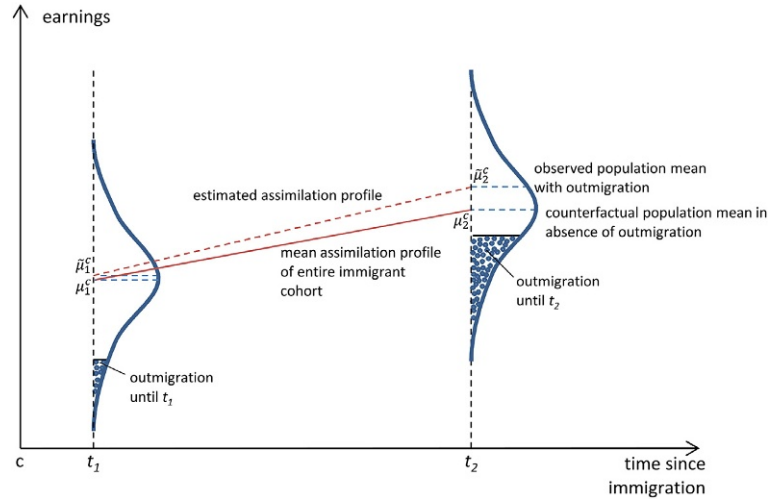


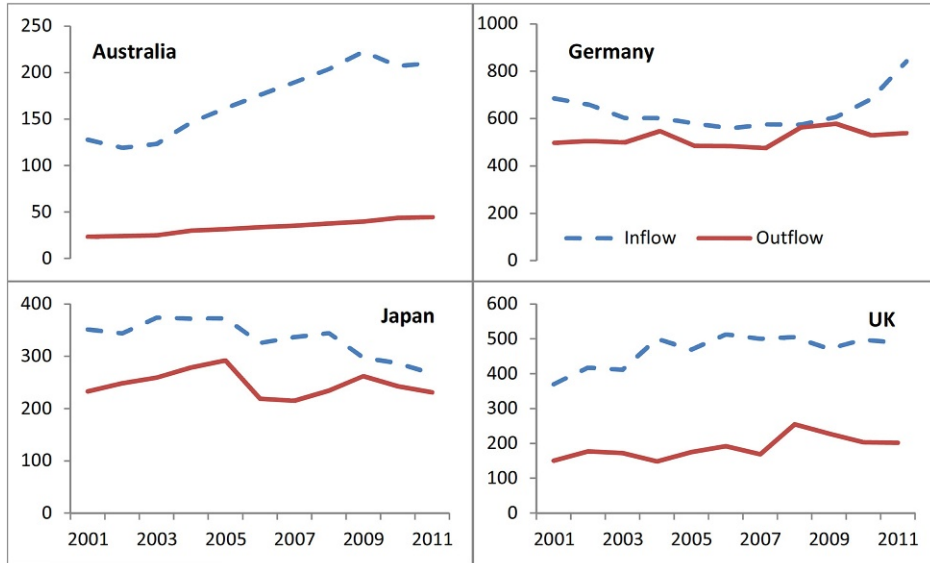
Figure 10.1 *Biased estimation of earnings profiles when out-migration is selective.*

from below, as in the figure. Hence, a simple OLS estimator that ignores this selective out-migration would indicate a steeper wage progression for this cohort.

This chapter, after first providing evidence on the scale of temporary migrations and their possibly selective character (Section 2), explains in more detail the methodological problems involved in estimating immigrant outcome profiles in the presence of selective out-migration (Section 3). We also outline the various ways in which to address these issues. We finally provide an example of how a simple model of endogenous return migration may lead to selection, and how this impacts on empirical estimates. Section 4 then provides an overview of the literature that estimates immigrant earnings profiles while accounting for the temporariness of many migrations, and discusses these papers within the framework we set out in Section 3. Finally, Section 5 summarizes the chapter contents and presents our final thoughts.

2. EVIDENCE ON TEMPORARY MIGRATION AND SELECTIVE OUT-MIGRATION

As shown in Figure 10.2 for a number of major OECD countries, immigration over the last decade has been accompanied by very sizeable out-migration (see OECD (2013) for the country-specific variable definitions). Not only are the profiles for other immigrant-receiving countries similar (OECD, 2013), but increasing evidence is emerging that permanent migrations are—and possibly always have been—the exception rather than the rule. Indeed, the temporariness of migration was stressed even in the early migration literature; for example, Piore (1979) estimated that over 30% of immigrants admitted



Source: OECD International Migration Outlook 2013

Figure 10.2 Immigrant in- and outflows in thousands for selected OECD countries. Source: *OECD International Migration Outlook (2013)*.

into the US in the early 1900s subsequently emigrated back to their countries of origin, a figure that actually may have been over twice as large (Bandiera et al., 2013). Such temporariness continues today: During the 2000–2010 period, almost 2.1 million foreign-born residents out-migrated from the US (Bhaskar et al., 2013), a pattern also characteristic of many other countries. A recent report by the OECD (2008), for instance, estimates out-migration rates after five years of residence of 60.4% for immigrants entering Ireland in 1993–98, 50.4% for immigrants entering Belgium in 1993–99, 39.9% for immigrants entering the UK in 1992–98, 39.6% for immigrants entering Norway in 1996–99, and 28.2% for immigrants entering the Netherlands in 1994–98.

Assessing out-migrations, however, is subject to a notable measurement problem: Although many countries carefully register the arrival of new immigrants, most keep no records of immigrants who leave, which greatly complicates the estimation of immigrants' career profiles. Nevertheless, the emergence of better data sources over recent decades has improved the documentation of foreign-born emigration.

2.1 Selective out-migration by country of origin

The most important question in estimating immigrant career profiles in destination countries is not the pure scale of out-migration but whether it is in any way selective and who the out-migrants are. A study by Dustmann and Weiss (2007), for example, used British

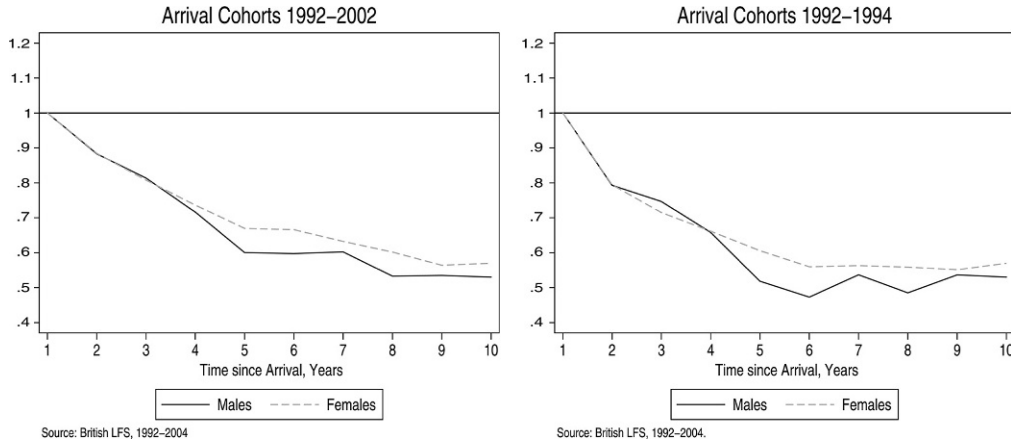


Figure 10.3 Immigrant survival rates by gender, from [Dustmann and Weiss \(2007\)](#). Source: British LFS, 1992–2004.

Labour Force Survey (BLFS) data on the year of first entry for different arrival cohorts to compute the fraction of each such cohort c that is still in the sample in year $c+j$ and thus estimated the extent of out-migration from the UK. Because the BLFS is not reliable for immigrants who are in the UK for less than a year, however, their base population is all immigrants who have been in the UK for at least one year (i.e., their analysis ignores the many migrations terminated within a year), so their figures underestimate the degree of out-migration.

[Figure 10.3](#) (see also [Dustman and Weiss, 2007](#), Figure 2) shows the survival rate of immigrants in Britain who stayed at least for one year from the first year after arrival until up to 10 years after arrival, with a distinction made between males (solid line) and females (dashed line). The right-hand panel also breaks out the cohorts that arrived between 1992 and 1994. In particular, the graph shows a large reduction in survival rates over the first five years, which, if interpreted as emigration, indicates that among those who stayed for at least one year, only about 60% of the male and 68% of the female foreign-born immigrants remained in Britain five years later.

[Figure 10.4](#) (see also [Dustmann and Weiss, 2007](#), Figure 3) pools male and female immigrants but distinguishes between origin (left panel) and ethnicity (right panel).² Overall, it shows a large variation in the emigration propensities across immigrants from different origin countries: The out-migration for immigrants from Europe, the Americas, Australasia, and the Middle East is highest (over 45% of those who are in the UK for at least one year left within five years of arrival), while out-migration for individuals from the Indian subcontinent and Africa is far less pronounced. In fact, there seems to be little

² Fractions larger than 1 in the left panel are due to sampling error.

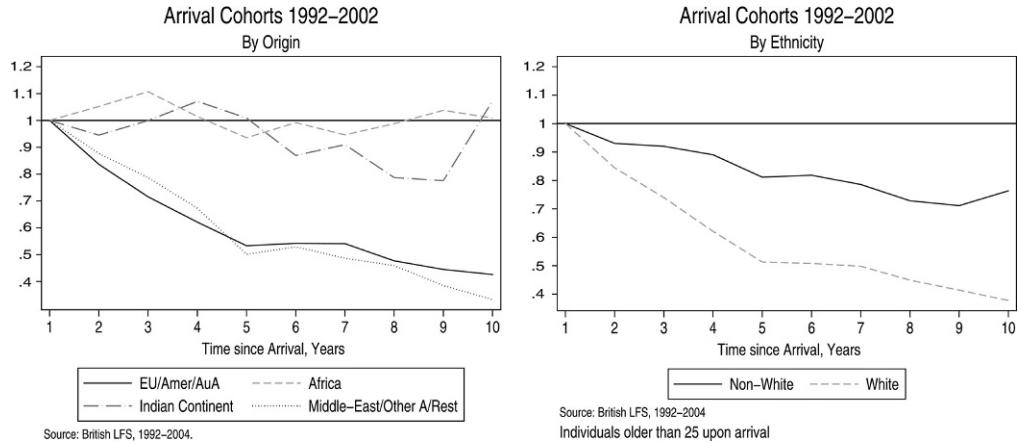


Figure 10.4 Immigrant survival rates by origin and ethnicity, from *Dustmann and Weiss (2007)*. Source: British LFS, 1992–2004.

indication of any emigration for immigrants from Africa and the Indian subcontinent. Rather, instead of being equally distributed across origin countries, out-migrants seem to come predominantly from English-speaking countries that are economically similar to the UK. *Rendall and Ball (2004)* reported a very similar picture: Whereas about 65% of immigrants from Canada and the US emigrate within five years, only about 15% of immigrants from the Indian subcontinent do so.

Borjas and Bratsberg (1996), using US data, found similar large variation in out-migration rates across origin countries, with the lowest out-migration rates reported for immigrants from Asia (cf. *Dustmann and Weiss, 2007*). Specifically, they estimated that only 3.5% of Asian immigrants who arrived in the US after 1975 had left the country by 1980, as compared to 18.4% of European immigrants, 24.8% of South American immigrants, and 34.5% of North American immigrants. Likewise, *Jasso and Rosenzweig (1990)*, using alien registration data for the 20 years between 1961 and 1980, found that Europeans have the highest propensity to leave the US and immigrants from Asia the lowest, with western hemisphere immigrants taking an intermediate place. In an earlier paper analyzing out-migration rates for the 1971 entry cohort, these same authors reported that immigrants from China, Korea, Cuba, the Philippines, and India had the lowest emigration rates, ranging from 14.5% to 41.6%, with emigration rates for Koreans and Chinese not exceeding 22% (*Jasso and Rosenzweig, 1982*). Canadian emigration, in contrast, was between 51% and 55%, and emigration rates for legal immigrants from Central America, the Caribbean (excluding Cuba), and South America were at least as high as 50% and possibly as high as 70%.

Patterns reported for Canada are not dissimilar. *Beaujot and Rappak (1989)* reported that out-migration for those arriving in the 1951–70 period was highest for individuals born in the US (50–62%) and lowest for those from Asia (only 1–17%). A similar profile

emerges for immigrants to Australia: Based on data from the 1973 Social Sciences Survey of Australian male residents aged 30–64 and the 1/100 census tapes from 1981, [Beggs and Chapman \(1991\)](#) reported that only 3–6% of immigrants from non-English-speaking countries are likely to leave as compared with 20–30% of immigrants from English-speaking countries (see also [Lukomskyj and Richards, 1986](#)).

Research for Scandinavian countries indicates that emigration rates for immigrants from industrialized—and in particular, Nordic countries—are far higher than those for immigrants from other regions. For Norway, [Bratsberg et al. \(2007\)](#), using data from the Norwegian population register, estimated that as many as 84% of immigrants from the US leave compared with only 9% from Vietnam. They interpreted this finding to mean stark differences in out-migration behavior based on the home country's economic development and distance from Norway. In addition, their administrative data included information on where foreign-born emigrants travel subsequently. They reported that of those who left Norway, at least 30% of the immigrants from Somalia, 40% from Iran, and two-thirds from Vietnam migrated to a third country rather than returning, while the great majority of out-migrants from neighboring Nordic countries returned home. Like [Bratsberg et al. \(2007\)](#), based on 1991–2000 emigrant data from *Statistics Sweden*, [Nekby \(2006\)](#) found that 28% of working-age emigrants are onward migrants emigrating to a third country. Moreover, whereas emigrants from Asia are as likely to be onward as return migrants, emigrants from Africa are more likely to move to third-country destinations, which suggests that many of the out-migrations observed are not return migrations but migrations that continue to other destination countries. Overall, she found that out-migration probabilities are highest for immigrants from North America and from Western European countries of origin.

The large out-migration propensity of Nordic migrants in comparison to other groups was confirmed by [Jensen and Pedersen \(2007\)](#) for Denmark, who found that although 80% of Turkish immigrants remained in Denmark for 10 years or more, only 20% of the Nordic immigrants did so. [Edin et al. \(2000\)](#) showed similar patterns for Sweden: About 44% of Nordic immigrants had left that country within five years of their arrival, a number that is significantly lower for immigrants from non-OECD countries. [Klinthäll \(2003\)](#) reported similarly large differences among the 1980 and 1990 arrival cohorts of non-Nordic European immigrants to Sweden, whose emigration rates are about twice as high as those of African and Asian immigrants.

The general picture that emerges from these studies is that migrants from developed countries are more likely to leave the host country than migrants from less developed countries, in particular those in Africa and Asia. This pattern is illustrated in [Figure 10.5](#), which combines the estimates—drawn from a large number of empirical studies—on the fractions of postwar immigrants that out-migrated after a certain period. More specifically, it plots the estimated fraction of immigrants who left the host country within a given time since

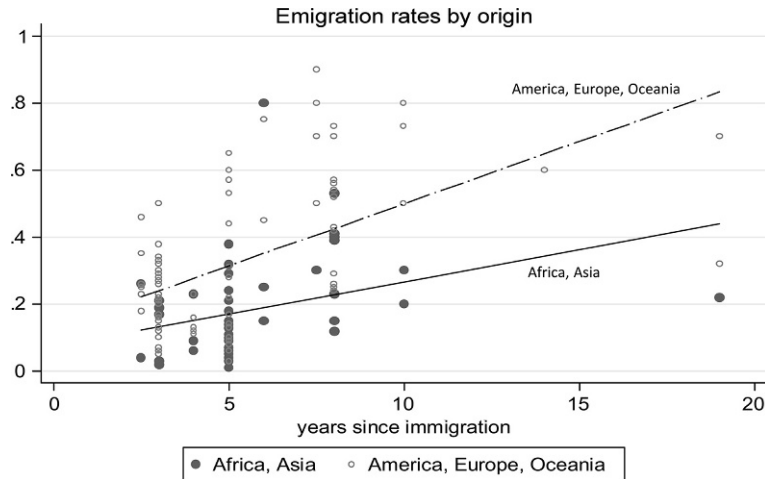


Figure 10.5 *Foreign-born emigration by origin region.*

arrival against the number of years since immigration.³ This pattern suggests higher emigration rates for migrants from the Americas, Europe, and the Pacific region than for migrants from Africa and Asia. Using this collection of estimates as observation points in a regression of the fraction of out-migrated immigrants on the years since immigration (ysm), we find that the out-migration rate of immigrants from the Americas, Europe, and the Pacific region increases on average by almost twice as many percentage points per year as that of immigrants from Africa and Asia (see Table 10.1). The estimated coefficients and the differences across immigrant groups are illustrated by the fitted lines in Figure 10.5.

2.2 Selection on education

Although a number of papers have examined the relation between educational attainment and out-migration, there is no clear pattern across the literature.⁴ For returnees

³ For Figure 10.5, we exclude annual emigration rate estimates that do not refer to a certain number of years spent in the host country, such as Van Hook et al.'s (2006) estimates of annual out-migration. If estimates refer to the fraction of migrants who entered in a certain time interval and left by the end of that interval (as in Bratsberg et al., 2007), the average year of immigration is approximated by the interval midpoint, a choice likely to overestimate emigration rates given that remigration propensities are generally higher during the early post-immigration years. The estimates are taken from Ahmed and Robinson (1994), Alders and Nicolaas (2003), Aydemir and Robinson (2008), Beaujot and Rappak (1989), Beenstock (1996), Bijwaard et al. (2011), Böhning (1984), Borjas and Bratsberg (1996), Bratsberg et al. (2007), Edin et al. (2000), Jasso and Rosenzweig (1982), Jensen and Pedersen (2007), Kirwan and Harrigan (1986), Klinthäll (2003, 2006), Lukomskyj and Richards (1986), Michalowski (1991), Rendall and Ball (2004), Reyes (1997, 2004), and Shorland (2006). The exact numbers used are available upon request.

⁴ See Dustmann and Glitz (2011) for a survey on the role played by skill accumulation and education not only in the selection of remigrants from a destination country's population of immigrants, but also on the selection of emigrants from their countries of origin.

Table 10.1 OLS coefficients of time since immigration for foreign-born emigration by origin region
Fraction that has emigrated by region of origin

	<i>Africa/Asia</i>	<i>Americas/Europe/Oceania</i>
<i>ysm</i>	0.019	0.037
<i>cons</i>	0.075	0.129
<i>N</i>	52	97

from the US among the Puerto Rican immigrant population, [Ramos \(1992\)](#) reported a positive selection on education, and [Zakharenko \(2008\)](#), working with CPS data, estimated the probability of emigration for immigrants to the US from any destination to be lower for highly educated immigrants. He also showed, however, that this result is largely driven by the strong association between higher education and emigration probabilities among longer-term migrants, a linkage that is statistically insignificant for short-term migrants. [Lam \(1994\)](#), on the other hand, using 1971 and 1981 Canadian census data, reported that younger and less educated immigrants to Canada are more likely to stay.

The literature reports similarly mixed results across European countries on the relation between educational attainment and the propensity to out-migrate. For example, [Jensen and Pedersen \(2007\)](#) found a positive relation between out-migration probabilities and educational attainment among immigrants in Denmark, while [Dustmann \(1996\)](#) found that intended migration durations are longer among less educated immigrants in Germany. Nevertheless, he also reported that the probability that immigration is intended to be permanent increases with years of schooling. Also for Germany, [Constant and Zimmermann \(2011\)](#), using information on multiple migration spells, showed that repeat migration is more likely among male and less educated individuals, whereas [Beenstock \(1996\)](#), using data for Israel, showed that the stays of more highly educated immigrants who arrived in the 1970s are more likely to be temporary. For Italy, [Coniglio et al. \(2009\)](#) confirmed that schooling increases the probability of out-migration even among undocumented immigrants. However, [Carrión-Flores \(2006\)](#), using data from the Mexican Migration Project 1982–1999, reported the opposite for Mexican immigrants in the US: They found a positive effect of high educational attainment on the likelihood of returning to Mexico. [Maré et al. \(2007\)](#), on the other hand, found for New Zealand that out-migration is highly likely for both unskilled and highly skilled immigrants.

These contradictory findings again raise the question of the direction of the selection of migrants from their societies of origin and out-migrants from the immigrant population in the respective destination countries, as well as how these two are related. We discuss theoretical models on this issue in [Section 3](#). One insight is provided by [Borjas and Bratsberg \(1996\)](#), who hypothesized that in a context of negative selection of immigrants from their origin societies, emigration by these migrants from the destination country

should be positively selected. This was confirmed for linked Finnish and Swedish data by [Rooth and Saarela \(2007\)](#), who found significant selection on educational attainment but no evidence of selection on earnings conditional on education. From a sending country's perspective, [Pinger \(2010\)](#) reported that among temporary emigrants from Moldova, a lower fraction has tertiary education than is the case among emigrants considered to have left permanently. Thus, overall, these studies suggest that the selection pattern of out-migration with respect to educational attainment is context-dependent.

2.3 Selection on earnings

A number of studies investigate the relation between out-migration and immigrant earnings,⁵ an association that is far from straightforward. [Dustmann \(2003\)](#), for instance, pointed out that changes in earnings may affect the optimal migration duration through either an income or substitution effect. Whereas the former increases the time a migrant may want to spend in the home country, the latter makes a return more costly. Empirical evidence is mixed, however, on which effect dominates: Most studies on the out-migration of US immigrants found that those earning high wages are less likely to leave, but findings differ for other immigration countries. Early work by [Massey \(1987\)](#) and [Borjas \(1989\)](#), for instance, identified a negative effect of wages on the probability that immigrants to the US will out-migrate, and [Abramitzky et al. \(2013\)](#) reached a similar conclusion for US immigrants even in the so-called age of mass migration. Using US census data from 1900, 1910, and 1920, together with the Integrated Public Use Microdata Series (IPUMS) for 1900, they constructed a panel of US natives and immigrants from a number of major sending countries who arrived between 1880 and 1900, and found that out-migrants were negatively selected on earnings. In [Abramitzky et al. \(2012\)](#), on the other hand, the same authors noted that for Norwegian immigrants who arrived around the same time there are no significant occupational differences between those who return to Norway and those who stay.

In more recent work, [Cohen and Haberfeld \(2001\)](#) assumed that in the absence of selective out-migration, period effects in earnings regressions for Israeli- and native-born workers in the US should be the same. Using a sample of such individuals from pooled 1980 and 1990 census data, they performed separate earnings regressions for the two groups. If out-migration were random, once years spent in the US and other individual human capital characteristics are controlled for, the coefficient on an indicator variable for being drawn from the 1990 sample should be the same in both regressions. In fact, the estimated coefficient is significantly lower for US-born workers, a finding that the authors interpret as evidence that Israelis who return from the US are negatively selected

⁵ See also [Reyes \(1997, 2001\)](#), [Constant and Massey \(2002, 2003\)](#), [Gundel and Peters \(2008\)](#), [Bijwaard \(2009\)](#), [Kirdar \(2009\)](#), [Yahirun \(2009\)](#), [Van Hook and Zhang \(2011\)](#), and [Bijwaard et al. \(2011\)](#) for their results on the effect of unemployment spells on the probability of out-migration.

among all Israeli immigrants. For the US, [Reagan and Olsen \(2000\)](#) also found that immigrants' potential wage as predicted by a number of observable characteristics of foreign-born workers included in the NLSY79 is negatively associated with the probability of emigrating. [Reyes \(1997\)](#) identified a negative relation between wages and the probability of return migration by Mexican immigrants in the US, but only during the first year of residence: This effect turns positive for immigrants who have remained in the US for longer. Interestingly, her data, taken from the Mexican Migration Project, suggest similarly sized wage effects for both male and female immigrants. The finding that foreign-born emigrants are negatively selected from the immigrant population is also supported by [Lubotsky \(2007\)](#), a study detailed in [Section 4](#).

Evidence does exist, however, of considerable differences in the relation between earnings and out-migration dependent on both origin and destination countries. [Longva \(2001\)](#), for instance, after dividing the immigrant population residing in Norway in 1980 and 1993 into those from OECD and those from non-OECD countries, found that OECD immigrants who left Norway between 1980 and 1992 or between 1993 and 1997 had higher earnings at the beginning of these periods than those who stayed. For non-OECD immigrants, however, he found the opposite. According to [Edin et al. \(2000\)](#), even though immigrants to Sweden who are economically more successful tend to stay for a shorter time, this dynamic is driven by the fact that these immigrants originate mostly from other Nordic countries. These authors established a negative association, conditional on the source country, between immigrants' incomes and the likelihood of leaving Sweden within five or 10 years after arrival.

Also for Sweden, [Nekby \(2006\)](#), by allowing out-migration rates to change non-linearly along the earnings distribution, showed that for both return and onward migrants, emigration rates are U-shaped with respect to earnings, with high probabilities of emigration for individuals in both low- and high-income groups. This finding is in line with [Dustmann's \(2003\)](#) results for Germany. In that study, using a simple theoretical framework to analyze immigrants' migration durations, he showed that for very low wages, durations increase with wages, while for intermediate- and high-income groups, the effect of wages on the time immigrants spend in Germany is negative. [Bijwaard and Wahba \(2014\)](#) also identified a U-shaped relation between incomes and out-migration probabilities by applying a competing risks model to register data on the entire population of labor immigrants from developing countries to the Netherlands during 1999–2007. Specifically, after modeling transitions between labor market states in the host country and the absorbing state of being in the country of origin, they computed out-migration probabilities for different income groups. They found that such probabilities are U-shaped with respect to income, with the highest probability of leaving among migrants in the lowest income group. They interpreted this finding as an indication that some immigrants leave because of disappointing economic outcomes in the host country, while the migration durations of others are governed by target saving behavior. Yet

for immigrants to Israel who were at least age 18 at the time of immigration, [Beenstock et al. \(2010\)](#) found no association whatsoever between immigrant earnings in 1983 and the individual still residing in the country in 1995.

2.4 Other characteristics and out-migration

Several studies investigated the relation between out-migration and individual characteristics other than earnings and education, including the possibility that out-migration probabilities increase around retirement age. [Waldorf \(1995\)](#), for instance, reported that among immigrants to Germany, the intention to return within the next four years increases close to retirement age, while [Cobb-Clark and Stillman \(2008\)](#) found among immigrants to Australia that actual out-migration rates are highest at retirement age. For the US, on the other hand, by computing annual emigration rates for various subgroups of immigrants (based on repeated CPS interviews with the same households), [Van Hook et al. \(2006\)](#) revealed that out-migration rates are higher for younger immigrants. A positive effect of age on the probability of remaining in the host country was also reported by [Bijwaard \(2010\)](#) for immigrants to the Netherlands, although [Edin et al. \(2000\)](#) found no such significant relation for immigrants to Sweden.

In a study of the relation between probability of leaving and age at immigration, [Aydemir and Robinson \(2008\)](#) found that immigrants who arrived in Canada at age 25–29 are slightly less likely to emigrate than immigrants who arrived at older ages (see also [Michalowski \(1991\)](#) for earlier work on the scale and composition of emigrant flows from Canada). [Jensen and Pedersen \(2007\)](#) also provided evidence for Denmark that the relation between age at immigration and out-migration differs by immigrant country of origin. That is, like [Aydemir and Robinson \(2008\)](#), they found that for immigrants from industrialized countries, age at immigration is negatively correlated with the probability of out-migration. They also showed, however, that the correlation is positive for men but insignificant for women from developing countries.

Other factors linked in the literature to emigration decisions include negative health shocks and economic prospects in the migrant's region of origin. As regards the former, [Sander \(2007\)](#), in an analysis of the effect of health shocks on emigration decisions among immigrants in Germany, revealed that negative health shocks do indeed affect the probability of out-migration, although the effect appears to depend on whether these shocks are transitory or permanent. That is, although the likelihood of out-migration increases when shocks to health are transitory, it decreases when they are more permanent. In terms of the relation between return migration and the economic prospects in migrants' region of origin, [Lindstrom \(1996\)](#) reported lower conditional return probabilities for Mexican immigrants in the US who are from economically more active regions in Mexico, an outcome he explained by the higher value that savings accumulated in the US have for these migrants on their return. [Reyes and](#)

Mameesh (2002) also found that economically more active and urban regions in the US attract more permanent migrants.

Naturally, out-migration rates vary by immigrant status: Refugees are considerably less likely to leave the host country than economic migrants, and part of the variations in out-migration across immigrants from different countries of origin can be explained by this (e.g., Duleep, 1994; Lundh and Ohlsson, 1994; Klinthäll, 2007; Aydemir and Robinson, 2008). The way in which emigration rates vary among different immigrant groups in the Netherlands is the subject of a study by Bijwaard (2010), who demonstrated that emigration hazards are higher for labor migrants than for individuals who immigrated for family reasons. Out-migration probabilities for various immigrant groups computed by Van Hook and Zhang (2011) from their CPS repeated household survey data also suggest that economic integration and social ties within the US play a major role in determining emigration probabilities, although the direction of causality is unclear.

Obviously, because of space constraints this list of studies is far from exhaustive; rather, this overview of research on the dimensions of foreign-born emigration and selection is a mere indication of the wide range of aspects investigated. In general, the papers reviewed suggest not only that migrants who leave are unlikely to be randomly drawn from the population of entry cohorts but that the direction of out-migration selection is far from homogeneous across either immigration countries or across different immigrant groups to the same country. This observation in turn suggests that any conclusions drawn for one country about the character of out-migration are unlikely to carry over to another country, and perhaps not even to another group of immigrants in the same country.

3. ESTIMATING IMMIGRANTS' CAREER PROFILES

Because immigrants' performance in the destination country has implications for many economic questions, its assessment is a key area of economic research. At the same time, however, it presents a challenging task, particularly given the nonpermanent nature of migration. One particularly problematic aspect is that individual immigrants make their economic decisions in conjunction with their migration plans while considering expected economic conditions in the country of origin (or countries to which they wish to migrate in the future), which adds substantial complexity to otherwise well-understood decisions like investment in human capital or labor supply issues. A second is that individual migrants out-migrate from and remigrate back to the host country in a way that leads to selective out-migration and remigration over the migration cycle. These two issues are inherently connected in that the nature of the selection through return and repeat migration is determined by the immigrants' dynamic life-cycle decisions.

In this chapter, we focus on the second problem, the estimation of immigrants' outcome profiles in the presence of selective out-migration. In doing so, we assume that immigrants' decisions on human capital investment, labor supply, and job choice are not determined by their migration and out-migration plans, a strong assumption but one made in almost all the literature.

Before discussing the issues more formally, however, we need to outline the major questions of interest to the researcher in estimating immigrants' earnings profiles. These questions can then serve as a point of reference in our subsequent discussion of different estimators and data sources, which also outlines the assumptions under which the questions can be answered.

3.1 Key research questions and immigrants' career profiles

As already emphasized, the economic performance of immigrants is of key interest to destination countries and often provides important information for migration policies. Here, to simplify the discussion, we focus on one particular aspect of economic performance: immigrants' earnings over their migration cycle. One key question in the debate over immigration is how much immigrants contribute to the tax and welfare system relative to what they receive in terms of benefits; in other words, what are immigrants' net fiscal contributions.⁶ Those at the lower end of the earnings distribution are typically net receivers of transfers, while those further up are net contributors. A typical query on immigrant earnings, therefore, refers to immigrants arriving in a particular year and those who remain in subsequent years:

Q1: *What is the growth in mean earnings of the populations of immigrants who arrived in the host country in a particular year and who are observed there in subsequent years?*

Because this question refers explicitly to migrants who live and work in the host country in subsequent years net of those who have left the host country, it concerns the selected populations of stayers, not the entire population defined by the arrival cohort. In [Figure 10.1](#), the earnings growth that responds to this question is given by the steeper of the two lines depicted, which refers to the increase in the mean earnings of the two truncated distributions. This parameter can be obtained from repeated cross-sectional data by simply computing the change in the mean earnings of a particular arrival cohort over time using individuals who are in the country at different points in time after the arrival year. The answer to Q1 is thus a combination of the earnings growth of immigrants who arrived in a particular year and stayed in the destination country in subsequent years and that of the population of surviving immigrants, which is an outcome of the compositional changes caused by (possibly selective) out-migration.

⁶ See, for example, [Dustmann and Frattini \(2013\)](#). See [Kerr and Kerr \(2011\)](#) for a recent survey and [Preston \(2013\)](#) for a discussion of the methodological challenges this literature faces.

Answering Q1 based on repeated cross-sectional data, however, reveals little about whether out-migration is selective because it does not allow compositional effects to be disentangled from individual wage growth. Nevertheless, in many instances, it may be valuable to better understand whether, and in which direction, out-migration is selective and how the earnings of an entry cohort of immigrants would have evolved over time in the host country had nobody left the country. For instance, if one study objective is to help design migration policies that regulate entry, then we need to understand the hypothetical career paths of immigrants in a particular arrival cohort had nobody left the country in subsequent years. If a further objective is to design policies ensuring that the highest performing immigrants remain in the country, then we want to understand the direction, and possibly the drivers, of selective out-migration. Thus, in many instances we may want to answer the following question:

Q2: *What would the growth in mean earnings of the population of immigrants who arrived in a particular year be if there was no subsequent out-migration?*

Unlike Q1, which can be answered based on data moments obtained from repeated cross-sectional data, Q2 requires the construction of a counterfactual scenario (designated in [Figure 10.1](#) by the flatter solid line) in which no migrants have left the destination country after arrival and the mean earnings of immigrants who arrived in a particular year can be observed for all immigrants in subsequent years. As discussed in more detail below, however, constructing this counterfactual situation under plausible assumptions requires more information than is available in repeated cross-sectional data.

A third question of possible interest refers to the mean earnings of a population of immigrants who all belong to the same entry cohort and all stayed in the destination country until T years later; for example, immigrants from Hong Kong living in Canada in 2010 who arrived in 1997, the year Hong Kong's sovereignty was transferred. This question therefore addresses the earnings growth of a cohort of immigrants who survive in the destination country until a certain date:

Q3: *What is the growth in mean earnings of the populations of immigrants who arrived in the host country in a particular year and who all stayed there until at least T years after arrival?*

Two scenarios exist under which the answers to all these questions are identical. The first, in which all migrations are permanent, is trivial. Yet such permanency is assumed in much of the literature on estimating immigrants' earnings profiles even though the studies reviewed in [Section 2](#) suggest that this assumption is implausible for almost all migration situations. Under the second scenario, the process that affects out-migration is exogenous to the process that affects immigrant earnings; in other words, out-migration is independent of earnings. This assumption, although it seems generally implausible, may be less restrictive for particular situations when conditioning on observables and when data on many individual characteristics of immigrants is available to the researcher.

In the next section, we will discuss the conditions under which we can identify the parameters that answer questions Q2 and Q3 above, and what information is needed over and above that contained in repeated cross-sectional data.

3.2 Estimation and identification of immigrant career profiles

In estimating immigrant career profiles, we first consider the following equation:

$$w_{it}^c = \mu_t^c + \varepsilon_{it}^c \quad (10.1)$$

where w_{it}^c is the log earnings of individual i of entry cohort c in year t , $t \geq c$, which can be decomposed into μ_t^c , the mean log earnings of individual i 's arrival cohort in period t if the entire immigrant cohort would stay in the host country, and ε_{it}^c the individual specific deviation of i 's earnings from the cohort mean, which depends on unobservable characteristics that affect earnings. We assume that the latter consist of an individual-specific and time-constant component plus a time-variant component, $\varepsilon_{it}^c = \alpha_i + e_{it}^c$. This model can easily be generalized by writing μ_t^c as a function of observable characteristics, so that, for example, $\mu_{it}^c = \bar{\mu}_t^c + x_{it}'b$.

We further assume that out-migration is an absorbing state; that is, once a migrant has left the country, he or she will never return. Out-migration in any year after arrival is then characterized by the following selection equation:

$$s_{it}^c = \prod_{c < \tau \leq t} 1[z_{i\tau}^c \beta + u_{i\tau}^c > 0]$$

where $\mathbf{1}[A]$ is an indicator function that takes the value $\mathbf{1}$ if A is true, and z_{it}^c and u_{it}^c are observed and unobserved characteristics respectively, that affect the migrant's decision to remain in the host country beyond year t . We also assume that $s_{it}^c = 1$ for $t = c$ (meaning that the population of interest is all immigrants who arrive in year c) and that $E(\varepsilon_{it}^c | s_{it}^c = 1) = 0$ for all t . The latter means that the expectation of the unobserved term in the outcome equation is equal to zero for the population of all immigrants who arrive at time c . This normalization not only defines the arrival cohort as the base population but implies that the conditional expectation will be zero for all subsequent periods t in the case that all migrations are permanent.

To focus on selection in out-migration, we also assume that conditional on s_{it}^c , ε_{it}^c is independent of μ_{it}^c , and that u_{it}^c is mean independent of z_{it}^c . Selection in this model is thus determined by the assumptions about the correlation between u_{it}^c and ε_{it}^c . This also implies that selection in this model is on earnings *levels*. In abstracting from selection on earnings *growth* we follow the vast majority of the literature. We will show, however, that even in this simpler case, the identification of the direction of selective out-migration requires considerably more model restrictions than generally acknowledged in the literature on the earnings assimilation of immigrants.

The assumption that out-migration is an absorbing state implies that $s_{it}^c = 1$ will always mean that $s_{it-1}^c = 1$; that is, a migrant observed in period t is also observed in period $t - 1$. To allow for more complex out-migration and remigration patterns (e.g., repeat migration) would require us to model not only the out-migration process but also the remigration process, which would add considerable complexity and additional identification issues. Hence, although such a model would certainly be interesting, it is beyond the scope of this chapter.

We first consider what could be identified in terms of the earnings growth of a cohort of immigrants that arrived in year c and were observed in two consecutive cross-sectional datasets, years $t_1 \geq c$ and $t_2 > t_1 \geq c$, if only cross-sectional data were available. In this case, the earnings growth of entry cohort c between t_1 and t_2 is given by

$$E\left(w_{it_2}^c | \mu_{t_2}^c, s_{it_2}^c = 1\right) - E\left(w_{it_1}^c | \mu_{t_1}^c, s_{it_1}^c = 1\right) = \left[\mu_{t_2}^c - \mu_{t_1}^c\right] + \left[E\left(\varepsilon_{it_2}^c | \mu_{t_2}^c, s_{it_2}^c = 1\right) - E\left(\varepsilon_{it_1}^c | \mu_{t_1}^c, s_{it_1}^c = 1\right)\right] \quad (10.2)$$

The first difference, $\mu_{t_2}^c - \mu_{t_1}^c = \Delta\mu_{t_1 t_2}^c$, is the earnings growth of the arrival cohort c from year t_1 to year t_2 had nobody out-migrated. Although this parameter answers Q2, it is identified only if the second difference, the expectation of the individual-specific deviations from the cohort means *conditional* on individuals residing in the host country in periods t_1 or t_2 , is equal to zero, which will trivially be the case if all migrations are permanent.

If migrations are nonpermanent, however, the last term in brackets will not be equal to zero, except for the special case in which ε is mean independent of both mean earnings and the selection rule: $E(\varepsilon | \mu, u, z) = 0$. In that case, conditional on any observables, the process that determines immigrants' earnings is unaffected by the process that determines out-migration. When out-migration is "exogenous" and the last two last terms in (10.2) equal zero, a simple OLS estimation of equation (10.1) identifies μ_t^c and yields answers to Q1, Q2, and Q3.

While this may be implausible in most applications, it may in some cases be justifiable to assume that selection is independent of ε *conditional* on a set of observable exogenous variables. For instance, if emigration is random *within* country-of-origin education cells ($E \cdot O$) but not *between* them, then conditioning on country of origin and education in a flexible way will eliminate the selection bias, as now $E(\varepsilon | u, z, E \cdot O) = 0$. Thus, conditional on $E \cdot O$, out-migration is independent of ε (see [Cameron and Trivedi \(2005, p. 863ff\)](#), or [Wooldridge \(2010, p. 908ff\)](#) for general discussions of such estimators).

In most cases, however, selection is a function of both observables and unobservables that are correlated with determinants of earnings. To illustrate, assume $\text{var}(u_{it}^c) = 1$, and that $E(\varepsilon_{it}^c | u_{it}^c, \dots, u_{it}^c) = \sigma_{eu}(t)u_{it}^c$ —as is the case if, for example, contemporaneous unobservables in the earnings and selection equations are jointly normally distributed, while

being uncorrelated across time—and consider the conditional expectation $E(\epsilon_{it}^\epsilon | \mu_t^\epsilon, s_{it}^\epsilon = 1)$: it follows from our assumptions that

$$\begin{aligned}
 E(\epsilon_{it}^\epsilon | \mu_t^\epsilon, s_{it}^\epsilon = 1) &= E(\epsilon_{it}^\epsilon | s_{it}^\epsilon = 1) \\
 &= E(\epsilon_{it}^\epsilon | u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta, \dots, u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta) \\
 &\quad + E(\alpha_i | u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta, \dots, u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta) \\
 &= \sigma_{eu}(t)E(u_{it}^\epsilon | u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta, \dots, u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta) \\
 &\quad + E(\alpha_i | u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta, \dots, u_{it}^\epsilon > -z_{it}^{\epsilon'}\beta)
 \end{aligned} \tag{10.3}$$

where $\sigma_{eu}(t)$ is the covariance between u_{it}^ϵ and e_{it}^ϵ .⁷ Thus, the bias we obtain in estimating $\Delta\mu_{t_1t_2}^\epsilon$ depends on the assumptions made about the correlation between unobservables in the earnings and selection equations. In the above case, the first conditional expectation on the right-hand side reflects selective out-migration determined through time-variant unobservables in the selection equation that are correlated with those in the outcome equation, while the second conditional expectation reflects a situation in which selection depends on time-constant individual-specific fixed effects. Whereas the latter corresponds to selection being systematically related to the immigrants' unobserved and time-invariant productivity, the former reflects selective out-migration that is determined by time-variant shocks to earnings being correlated with time-variant unobservables in the selection equation. Thus, when only repeated cross-sectional data are available, we cannot identify $\Delta\mu_{t_1t_2}^\epsilon$ without assuming that both α_i and e_{it}^ϵ are uncorrelated with selection. Nor will much information emerge about the direction of selection. Repeated cross-sectional analysis, therefore, provides no answer to Q2.

3.2.1 Stock sampled data

One way to move forward in the estimation of $\Delta\mu_{t_1t_2}^\epsilon$ is to use stock sampled data, which is becoming increasingly available either from surveys or administrative sources. One particular design links data on immigrants surveyed at a particular point in time to administrative data that allows the reconstruction of their past employment and earnings histories. Such data are likely to become more available as more efforts are made to link survey information with administrative data sources. In these datasets, the base population is all immigrants of a particular arrival cohort who remain in the host country until at least period T (determined by the year of the survey) and for whom longitudinal data are available for several years between t and T . Whereas the administrative data tend to contain no information on, for example, year of arrival or precise immigrant status, such information can be added from the survey data to produce an informative data source on migrant behavior. Nonetheless, these datasets do not typically provide longitudinal information on the entire entry cohort; rather, the migrant sample is determined by

⁷ We maintain the assumption of a unit variance for the residual in the selection equation throughout.

the migrants of any entry cohort that have remained in the host country at least until the survey year. Given the restriction that only those immigrants who stay until period T are observed, the advantage of stock sampled over repeated cross-sections for the purpose of identifying immigrant earnings profiles is not the longitudinal dimension, but the information that all individuals who are observed in earlier periods are known to stay until at least period T , as will become clear in our discussion below. Nevertheless, most datasets of that format used in the literature are longitudinal. For example, [Lubotsky \(2007\)](#)⁸ linked information on immigrants in the CPS sample who are still residing in the US in 1994 to social security earnings records, thereby providing longitudinal information on these migrants back to 1951. By construction, this dataset contains information on immigrants' earnings only for immigrants observed in the 1994 CPS.

What, then, do such data tell us? One key advantage of these combined datasets over (repeated) cross-sectional data is that they provide a sample of individuals who are all known to stay in the host country until at least some period T . Thus, in any period before T , the analyst observes a sample that will survive in the host country until at least T , which leads to the same selection criterion no matter in which period individuals are observed. As we show below, this provides additional identifying information, over and above that available in repeated cross-sectional data, even if individuals cannot be linked across waves. Nevertheless, whether this is sufficient to identify earnings growth of the original entry cohort depends again on the nature of selective out-migration up to year T . Denoting the survey year by T , the growth of immigrants' earnings belonging to entry cohort c between years t_1 and t_2 , with $c \leq t_1 < t_2 \leq T$, is

$$\begin{aligned} E\left(w_{it_2}^c \mid \mu_{t_2}^c, s_{iT}^c = 1\right) - E\left(w_{it_1}^c \mid \mu_{t_1}^c, s_{iT}^c = 1\right) &= \left[\mu_{t_2}^c - \mu_{t_1}^c\right] + \left[E\left(\varepsilon_{it_2}^c \mid \mu_{t_2}^c, s_{iT}^c = 1\right)\right. \\ &\quad \left.- E\left(\varepsilon_{it_1}^c \mid \mu_{t_1}^c, s_{iT}^c = 1\right)\right] \end{aligned} \quad (10.4)$$

Note the difference between equations (10.2) and (10.4). Whereas in (10.2) the selection indicator refers to the particular year in which the cross-section was collected, in (10.4) selection for each wave refers to the same year, year T , reflecting that we look at a sample of immigrants who all remained in the destination country at least until period T . Thus, if the dataset is longitudinal, it will be a balanced panel of immigrants. No matter the nature of the selection, the expression in (10.4) always identifies the earnings growth of immigrants belonging to cohort c who remained in the country until period T . Estimations based on these data, therefore, answer Q3 for immigrants from cohort c who stayed in the host country until year T .

Yet do such data tell us anything about $\Delta\mu_{t_1 t_2}^c$; that is, the hypothetical earnings growth of entry cohort c if all individuals who entered in that year were observed in both

⁸ Earlier studies of immigrant earnings growth that use stock-based longitudinal data include [Hu \(2000\)](#) and [Duleep and Dowhan \(2002\)](#). See also our discussion in [Section 4](#).

periods t_1 and t_2 , and thus answer Q2, except for the trivial cases in which all migrations are permanent or $E(\varepsilon|\mu, u, z) = 0$? To examine this, we reconsider the selection term in (10.3) and assume that $E(e_{it}^c | u_{it}^c, \dots, u_{iT}^c) = \sigma_{eu}(t)u_{it}^c$, to rewrite $E(\varepsilon_{it}^c | \mu_{it}^c, s_T^c = 1)$ as

$$\begin{aligned} E(\varepsilon_{it}^c | \mu_{it}^c, s_T^c = 1) &= E(e_{it}^c | u_{it}^c > -z_{it}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) \\ &\quad + E(\alpha_i | u_{it}^c > -z_{it}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) \\ &= \sigma_{eu}(t)E(u_{it}^c | u_{it}^c > -z_{it}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) \\ &\quad + E(\alpha_i | u_{it}^c > -z_{it}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) \end{aligned} \quad (10.3')$$

where, because of the nature of the data, we condition on the same selection rule for each period. Suppose first that selection depends only on time-constant unobservables in the outcome equation, α_i , so that $\sigma_{eu}(t) = 0$. In this case, the selection term will be constant over time for the same individual, and an OLS estimation of equation (10.1) on the pooled t_1 to T cross-sections with years since migration indicators included would yield the following estimation equation:

$$w_{it}^c = \mu_{t_1}^c + \Delta\mu_{t_1(t_1+1)}^c \mathbf{1}[t = t_1 + 1] + \dots + \Delta\mu_{t_1 T}^c \mathbf{1}[t = T] + \varepsilon_{it}^c \quad (10.1')$$

where $\mathbf{1}[\cdot]$ is an indicator function, equal to 1 if its argument is true and zero otherwise. An OLS regression can then be run on (10.1') to identify wage growth $\Delta\mu_{t_1 t_2}^c$ (thereby also answering Q2), although doing so will not identify the entry-level earnings for the original cohort that arrived in period c . Rather, because the last term in (10.3') is unequal to zero for $\text{corr}(\alpha_i, u_{it}^c) \neq 0$, it will identify only the mean wage of those individuals of arrival cohort c that remained in the country until period T for the first period in which they were observed.

Assume now that the selection of out-migrants is related not only to the unobservable time-constant variables α_i but also to contemporaneous time-variant unobservables in the earnings equation,⁹ a quite plausible assumption for many applications. For instance, negative shocks to wages in a particular period may be correlated with unobservables in the selection equation and trigger an out-migration. In that case, an OLS estimator using stock-based data will not identify $\Delta\mu_{t_1 t_2}^c$ (neither will a difference (or fixed effects) estimator in case the dataset is longitudinal; see below). To illustrate, we consider the conditional expectation of earnings growth between periods t_1 and t_2 :

$$\begin{aligned} E(w_{it_2}^c - w_{it_1}^c | \Delta\mu_{t_1 t_2}^c, s_T^c = 1) &= \Delta\mu_{t_1 t_2}^c + \left[\sigma_{eu}(t_2)E(u_{it_2}^c | u_{it_2}^c > -z_{it_2}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) \right. \\ &\quad \left. - \sigma_{eu}(t_1)E(u_{it_1}^c | u_{it_1}^c > -z_{it_1}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) \right] \end{aligned}$$

The bias term in brackets will only vanish if $\Delta\sigma_{eu}(t) = \sigma_{eu}(t_2) - \sigma_{eu}(t_1) = 0$ and $E(u_{it_2}^c - u_{it_1}^c | u_{it_2}^c > -z_{it_2}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) = 0$, meaning that both the covariance σ_{eu} and (whenever $\sigma_{eu} \neq 0$) the selection threshold $-z_{it}^c{}'\beta$ are constant over time. In

⁹ Restricting the correlation in time-variant unobservables to contemporaneous realizations between ε_{it}^c and u_{it}^c simplifies the exposition, but this correlation could be generalized.

particular, the latter will be violated in most scenarios, as selection will depend on non-constant individual characteristics such as age or time already spent in a host country, so that these variables should be included in z_{it}^c . Hence, if selective out-migration is related to time-variant unobservables in the earnings equation, stock-based sampled data will in general not allow us to answer Q2. As stock-based samples are increasingly selective subsamples of the initial immigrant cohort for large T , individuals with on average high realizations of u are more likely to be contained in the sample. For these immigrants, the selection conditions $u_{it}^c > -z_{it}^{c'}\beta, \dots, u_{iT}^c > -z_{iT}^{c'}\beta$ are less binding, so that selection is less affected by changes in z or σ over time, and hence the bias from the correlation in time-variant unobservables decreases with T .

Can we, then, identify the *direction* of selective out-migration by comparing estimates from stock-based data with those for repeated cross-sectional data for a particular entry cohort, as is done for $\Delta\mu_{t_1 t_2}^c$ in a number of empirical studies (see below)? Remember first the bias arising when cross-sectional data are used. Assuming that the earnings residual of stayers is given as in equation (10.3), the bias can be decomposed into two parts: The first part derives from selection on time-constant unobservables, which is equal to

$$E\left(\alpha_i \mid u_{it_2}^c > -z_{it_2}^{c'}\beta, \dots, u_{it_1}^c > -z_{it_1}^{c'}\beta\right) - E\left(\alpha_i \mid u_{it_2}^c > -z_{it_2}^{c'}\beta, \dots, u_{it_1}^c > -z_{it_1}^{c'}\beta\right)$$

for unrestricted repeated cross-sections, and equals zero if a stock-based sample is available, for which the conditioning set in periods t_2 and t_1 is the same. Thus, if out-migrants are selected on time-constant unobservables only, OLS estimates on pooled cross-sections are likely to be larger (smaller) than those obtained from stock-based longitudinal data whenever out-migrants are negatively (positively) selected. If in addition selection is on time-varying unobserved determinants of immigrant earnings, the bias from cross-sectional data is augmented by

$$\sigma_{eu}(t_2)E\left(u_{it_2}^c \mid u_{it_2}^c > -z_{it_2}^{c'}\beta, \dots, u_{it_2}^c > -z_{it_2}^{c'}\beta\right) - \sigma_{eu}(t_1)E\left(u_{it_1}^c \mid u_{it_1}^c > -z_{it_1}^{c'}\beta, \dots, u_{it_1}^c > -z_{it_1}^{c'}\beta\right)$$

which generally will be non-zero. Taken together, estimates on cross-sectional and stock-based samples, if observed for the same time periods, will differ by

$$\begin{aligned} & E\left(\alpha_i \mid u_{it_2}^c > -z_{it_2}^{c'}\beta, \dots, u_{it_2}^c > -z_{it_2}^{c'}\beta\right) - E\left(\alpha_i \mid u_{it_2}^c > -z_{it_2}^{c'}\beta, \dots, u_{it_1}^c > -z_{it_1}^{c'}\beta\right) \\ & + \sigma_{eu}(t_2)E\left(u_{it_2}^c \mid u_{it_2}^c > -z_{it_2}^{c'}\beta, \dots, u_{it_2}^c > -z_{it_2}^{c'}\beta\right) - \sigma_{eu}(t_1)E\left(u_{it_1}^c \mid u_{it_1}^c > -z_{it_1}^{c'}\beta, \dots, u_{it_1}^c > -z_{it_1}^{c'}\beta\right) \\ & - \left[\sigma_{eu}(t_2)E\left(u_{it_2}^c \mid u_{it_2}^c > -z_{it_2}^{c'}\beta, \dots, u_{iT}^c > -z_{iT}^{c'}\beta\right) - \sigma_{eu}(t_1)E\left(u_{it_1}^c \mid u_{it_1}^c > -z_{it_1}^{c'}\beta, \dots, u_{iT}^c > -z_{iT}^{c'}\beta\right) \right] \end{aligned} \quad (10.4')$$

the sign of which is informative about the direction of selection only in very special cases. Supposing, for example, that immigrants are target savers and hence are more likely to

out-migrate if they experience a positive earnings shock ($\sigma_{eu} < 0$), and that older individuals are more likely to leave ($-z_{it}^{\epsilon} \beta$ increases over time). In this quite plausible scenario, the above difference may be negative even if more productive individuals are generally more likely to stay ($\text{corr}(\alpha_i, u_{it}^{\epsilon}) > 0$), e.g., because they face lower integration costs. A comparison between estimates obtained from unrestricted cross-sectional and stock-based samples in this case is not informative about the direction of selection.

To address such more general types of selection, we need to model the process that determines selection and obtain an estimate for the selection term. Such modeling, however, is generally not possible with stock-based data, which do not indicate who leaves the country but only who has survived until period T . What is needed, therefore, is information on those individuals who leave the country.

3.2.2 Complete longitudinal data

Administrative datasets, which are now available for many countries, allow immigrant cohorts to be followed from entry onward and throughout their migration history. For example, assuming that complete longitudinal data are available for a cohort of immigrants who entered the country at time c , each individual in that dataset would be observed for a maximum number of years $\bar{T}$ (determined by the last year for which the survey is available) or until the year that individual leaves the country. Contrary to the stock sampled data discussed above, such data provide information on who left the country between $t-1$ and t .¹⁰ Furthermore, if immigrants are observed from the time of arrival, the mean earnings level μ_c^{ϵ} of the initial immigrant cohort c can be determined. Further, we can always construct a stock-based sample from complete longitudinal data by conditioning on survival until any year T . In other words, we can answer Q1 and Q3 for different survival cohorts and answer Q2 under the same assumptions as made in the previous sections.

Longitudinal data, however, also allow us to give up some of the restrictive assumptions on $\sigma_{eu}(t)$ necessary for identification of $\Delta\mu_{t_1 t_2}^{\epsilon}$. To illustrate, we first remember our assumption that emigration is an absorbing state; that is, immigrants who have left the country will never return. Even if this is not the case, we can construct such a dataset from longitudinal data by discarding all individuals who dropped out of the sample for some periods. As with out-migration, some migrants who were living in the country in period t_1 have disappeared by period t_2 . Hence, for these migrants, we observe no time-variant characteristics for period t_2 that are not changing predictably. We do, however, observe all time-invariant characteristics, as well as characteristics that change systematically (e.g., age). We therefore need to assume that the process determining

¹⁰ Of course, there may be other reasons why individuals drop out of the data, such as panel attrition or transitioning into sectors not covered by administrative data. We ignore these problems in the present analysis.

out-migration does not depend on time-variant observables that refer to period t_2 and change unpredictably.¹¹

To focus on selection correlated with the time variant unobservables e_{it}^c , consider the earnings equation in (10.1) written in differences:

$$\Delta w_{it}^c = \Delta \mu_{t_1 t_2}^c + \Delta e_{it}^c \quad (10.1'')$$

This eliminates α_i from the earnings equation and any selection related to it. Given that being in the country in period t implies $s_{it-1}^c = 1$ (i.e., the individual was in the country in the previous period), a selection equation conditional on $s_{it-1}^c = 1$ can be written as

$$s_{it}^c = \mathbf{1}[z_{it}^{c'} b_t + v_{it}^c > 0] \quad (10.5)$$

Equation (10.5) can thus be seen as a reduced form selection equation in which all explanatory variables from (10.1), and possibly their leads and lags, can be included in z_{it}^c . As explained above, we cannot include in z_{it}^c time-variant variables realized in period t that change unsystematically because these are not observed if the migrant has left the country in t ; thus, we need to assume that such variables do not affect out-migration, which may be restrictive in certain applications.

Assuming now that $v_{it}^c | (z_{it}^c, s_{it-1}^c = 1) \sim N(0, 1)$, and $E(\Delta e_{it}^c | \Delta \mu_{t, t-1}^c, z_{it}^c, v_{it}^c, s_{it}^c = 1) = r_{\Delta e, v_t} v_{it}^c$, where $r_{\Delta e, v_t}$ is the covariance between Δe_t and v_t , the expectation of wage growth in (10.1'') for individuals for whom $s_{it-1}^c = 1$, conditional on that individual also being observed in period t , can now be written as

$$\begin{aligned} E(\Delta w_{it}^c | \Delta \mu_{t, t-1}^c, z_{it}^c, s_{it}^c = 1) &= \Delta \mu_{t, t-1}^c + E(\Delta e_{it}^c | \Delta \mu_{t, t-1}^c, z_{it}^c, v_{it}^c, s_{it}^c = 1) \\ &= \Delta \mu_{t, t-1}^c + r_{\Delta e, v_t} \lambda(z_{it}^{c'} b_t) \end{aligned}$$

where the last equality follows from the normality assumption and $\lambda(z_{it}^{c'} b_t) = \phi(z_{it}^{c'} b_t) / \Phi(z_{it}^{c'} b_t)$ is the inverse Mills ratio (Heckman, 1979). Because $s_{it-1}^c = 1$, whenever $s_{it}^c = 1$, there is no need to condition on $s_{it-1}^c = 1$.

Wooldridge (2010, p. 837f) suggested estimating this model by first estimating simple probit models for each time period to obtain estimates of the inverse Mills ratios and then estimating pooled OLS of Δw_{it}^c on $\Delta \mu_{t, t-1}^c$ and $\mathbf{1}[t = c + 1] \hat{\lambda}_{it}, \dots, \mathbf{1}[t = T] \hat{\lambda}_{it}$, where the $\mathbf{1}[t = \tau]$ are dummy variables, with $\mathbf{1}[t = \tau] = 1$ if $t = \tau$. Doing so yields the following equation, which can be estimated using pooled OLS:

$$\Delta w_{it}^c = \Delta \mu_{t, t-1}^c + \mathbf{1}[t = c + 1] r_{\Delta e_{(c+1)} v_{(c+1)}} \hat{\lambda}_{it} + \dots + \mathbf{1}[t = T] r_{\Delta e_T v_T} \hat{\lambda}_{it} + \xi_{it}$$

¹¹ It should be noted that this case is different from standard panel data applications in which we usually observe all variables for individuals who have selected or not selected into a particular state in a particular period, as, for example, in the estimation of wage equations when labor force participation is selective (see, e.g., Wooldridge (1995) and Kyriazidou (1997) for estimators of these models, and Dustmann and Rochina-Barrachina (2007) for a comparison and application).

We can now use a simple F -test of the null hypothesis that $r_{\Delta e_t v_t}$ are jointly equal to zero to test whether out-migration depends on time-variant characteristics that affect earnings. If the null hypothesis is rejected, the estimated OLS standard errors must be adjusted for generated regressor bias.

Although the literature continues to be dominated by selection corrections derived from the assumption of jointly normally distributed unobservables, following Newey (2009), the above attrition correction can be extended to semiparametric estimators in which the predicted inverse Mills ratio $\lambda(z_{it}'\hat{b}_t)$ is replaced in the differenced wage equation by an unspecified function $\varphi(z_{it}'\hat{b}_t)$ of the linear index of the selection equation. The latter can be estimated using the semiparametric estimator suggested by Klein and Spady (1993), which does not rely on normally distributed residuals. $\varphi(z_{it}'\hat{b}_t)$ can then be approximated with a polynomial of $z_{it}'\hat{b}_t$ (see, e.g., Melenberg and van Soest (1996) and Martins (2001) for applications). Table 10.2 summarizes the discussion of identification of immigrants' earnings profiles in terms of the simple model above.

It should be noted that in order to focus on the key issues related to selective out-migration, we have abstracted from several problems that may also affect the estimation of earnings equations. First, if fixed effects estimators are to be used, the explanatory variables in the earnings equation must be strictly exogenous.¹² Otherwise, their predetermination would lead to an endogenous regressor bias in the differenced equation. For example, this assumption of strict exogeneity would be violated were tenure introduced into the level earnings equation, as shocks to wages in the last period might induce individuals to change firms, which would reset their tenure clocks (see Dustmann and Meghir (2005) for a discussion). Addressing this problem requires IV-type estimators.

Table 10.2 Identification of immigrant earnings profiles under selective out-migration

	Selection on time-constant unobservables	Selection on both time-constant and time-varying unobservables
Unrestricted repeated cross-sections		
μ_c	Not identified	Not identified
$\Delta\mu_{t_1 t_2}$	Not identified	Not identified
Stock-based samples		
μ_c	Not identified	Not identified
$\Delta\mu_{t_1 t_2}$	Identified	Not identified
Longitudinal data		
μ_c	Identified if panel starts at $t=c$	Identified if panel starts at $t=c$
$\Delta\mu_{t_1 t_2}$	Identified	Identified

Unrestricted repeated cross-sections

μ_c	Not identified	Not identified
$\Delta\mu_{t_1 t_2}$	Not identified	Not identified

Stock-based samples

μ_c	Not identified	Not identified
$\Delta\mu_{t_1 t_2}$	Identified	Not identified

Longitudinal data

μ_c	Identified if panel starts at $t=c$	Identified if panel starts at $t=c$
$\Delta\mu_{t_1 t_2}$	Identified	Identified

¹² The difference estimator maintains its consistency under the slightly weaker condition that $E(\Delta e_{it}^c | \Delta \mu_{it}^c) = 0$.

Second, it may be desirable to distinguish between the different factors that affect immigrant wage growth; for example, labor market experience, time in the destination country, and time effects. Yet a level equation focused on one particular entry cohort does not separately identify the effects of time and period of residence in the host country unless further assumptions are made. One common practice is to assume the same time effects for immigrants and natives (Borjas, 1985); however, there is increasing evidence that this assumption may be violated (see, e.g., Borjas and Bratsberg, 1996; Barth et al., 2004; Dustmann et al., 2010). Hence, it is impossible to distinguish between wage growth stemming from time effects and that resulting from period of residence in the host country without additional identifying assumptions.¹³ Additional assumptions are also needed when estimating earnings equations for immigrants in differences if the level effects of years of residence and potential experience are to be identified separately.

3.3 Numerical example

We now illustrate the possible biases from selective out-migration given different assumptions about the correlation between unobservables in the earnings and selection equations, when these are not sufficiently accounted for. We use a simple Monte Carlo experiment that considers only one immigrant cohort. Suppose that the log earnings for immigrant i in one particular entry cohort, net of the effect of observable characteristics other than time spent in the host country, evolve as

$$w_{it} = \mu_t + \alpha_i + e_{it} \equiv \mu_o + \gamma ysm_{it} + \alpha_i + e_{it} \quad (10.6)$$

where for simplicity log earnings w_{it} of immigrant i in period t are specified as linear in years since immigration (ysm), so that $\Delta\mu_{t_1 t_2} = \gamma$ for all t . α_i is a time-constant individual-specific component, and e_{it} includes unobserved factors, which are assumed to be independent and identically distributed across individuals and time and independent of anything else on the right-hand side of the equation. Suppose also that the selection rule for an immigrant remaining in the host country is given by

$$s_{it} = \prod_{\tau \leq t} \mathbf{1}[z_{i\tau}'\beta + u_{i\tau} \equiv \beta_0 + \beta_1 ysm_{i\tau} + u_{i\tau} > 0]$$

then out-migration is an absorbing state (i.e., once out-migrated, an individual will not reappear in the dataset at a later point in time).¹⁴

¹³ No such problem exists for natives because the quality of new entry cohorts is usually assumed not to change over time.

¹⁴ Throughout this simulation exercise, we specify that $\mu_o = 2$, $\gamma = 0.02$, $\beta_0 = 0.5$, $\beta_1 = 0.05$, $\alpha \sim N(0, 0.2)$, $e \sim N(0, 0.2)$, and $u \sim N(0, 1)$. We generate a sample of 100,000 individuals (100 for Figure 10.6) who, to abstract from other issues, are all assumed to be observed from the date of immigration up to 30 years for those who do not out-migrate.

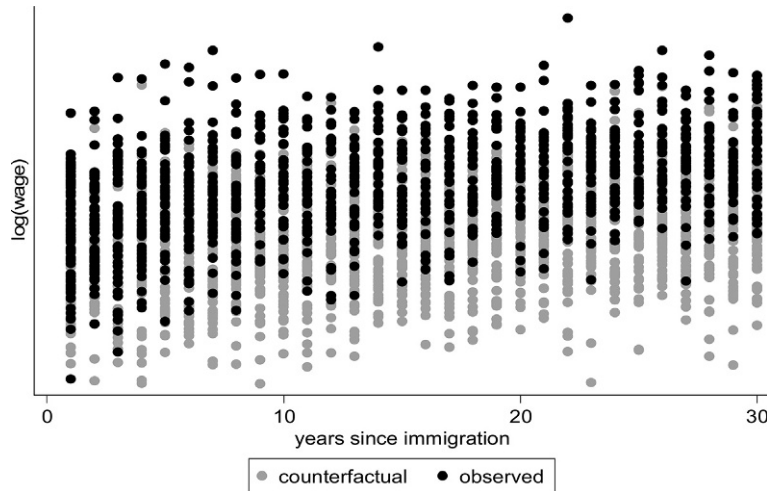


Figure 10.6 *Observed and counterfactual log-earnings.*

Assuming first that u_{it} is correlated only with α_i and not with e_{it} ¹⁵ (i.e., selection is on individual-specific time-constant unobservables in the earnings equation), then data generated in our simulation are as shown in Figure 10.6, where the black dots represent the observed log earnings of immigrants still residing in the host country and the gray dots the immigrants from the original arrival cohort who out-migrated. Given our assumptions about the nature of selection (with selection into staying being positively correlated with unobserved productivity α_i), out-migration is negatively selective.

Table 10.3 lists the results of using the OLS estimator for equation (10.6), assuming that the data we have available are either repeated cross-sections (column 2), stock-based data, where immigrants who remain at least 5, 15, or 25 years are observed throughout (columns 3–5), or complete longitudinal data (column 6). The OLS estimates in column 2 show a strong upward bias when applied to the pooled cross-sections. Estimating (10.6) using stock-based data restricted to immigrants that stay for at least a pre-specified number of years and allowing for different cut-off years T yields estimates of the slope parameters that are close to the true parameter values.¹⁶ Hence, if selection occurs on time-constant individual fixed effects in the earnings equation only, OLS on stock-based samples produces parameter estimates that answer both Q2 and Q3.

However, the restriction to individuals remaining for a minimum number of years and the positive correlation between u_{it} and α_i tend to exclude immigrants with the lowest realizations of α_i (and increasingly so when we constrain the sample to survival at

¹⁵ We assume that $\text{corr}(\alpha, u) = 0.7$.

¹⁶ For small sample sizes, restricting the sample to very short panels may leave too little variation in the explanatory variable and incline estimates to attenuation bias.

Table 10.3 Selection on time-constant unobservables only

	True coefficients	OLS, all observations	OLS, stayed for 5 or more years	OLS, stayed for 15 or more years	OLS, stayed for 25 or more years	FE estimates
<i>cons</i>	2	2.277921 (0.0010001)	2.341263 (0.0028334)	2.438709 (0.0016322)	2.450425 (0.0012547)	—
<i>ysm</i>	0.02	0.0278496 (0.0000591)	0.021146 (0.0008531)	0.0199223 (0.0001797)	0.0200594 (0.0000845)	0.0199664 (0.0000534)

higher T values). Hence, the intercept of the earnings equation (reflecting the entry-level earnings of the respective arrival cohort) is overestimated relative to the intercept of the original arrival cohort. Nevertheless, this parameter does provide an estimate for the entry-level earnings of immigrants from a particular arrival cohort that survived until period T in the destination country. Finally, the last column of Table 10.3 reports the estimates from a within-group estimation, which, similar to estimation on the stock-based samples, eliminates selection on time-constant fixed effects.

The estimators discussed above, however, although widely used in the literature, produce no consistent estimates if we relax the assumption that selection is correlated with time-constant unobservables only. To illustrate this problem, we assume that in addition to α_i , there is contemporaneous correlation between u_{it} and e_{it} ,¹⁷ and that only the stock-based sample data are available. In this case, the mean earnings observed in period t , conditional on being observed in $t-1$, are given by

$$\begin{aligned}
 E(w_{it} | s_{iT} = 1) &= \mu_t + E(\alpha_i | s_{iT} = 1) + E(e_{it} | s_{iT} = 1) \\
 &= \mu_t + E(\alpha_i | u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta) \\
 &\quad + E(e_{it} | u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta)
 \end{aligned} \tag{10.7}$$

and given our normality assumption, OLS identifies $\Delta\mu_{t_1 t_2} + \sigma_{eu}(t_2)E(u_{it_2} | u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta) - \sigma_{eu}(t_1)E(u_{it_1} | u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta)$.

While, as discussed in Section 3.2.1, the conditional expectation of α_i is constant over time, the conditional expectation of e_{it} may change if (for $\sigma_{eu} \neq 0$) either z_{it} (and thus the attrition probability) or the correlation between out-migration and time-variant unobservables in the earnings equation changes over time. Although either case is sufficient to bias estimates, in fact both apply in our simulated data.

In our simulations we have assumed that the (positive) correlation between e and u , $\sigma_{eu}(t)$, decreases over time, which induces a negative bias. Further, the increase in the probability of a migrant choosing not to leave the host country in a given period implies a reduction of the threshold $-z_{it}'\beta$, above which realizations of u_{it} are required for

¹⁷ We assume in the simulation that upon immigration, $\text{corr}(e_s, u_t) = 0.7$ for $s = t$ but $\text{corr}(e_s, u_t) = 0$ for $s \neq t$, with the correlation decreasing over time by 10% per year.

the migrant to stay, so that $E(u_{it_2} | u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta) < E(u_{it_1} | u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta)$, reinforcing the negative bias. Columns 3–5 of Table 10.4 show these downward biased estimates for our illustrative simulation. Because the remaining sample becomes increasingly selected with respect to the initial immigrant cohort when T is larger, the bias in the estimated slope parameter becomes less severe as the time that immigrants must stay to be included in the stock-based sample increases. This is because a subsample with a high T has, on average, high realizations of α_i and u_{it} , so that the selection conditions $u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta$ become increasingly less binding and the bias decreases—which is particularly visible in our estimates in Table 10.4 when we move from $T=5$ to $T=15$. On the other hand, the relatively small bias of the within-group estimator is specific to our simulated population and depends on the model parameters.

Can we learn something about the direction of the selection of immigrants by comparing estimates based on repeated cross-sectional data, and estimates based on stock-based samples? Estimates on the unrestricted repeated cross-sections continue to be above those obtained from stock-based data (column 2). However, other than in the previous case (Table 10.3), the bias consists now of two opposing parts: First, the same upward bias resulting from selection on time-constant unobservables, as in Table 10.3. Second, a downward bias resulting from selection on time-variant unobservables. Thus, the overall repeated cross-sections bias of the OLS estimates in column 2 is smaller in Table 10.4 than in Table 10.3. In this example, the downward bias again arises from both the decrease in σ_{eu} and the decrease in $-z_{it}'\beta$ over time, and its magnitude differs for repeated cross-sections (obtained as the difference in the estimated slope coefficients in Tables 10.2 and 10.3, $0.02785 - 0.02427 = 0.00358$) and for stock-based samples (ranging from $0.02 - 0.008605 = 0.011395$ for $T=5$ to $0.02 - 0.018289 = 0.001711$ for $T=25$), due to the different conditioning rules (see equation (10.4') above). Hence, when comparing estimates obtained from stock-based samples and from unrestricted repeated cross-sections, this part of the bias does not simply cancel out, and such comparisons are uninformative about the direction of immigrant selection, unless one assumes that selection is on time-constant unobservables only (see also our discussion in Section 3.2.1).

Table 10.4 Selection on time-constant and time-varying unobservables

	True coefficients	OLS, all observations	OLS, stayed for 5 or more years	OLS, stayed for 15 or more years	OLS, stayed for 25 or more years	FE estimates	Corrected estimates
<i>cons</i>	2	2.352842 (0.0008919)	2.436231 (0.0024729)	2.485823 (0.0015218)	2.48313 (0.0011942)	—	—
<i>ysm</i>	0.02	0.0242654 (0.000055)	0.0086049 (0.0007664)	0.016204 (0.0001715)	0.0182887 (0.0000818)	0.0184484 (0.0000528)	0.0209794 (0.001225)

The same holds true for fixed effects estimates (column 6). Like for stock-based samples, the bias here is only due to selection on time-varying unobservables. However, for the same reason as above, a comparison of estimates with those obtained from repeated cross-sections is uninformative about the direction of immigrant selection.

As previously explained, addressing selective out-migration that works through a correlation between time-variant unobservables in the selection and earnings equations requires the specification of a selection rule. Such specification is impossible, however, with stock sampled data because those who leave the country are not observed. Longitudinal data, on the other hand, do allow us to observe those who emigrate in period t in all periods $\tau < t$, which enables specification of a selection equation. We thus continue to assume that data are generated by the process specified above but that the complete (unbalanced) panel is observed, so that an individual is observed until he or she leaves the host country.

Because the assumption that out-migration is an absorbing state means that $s_{it}=1$ always implies $s_{it-1}=1$, to identify the slope parameter $\Delta\mu_{t_1t_2}^c$, we can specify the following selection equation for the differenced earnings equation conditional on $s_{it-1}=1$:

$$s_{it} = \mathbf{1}[b_0 + b_1 w_{it-2} + v_{it} > 0]$$

where w_{it-2} is chosen as an instrument that satisfies the exclusion restriction $E(\Delta e_{it} | w_{it-2}) = 0$ while being correlated with selection via the individual-specific effect α_i . As we discuss in [Section 3.2.2](#), this reduced form selection equation can contain any variable that is observed for all individuals for whom $s_{it-1}=1$ and that is informative about selection between periods $t-1$ and t . This also can include realizations of variables in period t if these variables change systematically over time. While in many applications age and other individual characteristics are likely to be determinants of the decision to stay in the host country, our simple simulation example specifies that only years since immigration and the unobserved components in u_{it} affect selection. For expositional purposes, we assumed all immigrants to arrive at the same point in time. Hence, since we estimate a probit model for each year separately, years since migration do not contain additional information for selection. However, since stayers are selected not only on time-varying unobservables e_{it} , but also on the time-constant unobservables α_i , earnings lagged by two periods will help identify the selection equation, while being uncorrelated with $\Delta e_{it} = e_{it} - e_{it-1}$.

As explained in [Section 3.2.2](#), for each period t , we use a probit estimator to estimate the selection equations and compute the inverse Mills ratio, which we then insert into the differenced earnings equation as additional regressors:

$$\Delta w_{it} = \gamma \Delta ysm_{it} + r_t \frac{\phi(\hat{b}_0 + \hat{b}_1 w_{it-2})}{\Phi(\hat{b}_0 + \hat{b}_1 w_{it-2})} + \Delta e_{it} \quad (10.8)$$

By estimating this model, we obtain the results given in the last column of [Table 10.4](#). Time-varying coefficients r_t allow for the covariance between the residual of the differenced earnings equation and the selection equation to change with time since

immigration. An F -test on the OLS estimates obtained from equation (10.8) rejects joint insignificance of r_t , i.e., that the selection correction terms have no effect on the change in earnings, at the 1% level.¹⁸

3.4 Interpretation: A simple model of return migration

To interpret the estimated direction of selectivity and to fix ideas about possible sources of selective out-migration driven by skill endowment and accumulation, we extend the work of Borjas and Bratsberg (1996) to produce a simple model of temporary migration.¹⁹ In this model, log earnings in the origin country o and destination country d take the following form:²⁰

$$w_l = \mu_l + \varepsilon_l \quad (10.9)$$

where μ_l , $l = o, d$, denotes the mean log earnings or rental rate of human capital in location l , and ε_l is the deviation in the productive capacity (or human capital) of an individual-working in country l from the mean, which is determined by an individual's (observed and unobserved) human capital. We abstract from variation in ε_l over time, so that it corresponds to the time constant α in the previous subsections. We further assume that the ε_l have a mean of zero and that $\varepsilon_d = \pi \varepsilon_o$, where π is the price of skills in the destination country relative to that in the origin country, implying that all skills are perfectly transferable across countries. It follows that $\text{corr}(\varepsilon_d, \varepsilon_o) = 1$. If the price for skills is higher in the destination than in the origin country (i.e., $\pi > 1$), this implies that the variance in earnings is higher in the host country. In this model, skills are *one-dimensional* so that individuals can be ranked on them, and the ranking of individuals on productive capacity is the same in both countries. Individuals know both their skills and the relative skill price π .

According to Borjas and Bratsberg (1996), workers have three options: to stay at home and not migrate at all, to migrate temporarily, or to migrate permanently. Of these, temporary migration may be optimal when having been abroad increases human capital that is valuable at home by a certain amount $\kappa = \tau k$. We extend these authors' model by assuming that this gain in human capital varies with the period of stay in the destination country, in which case earnings when emigrating and returning are given by

¹⁸ As would be expected, it does not reject joint insignificance in the differenced earnings equation in the case where selection is on time-constant unobservables only.

¹⁹ See Dustmann et al. (2011) for a generalization of this model to multiple skills and a dynamic setting and Dustmann and Glitz (2011) for a simplified version.

²⁰ As first discussed in Dustmann (1995), we focus on a human capital accumulation in the host country that has a higher value in the home country as a motive for return. Borjas and Bratsberg (1996) also considered a lower than expected return in the host country as a reason for return migration.

$$w_{do} = \tau(\mu_d + \varepsilon_d) + (1 - \tau)(\mu_o + \varepsilon_o + \tau k) \quad (10.10)$$

where τ denotes the fraction of an individual's working life spent abroad. Assuming that individuals try to maximize income, they will choose to stay in the country of origin if $w_o > w_d$ and $w_o > w_{do}$, choose to migrate permanently if $w_d > w_o$ and $w_d > w_{do}$, and choose to migrate temporarily if $w_{do} > w_o$ and $w_{do} > w_d$. Hence, ignoring the costs of migration, the skill thresholds for which no migration, temporary migration, and permanent migration are optimal can be easily derived by substituting (10.9) and (10.10) into the conditions above.

In Figure 10.7, we illustrate the case in which $\pi > 1$ (i.e., the variance in earnings, as well as the price of skills, is higher in the destination country). Given the skill distribution shown here, those with the lowest skills (with ε_o below the threshold $(\mu_o - \mu_d - k)/(\pi - 1)$) will decide to stay in the country of origin, those with the highest skills (above the threshold $(\mu_o - \mu_d + k)/(\pi - 1)$) will decide to emigrate and remain permanently, and those between the two thresholds will decide to emigrate but will return migrate after spending some time τ abroad. For $\pi > 1$, therefore, selection in this model leads those with higher productive capacity to emigrate, and among those who decide to emigrate, motivates those with the highest productive capacity to stay permanently. It should also be noted that in this simple model, the migration selection depends only on the relative skill price π . The selection of emigrants and remigrants in this model is the exact opposite when the price of skills is higher in the sending country (i.e., $\pi < 1$). Thus, in this model, temporary migrants are always predicted to be drawn from the middle of the skill distribution.

The model further predicts that an increase in the rental rate of human capital in the destination country, μ_d , will cause the two thresholds to shift to the left, whereas an increase in the home country of the value of human capital acquired abroad, κ , will result in a widening of the distance between them. Because in this simple illustrative model gains from a stay abroad can only be realized in the country of origin, a large value for κ makes temporary migration an attractive choice, so that only individuals

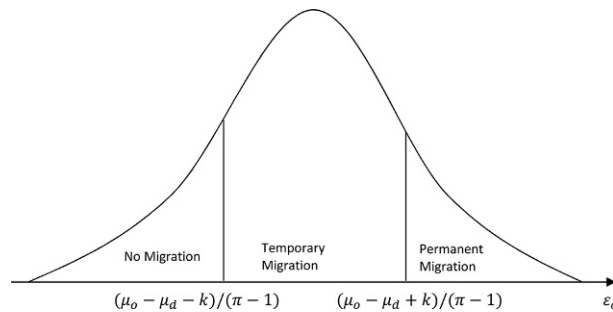


Figure 10.7 Selection of emigrants and remigrants under higher ε returns to skills in the host country.

with very low (very high) ε_0 choose to stay permanently in the country of origin (destination). On the other hand, a high relative return to skills in the host country, π , brings the two thresholds closer together, implying that for most individuals with below (above) average skill endowment, staying permanently in the country of origin (destination) is the preferred option. In this model, therefore, temporary migrants can be hierarchically sorted (Willis, 1986); that is, clearly ranked on skills relative to non-migrants and permanent migrants dependent on the skill prices in the two countries.

In this model, the optimal migration duration for those who decide to emigrate but remain in the host country only temporarily is determined by the first-order condition of w_{do} with respect to τ :

$$\tau = \frac{[\mu_d - \mu_o] + [\pi - 1]\varepsilon_0 + k}{2k}$$

Hence, for $\pi > 1$, the optimal migration duration for temporary migrants increases with initial skill endowment ε_0 . Given the opposite case, however (i.e., the price of skills is higher in the country of origin), it will be higher for low-productivity individuals.

As emphasized in other sections, this relation between the time immigrants choose to stay in the host country and their productive capacity has important implications for estimating their earnings profiles. According to our simple model, three dynamics can be predicted: (i) if skill prices are higher abroad, then immigrants will be positively selected from the population of the origin country; (ii) the migration duration for those who emigrate will increase with productivity; and (iii) those with the highest levels of productivity will decide to migrate permanently. This scenario is thus compatible with negatively selective return migration in the sense that of those who emigrate, the lowest productivity individuals will return first. If $\pi < 1$ (i.e., the return to skills is higher in the destination country), those with higher skills will return sooner.

To explicitly relate these observations to the earlier discussion on earnings profile estimation, we assume that skills are normally distributed, with $\varepsilon_o \sim N(0, 1)$ and $\varepsilon_d \sim N(0, \pi^2)$, and $Cov(\varepsilon_d - \varepsilon_o, \varepsilon_o) = (\pi - 1)$. Consider now the mean earnings of those who have emigrated to the destination country and who are observed there in period t :

$$E(w_d | \tau > t) = \mu_d + \pi E(\varepsilon_o | \tau > t) = \begin{cases} \mu_d + \pi E(\varepsilon_o | \varepsilon_o > q(t)) = \mu_d + \pi \lambda(-q(t)) & \text{if } \pi > 1 \\ \mu_d + \pi E(\varepsilon_o | \varepsilon_o < q(t)) = \mu_d - \pi \lambda(q(t)) & \text{if } \pi < 1 \end{cases}$$

where $q(t) = k(2t - 1) - [\mu_d - \mu_o]/(\pi - 1)$, and λ is the inverse Mills ratio. This calculation raises a number of important issues. First, it is clear that here the entry wage of a particular cohort is composed of the mean wage obtainable by the average individual from the

home country who migrates to the destination country, μ_d , and a term that reflects the selection of migrants from the overall population:

$$E(w_d | \tau > 0) = \begin{cases} \mu_d + \pi\lambda(-q(0)) > \mu_d & \text{if } \pi > 1 \\ \mu_d - \pi\lambda(q(0)) < \mu_d & \text{if } \pi < 1. \end{cases}$$

Hence, the mean entry wage of the original arrival cohort (denoted as μ_t^c above) depends on the degree of initial immigrant selection (see [de Coulon and Piracha \(2005\)](#) for a similar model on the selection of emigrants and return migrants from their origin societies). If $\pi > 1$, the entry wage will be larger than that a randomly drawn individual from the home country would earn in the host country, meaning that emigration is positively selective. Also worth noting is that the rule governing return migration depends on the same unobserved productivity term as the earnings equation, ε_o . Thus, not only non-migrants and permanent migrants but also temporary migrants with lower or higher durations of stay can be strictly ordered in terms of their underlying skills to produce an especially simple selection mechanism in which out-migration truncates the skill distribution. This case is the one shown in [Figure 10.1](#), where selective out-migration increasingly eliminates immigrants from the lower part of the earnings distribution. Therefore, the earnings of immigrants in the host country at any period t follow a truncated distribution, one whose truncation is based on the outcome variable, log earnings.

Next, we consider determining wage growth by estimating earnings regressions based on repeated cross-sectional data. Because in this simple model μ_d is constant, the wage growth of the original arrival cohort if all migrations were permanent, $\Delta E(w_d(t))$, equals zero. Hence, when $\pi > 1$ (i.e., out-migration is negatively selective), the wage growth between periods t_1 and $t_2 > t_1$ obtained from repeated cross-sectional data is

$$E(w_d(t_2) | \tau > t_2) - E(w_d(t_1) | \tau > t_1) = \pi[\lambda(-q(t_2)) - \lambda(-q(t_1))]$$

The last term in brackets is always positive for $\pi > 1$: The inverse Mills ratio decreases in its argument, and $q(t)$ increases in t . This leads to an overestimation of the wage growth of the original arrival cohort had nobody return migrated if return is negatively selected. When $\pi < 1$, on the other hand, the term is negative, generating an underestimation if return migration is positively selective.

If we re-examine these same issues using stock sampled data, then

$$E(w_d(t_2) | \tau > T) - E(w_d(t_1) | \tau > T) = \pi[\lambda(-q(T)) - \lambda(-q(T))]$$

where $w_d(\tau)$ is the migrants' earnings in the host country during period τ . Because the last term in brackets equals zero, our stock sampled data produces an unbiased estimate of $\Delta\mu_t^c$ thanks to the special selection type induced in this simple model, in which the time-constant unobservables governing selection are the same as those affecting wages.

4. EXISTENT STUDIES ON THE ESTIMATION OF EARNINGS EQUATIONS WHEN OUT-MIGRATION IS NONRANDOM

Analyzing immigrant assimilation in terms of wages and other native population outcomes has been at the core of economic migration research for many decades (see [Dustmann and Glitz \(2011\)](#) for a survey). Beginning with [Chiswick's \(1978\)](#) analysis of the earnings adjustment of male immigrants to the US and [Long's \(1980\)](#) similar investigation for foreign-born women, a large body of literature has emerged on the estimation of immigrants' earnings profiles. Yet these early studies are often criticized for their reliance on single cross-sectional sample data (e.g., [Chiswick, 1978](#); [Long, 1980](#); and [Carliner, 1980](#), all use 1970 US census data), which do not allow a differentiation to be made between cohort effects and time of residence effects as an individual in 1970 that has been in the US for 10 years must have arrived in 1960, and an individual that has been in the US for 20 years in 1950. Such data thus permit no distinction between wage growth after arrival in the US and differences in immigrants' initial earnings positions after arrival. Moreover, whereas the early literature implicitly assumes that these cohort differences equal zero, [Borjas \(1985\)](#) stressed that, if the quality of successive cohorts deteriorates, this may lead to overestimation of assimilation profiles. [Borjas \(1985\)](#), [LaLonde and Topel \(1992\)](#), [Chiswick and Miller \(2010\)](#), and other researchers addressed this problem by employing repeated cross-sections, which allows addressing of this problem under some assumptions.

As we have already demonstrated, similar issues arise when immigrant out-migration is selective. Even with repeated cross-sectional data, assimilation estimates can still be biased when selective out-migration drives what seem to be changes in cohort quality (see, e.g., [Chiswick, 1986](#)). Several early studies took advantage of longitudinal data to address this problem. [Borjas \(1989\)](#), for example, used longitudinal information from the 1972–78 Survey of Natural and Social Scientists and Engineers (which is based on immigrant listings in the 1970 US census) to analyze the earnings paths of a particular immigrant group. He found that among foreign-born scientists and engineers in the sample, attrition—which is assumed to be largely driven by emigration from the US—is more likely for individuals with less favorable economic outcomes. Because this dataset includes earnings information dating back to 1969, it allowed him to perform separate estimations for the 1969–71 earnings of immigrants who stayed until 1978 versus those who left between 1972 and 1978. The results suggest that both initial earnings and earnings growth are lower for immigrants who leave the sample later. In subsequent work, [Pischke \(1992\)](#) addressed the potential bias in the estimation of the earnings assimilation of immigrants in Germany by including fixed effects in his panel estimates, while [Lindstrom and Massey \(1994\)](#) estimated log wage regressions using different samples of Mexicans residing in the US and Mexicans who returned to their native country. Based on their comparison of estimates from single versus repeated cross-sectional data, however, they concluded that selective emigration is unlikely to affect estimates of wage assimilation.

4.1 Studies using stock sampled longitudinal data

A number of more recent studies use stock sampled longitudinal data on immigrants in the US. [Hu \(2000\)](#), for example, compared assimilation profiles from longitudinal social security records matched to the Health and Retirement Survey (HRS) with profiles from the cross-sectional decennial census. The lower immigrant earnings growth suggested by the stock-based longitudinal data (as compared to the repeated cross-sectional data) is consistent with negatively selective out-migration. This interpretation, however, relies on the assumption that the selection of out-migrants occurs on time-constant unobservables only (see our discussion in [Section 3.2](#)). Moreover, whereas the census estimates suggest that net of age, education, and time effects, non-Hispanic white immigrants experience an earnings increase, with the 1950–59 and 1965–69 immigrant cohorts catching up with US-born workers within 10 years of arrival, the longitudinal data indicate a decline in earnings residuals by time spent in the US for this population both in levels and relative to the US-born population. For Hispanic immigrants, both data sources indicate an earnings increase, although the HRS data suggest it takes about 35 years to catch up with US-born workers, much longer than the 20 or so years suggested by the census estimates. Nevertheless, because the HRS follows a relatively narrow birth cohort (1931–41), separating the effects of age at immigration or pre-migration experience and changes in average immigrant cohort earnings is difficult.

[Lubotsky \(2007\)](#) addressed this difficulty by using a broader sample, based on 1951–97 social security earnings records matched to the 1990 and 1991 Survey of Income and Program Participation (SIPP) and March supplement to the 1994 CPS, to estimate a log earnings equation with cohort fixed effects, time since immigration indicators, and a number of human capital variables on the right side. He then compared the estimates of the time since immigration effects with those obtained from the 1970–90 cross-sectional census data. Consistent with the hypothesis of negatively selective out-migration, the wage profiles obtained from the longitudinal data are flatter.²¹ Like Hu, however, [Lubotsky \(2007\)](#) used stock-based longitudinal data, and—recalling our discussion in [Section 3.2.2](#)—this interpretation relies on the assumption that selection occurs only on time-constant unobservables that affect earnings. Under this assumption and given the format of his data, Lubotsky estimated his earnings equation in levels and only included cohort- rather than individual-specific indicators. Consistent with the greater likelihood that it is low-earning immigrants who will leave, the stock-based longitudinal data indicate a less pronounced deterioration in cohort earnings on entry to the

²¹ A second concern addressed in [Lubotsky \(2007\)](#) is that because of alternative arrival cohort definitions, it cannot be ruled out that migrants, when moving back and forth, have in fact spent time in the US before the stated date of entry. If the incidence of repeat migration increases more over time among low-earnings migrants, then the deterioration in average cohort earnings will be overstated.

US. In an earlier study that used part of the same stock-based longitudinal data as Lubotsky (2007), Duleep and Dowhan (2002) analyzed immigrant assimilation in the US at different quantiles of the earnings distribution and for different years of immigration (cohorts arriving between 1960 and 1983). As estimates are not compared to, e.g., results that would be obtained from repeated cross-sectional data, the direction of selection cannot be determined from their analysis even under the assumption that out-migrants are selected only on time-constant unobservable components of the earnings equation. In their interesting paper using historical data of immigration to the US, Abramitzky et al. (2013) constructed a panel of US residents from 1900–20 census data. Since the panel is restricted to individuals who are still observed in 1920, the data on immigrants amounts to a stock-based sample. Their smaller assimilation estimates obtained from the stock-based panel compared to those from cross-sectional data are consistent with out-migrants being negatively selected, but this conclusion rests on the assumptions regarding the selection process that we discuss above.

For Canada, Picot and Piraino (2012), in large part following Lubotsky (2007), reported that although earnings growth rates based on cross-sectional data are overestimated, the earnings *gap* between Canadian and foreign-born workers appears to evolve similarly whether cross-sectional or stock-based longitudinal data are used, the reason being that at the lower end of the earnings distribution, attrition tends to increase to a similar degree among both immigrants and Canadians, leaving no obvious differences between cross-sectional and longitudinal data in terms of earnings gap evolution.

4.2 Studies using longitudinal data

Whereas the US data used by Lubotsky (2007) and others generally do not follow immigrants from the beginning of their stay in the host country, a number of European studies use datasets that do, thereby enabling analysis of the differences in earnings assimilation between short- and longer-term migrants. In addition, because longitudinal data that are not restricted to migrants residing in the host country until some period T are informative about which migrants leave the host country, in principle they allow selection to be explicitly modeled. Most existing studies, however, in using individual fixed effects to address the potential inconsistency in estimated earnings profiles, maintain the assumption that selection occurs on time-constant unobservables only.

Such an approach was taken in an early paper by Pischke (1992), who estimated the earnings assimilation of immigrants in Germany using data from the German Socio-Economic Panel (GSOEP), which should eliminate any possible bias from selection on individual-specific unobservables. Likewise, Edin et al. (2000) used Swedish register data on immigrants arriving in Sweden between 1970 and 1990 to show that over a quarter of immigrants aged between 18 and 55 on arrival leave the country within five years. In line with the US literature, they also found that emigration is more likely among

economically less successful migrants. Then, under the assumption that emigration may vary with earnings levels but not with earnings growth, they demonstrated that if out-migration is not taken into account, earnings assimilation by OECD immigrants is overestimated by about 90%, a figure largely in line with Lubotsky (2007). Arai (2000), however, criticized their results on negatively selective emigration from Sweden on the grounds that much of this finding is driven by the higher mobility of young migrants and the positive correlation of age with earnings, together with a number of sampling issues.

Sarvimäki (2011) analyzed immigrant assimilation in Finland using longitudinal data that follow immigrants from the time of their arrival. To address selective emigration, he compared OLS estimates of immigrant earnings growth based on the whole sample with estimates based only on immigrants that stay for at most five years. Although he does show that short-term migrants experience no earnings growth, he acknowledged that the direction of selection is unclear and his results merely indicate that immigrants who stay are not a random sample of their initial arrival cohort. In terms of the model we use in Section 3, and given that he estimated earnings equations in levels, the differences in earnings growth would be consistent with a variety of scenarios. If, for instance, most of the selection occurs during the first few years and out-migrants are positively selected on time-constant unobservables only, then we expect the OLS estimates of earnings growth to be strongly downwardly biased in any sample that is restricted to short-term migrants. In the presence of negative selection of out-migrants on time-constant unobservables, on the other hand, the differences in the estimates of earnings growth could arise if during the first years after immigration, immigrants tend to leave when facing negative earnings shocks and the effect of time-variant unobservables on selection decreases with the time immigrants have spent in the host country ($\sigma_{eu}(t) > 0$ and $d\sigma_{eu}/dt < 0$). The positive but decreasing covariance between the time-variant components e_{it} and u_{it} induces a negative bias from selection on time-varying unobservables that may dominate the positive bias due to selection on time-constant individual effects, implying lower estimated earnings growth for short-term migrants.

Barth et al. (2012), on the other hand, in their analysis of the role of native-immigrant differences in job mobility, found little difference between short- and long-term migrants in either between- or within-firm wage growth when restricting the sample by dropping immigrants who leave within the first five years after arrival. Nevertheless, their findings permit no conclusion that out-migration is random. Skuterud and Su (2009), using a rotating panel of immigrants and natives in Canada, found little difference between OLS and fixed-effect estimates of immigrant wage assimilation. They reconciled this observation with the contrary findings for stock-based samples from the US (e.g., Hu, 2000; Lubotsky, 2007) in two ways: First, they argued that in the Canadian case, emigrant selection may be less clear because many more able immigrants may move onward to the US. Second, they pointed out that if emigration is correlated with heterogeneity in wage

growth rather than wage levels, then it is possible that $E(\alpha_i | u_{it} > -z_{it}'\beta, \dots, u_{it} > -z_{it}'\beta) = 0$, so including fixed effects need not change the estimates of $\mu_t - \mu_{t-1}$ from an unbalanced panel like Skuterud and Su's (2009), which initially also contained migrants who later leave. However, if at least some immigrants can correctly anticipate wage growth and leave when expected wage growth is low, then, in a comparison of estimates based on cross-sectional and stock-based longitudinal data, $E(\mu_t | s_t = 1) - E(\mu_{t-1} | s_{t-1} = 1) > E(\mu_t - \mu_{t-1} | s_T = 1)$, so stock-based longitudinal data will predict flatter wage profiles for immigrants than estimates based on cross-sectional data, which would be an alternative explanation for the difference between repeated cross-sectional and stock-based longitudinal sample estimates reported by Lubotsky and others. Cobb-Clark et al. (2012) analyzed immigrant earnings in Australia; however, their very brief discussion paper did not specify which immigrant cohorts are analyzed or how they change over time. Nevertheless, their results do suggest that compared to longitudinal data, selective emigration tends to bias estimates of employment assimilation upwards when repeated cross-sectional data are used, but that this has less of an effect on wage assimilation profiles.

To analyze earnings assimilation between German-born workers and different immigrant cohorts, Fertig and Schurer (2007) used the German Socio-Economic Panel (GSOEP). As a panel, this dataset enables explicit modeling of out-migration. Specifically, these authors formulated a multiple equations selection model of individual earnings together with the probabilities of employment, survey participation, and staying in the host country under the assumption that unobservables are jointly normally distributed (see our discussion in Section 3.2.2) and include correction terms in the earnings regression. As exclusion restrictions, Fertig and Schurer included family and country-of-origin characteristics as explanatory variables in the selection equation. They only found evidence for assimilation, however, in some of their immigrant groups, possibly because most immigrants in their sample are longer-term migrants who arrived in the early 1970s, whereas the wage observations only begin with the first year of the survey in 1984. In addition, unlike our exposition in Section 3.2.2, these authors assumed that the covariance between unobservables in the earnings and selection equations is constant over time, although they did allow it to be immigrant cohort specific. More important, they included selection correction terms in the levels rather than the differenced earnings equation. This assumes that selective out-migration can be reduced to a static problem where—given observable control variables—selection in each period does not depend on past selection. This ignores that an immigrant being in the country in a given period depends on decisions made in previous periods. If, for instance, out-migration is an absorbing state then such an estimating equation would not appropriately correct for selective out-migration because it ignores whether or not an individual is observed depends on selection in earlier periods. It is important to note that by specifying selection

corrections for a differenced earnings equation, as we suggest above, one conditions on past selection, while this is not the case when corrections from static selection equations are simply applied to an earnings equation in levels. This same caveat applies to [Venturini and Villosio \(2008\)](#) and [Faini et al. \(2009\)](#), who used a similar framework for Italy.

[Biavaschi \(2013\)](#), working with US data, challenged the view that emigration is negatively selective. She formulated a selection model and then, using data from the 2000 US and Mexican censuses, semiparametrically estimated the counterfactual density of the wage residuals of Mexican immigrants in the US had there been no out-migration. Biavaschi used an identification at infinity argument to recover this counterfactual distribution if only one cross-section of data is available. She found that emigration is more likely among Mexican-born workers at the upper part of the wage distribution, which implies that in the absence of emigration, the Mexican-born population in the US would have higher wages. Assuming that out-migration is not correlated with unpredictably changing variables that affect wages, an alternative method of accounting for selective out-migration was proposed by [Kim \(2009\)](#). Specifically, he applied a weighting procedure to an overlapping rotating panel constructed from the merged outgoing rotation groups from 1994–2004 CPS data to produce a larger sample size than usually available from true longitudinal datasets. Like [Hu \(2000\)](#), he found that estimates of immigrants' economic assimilation based on repeated cross-sectional data are positive and upwardly biased but that the reverse is likely to be true when selective out-migration is taken into account.

The problem of selective out-migration is recognized well beyond studies of immigrant earnings assimilation. [Kaushal's \(2011\)](#) analysis of the returns to US versus overseas education, for example, showed that, based on longitudinal data from the National Survey of College Graduates, US-educated science and engineering professionals who stay in the country earn more than those who leave, which indicates a bias in cross-sectional estimates of education-dependent earnings trajectories. In this study, however, whether foreign-born individuals acquire their education in the US or abroad does not matter for emigration propensity. Examining the variation in earnings and the change in returns to skills rather than simply the earnings level, [Lubotsky \(2011\)](#) re-emphasized the need to use longitudinal data when emigration is selective. Using a similar longitudinal stock-based sample of immigrants to the US as in [Lubotsky \(2007\)](#), he argued that not only will estimates of wage growth be inconsistent if out-migration is selective, but that estimates of the impact of changes in the wage structure—e.g., due to the increase in the returns to skills since the 1980s—on the earnings gap between native and foreign-born workers will be affected as well. Other studies investigating various effects of selective out-migration on empirical estimates have been conducted by [Bratsberg et al. \(2010\)](#), who analyzed immigrant employment in Norway, by [de Matos \(2011\)](#) for immigrant career paths in Portugal, and [Kaushal and Shang \(2013\)](#) for wage assimilation among US immigrants in different destination areas.

5. CONCLUSIONS

In this chapter, we address selective out-migration—a key problem in estimating immigrant career profiles—by first giving evidence of the temporariness of many migrations and then reviewing the literature that assesses the degree to which emigration is selective along various dimensions. This literature review provides evidence not only that out-migration is heterogeneous with respect to country of origin, immigrant education, and immigrant earnings in the destination country, but that the direction and degree of selective out-migration, far from being uniform, differ across both immigration countries and different groups in the same immigration country.

To outline the potential methodological problems in estimating immigrant earnings profiles, we distinguish three important research questions related to immigrants' economic careers in the destination country. Although two of these are answerable by computing means from observable data, the third requires the construction of a counterfactual scenario. The first question, which is of particular interest to many researchers in the field, explores the evolution of mean earnings of a particular immigrant arrival cohort c that is part of a population that decides to stay in the destination country. Because this question refers to the population of immigrants that remains t years after immigration, it can be answered using repeated cross-sectional data. Researchers may also be interested, however, in such evolution for an entry cohort from among all migrants who survive in the host country until T years after migration. Because answering this second question requires the identification of all immigrants who survived until T in all years between c and T , it is dependent on the availability of stock sampled data of the type we describe above.

A third question concerns the earnings paths in the host country of the original arrival cohort if nobody out-migrates. If selection is correlated with unobservables in the outcome equation, estimates based on repeated cross-sectional data do not answer the question. Under restrictive assumptions about the nature of the selection process, stock-based sampled data will provide an answer to that question, and availability of both stock-based sampled data and repeated cross-sections allows signing the selection of out-migrants. However, given more general assumptions, data are needed that allow modeling of the out-migration selection process; for example, longitudinal data that are not stock sampled. In our discussion, therefore, we suggest selection patterns and various estimators of the parameter underlying this question.

The overall purpose of the chapter is to discuss the different research questions different types of data allow addressing, and under which assumptions. We illustrate how misleading many estimates of immigrant earnings assimilation may be in providing answers to particular questions if the research design does not take into account the possibility of selective out-migration. We also demonstrate the impossibility of making any general statement about the direction of selective out-migration, unless longitudinal data

are available that allow modeling out-migration, and that the direction and magnitude of selection may differ both across countries and across immigrant groups within the same country. Unfortunately, the latter implies that we may know far less about immigrant career profiles than the vast literature suggests. On the other hand, the availability of increasingly better data raises hopes that, in the near future, we will be able to more accurately assess immigrant progress in destination countries and the selection on their out-migration.

Overall, because of its strong consequences for all types of policy and such related areas as migration's impact on natives, assessment of immigrants' career profiles is vital to the economics of migration. Even if longitudinal data are available that allow assessment of selective out-migration, the key assumption that we made here is that the process that governs out-migration is independent of decisions that may determine the individual's investments into human capital or other labor market decisions. Hence, one slowly emerging body of literature not covered in this chapter, in estimating immigrant career paths, allows migrants to make their migration plans in conjunction with their economic decisions, including labor supply and human capital investments. Such estimation, however, requires that these decisions be modeled jointly with migration choices (as in Bellemare, 2007), a methodological challenge typical of the many problems that riddle research in this area despite steady progress in recent years. These very challenges, however, open up myriad promising avenues for future study on aspects that still need to be (and can be) addressed.

ACKNOWLEDGMENTS

We thank Albrecht Glitz and Martin Weidner for helpful comments, and Shanka Fernando for excellent research assistance. We acknowledge the support by the Norface programme on Migration. Dustmann acknowledges support by the European Research Council (ERC) Advanced Grant No. 323992.

REFERENCES

- Abramitzky, R., Boustan, L.P., Eriksson, K., 2012. Europe's tired, poor, huddled masses: self-selection and economic outcomes in the age of mass migration. *Am. Econ. Rev.* 102 (5), 1832–1856.
- Abramitzky, R., Boustan, L.P., Eriksson, K., 2013. A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration, Working Paper.
- Ahmed, B., Robinson, J.G., 1994. Estimates of Emigration of the Foreign-Born Population: 1980–1990, US Census Bureau Population Division Working Paper No. 9.
- Alders, M., Nicolaas, H., 2003. One in three immigrants leave within six years. Statistics Netherlands, online, 20 January.
- Arai, M., 2000. Comments on Edin, LaLonde and Åslund: emigration of immigrants and measures of immigrant assimilation: evidence from Sweden. *Swed. Econ. Pol. Rev.* 7, 205–211.
- Aydemir, A., Robinson, C., 2008. Global labour markets, return, and onward migration. *Can. J. Econ.* 41 (4), 1285–1311.

- Bandiera, O., Rasul, I., Viarengo, M., 2013. The making of modern America: migratory flows in the age of mass migration. *J. Dev. Econ.* 102, 23–47.
- Barth, E., Bratsberg, B., Raaum, O., 2004. Identifying earnings assimilation of immigrants under changing macroeconomic conditions. *Scand. J. Econ.* 106 (1), 1–22.
- Barth, E., Bratsberg, B., Raaum, O., 2012. Immigrant wage profiles within and between establishments. *Lab. Econ.* 19 (4), 541–556.
- Beaujot, R., Rappak, J.P., 1989. The link between immigration and emigration in Canada, 1945–1986. *Can. Stud. Popul.* 16 (2), 201–216.
- Beenstock, M., 1996. Failure to absorb: remigration by immigrants into Israel. *Int. Migrat. Rev.* 30 (4), 950–978.
- Beenstock, M., Chiswick, B.R., Paltiel, A., 2010. Testing the immigrant assimilation hypothesis with longitudinal data. *Review of Economics of the Household* 8 (1), 7–27.
- Beggs, J.J., Chapman, B., 1991. Male immigrant wage and unemployment experience in Australia. In: *Immigration, Trade and the Labor Market*. University of Chicago Press, pp. 369–384.
- Bellemare, C., 2007. A life-cycle model of outmigration and economic assimilation of immigrants in Germany. *Eur. Econ. Rev.* 51 (3), 553–576.
- Bhaskar, R., Arenas-Germosén, B., Dick, C., 2013. Demographic Analysis 2010: Sensitivity Analysis of the Foreign-Born Migration Component. U.S. Census Bureau Population Division Working Paper No. 98, May 2013.
- Biavaschi, C., 2013. Recovering the Counterfactual Wage Distribution with Selective Return Migration, Working Paper.
- Bijwaard, G.E., 2009. Labour Market Status and Migration Dynamics, IZA Discussion Paper No. 4530.
- Bijwaard, G.E., 2010. Immigrant migration dynamics model for the Netherlands. *J. Popul. Econ.* 23 (4), 1213–1247.
- Bijwaard, G.E., Wahba, J., 2014. Do high-income or low-income immigrants leave faster? *J. Dev. Econ.* 108, 54–68.
- Bijwaard, G.E., Schluter, C., Wahba, J., 2011. The impact of labor market dynamics on the return migration of immigrants. *Rev. Econ. Stat.*
- Böhning, W.-R., 1984. *Studies in International Labour Migration*. Macmillan, London.
- Borjas, G.J., 1985. Assimilation, changes in cohort quality, and the earnings of immigrants. *J. Labor Econ.* 3 (4), 463–489.
- Borjas, G.J., 1989. Immigrant and emigrant earnings: a longitudinal study. *Econ. Inq.* 27 (1), 21–37.
- Borjas, G.J., Bratsberg, B., 1996. Who leaves? The outmigration of the foreign-born. *Rev. Econ. Stat.* 78 (1), 165–176.
- Bratsberg, B., Raaum, O., Sørli, K., 2007. Foreign-born migration to and from Norway. *International Migration, Economic Development and Policy* 259–291.
- Bratsberg, B., Raaum, O., Røed, K., 2010. When minority labor migrants meet the welfare state. *J. Labor Econ.* 28 (3), 633–676.
- Cameron, A.C., Trivedi, P.K., 2005. *Microeconometrics: Methods and Applications*. Cambridge University Press.
- Carliner, G., 1980. Wages, earnings and hours of first, second, and third generation American males. *Econ. Inq.* 18 (1), 87–102.
- Carrión-Flores, C.E., 2006. What makes you go back home? Determinants of the duration of migration of Mexican immigrants in the United States. In: *Society of Labor Economists Annual Meeting*, Cambridge, MA.
- Chiswick, B., 1978. The effect of Americanization on the earnings of foreign-born men. *J. Polit. Econ.* 86 (5), 897–921.
- Chiswick, B.R., 1986. Is the new immigration less skilled than the old? *J. Labor Econ.* 4, 168–192.
- Chiswick, B.R., Miller, P.W., 2010. Negative assimilation of immigrants: a special case. *Industrial and Labor Relations Review* 64 (3), 502–525.
- Cobb-Clark, D., Stillman, S., 2008. Emigration and the Age Profile of Retirement among Immigrants. IZA Discussion Paper No. 3874, Institute for the Study of Labor (IZA).
- Cobb-Clark, D., Hanel, B., McVicar, D., 2012. Immigrant Wage and Employment Assimilation: A Comparison of Methods. IZA Discussion Paper No. 7062, Institute for the Study of Labor (IZA).

- Cohen, Y., Habersfeld, Y., 2001. Self-selection and return migration: Israeli-born Jews returning home from the United States during the 1980s. *Popul. Stud.* 55 (1), 79–91.
- Coniglio, N.D., De Arcangelis, G., Serlenga, L., 2009. Intentions to return of clandestine migrants: the perverse effect of illegality on skills. *Rev. Dev. Econ.* 13 (4), 641–657.
- Constant, A., Massey, D.S., 2002. Return migration by German guestworkers: neoclassical versus new economic theories. *Int. Migrat.* 40 (4), 5–38.
- Constant, A., Massey, D.S., 2003. Self-selection, earnings, and out-migration: a longitudinal study of immigrants to Germany. *J. Popul. Econ.* 16 (4), 631–653.
- Constant, A.F., Zimmermann, K.F., 2011. Circular and repeat migration: counts of exits and years away from the host country. *Popul. Res. Pol. Rev.* 30 (4), 495–515.
- Cortes, K.E., 2004. Are refugees different from economic immigrants? Some empirical evidence on the heterogeneity of immigrant groups in the United States. *Rev. Econ. Stat.* 86 (2), 465–480.
- de Coulon, A., Piracha, M., 2005. Self-selection and the performance of return migrants: the source country perspective. *J. Popul. Econ.* 18 (4), 779–807.
- de Matos, A.D., 2011. The Careers of Immigrants, **Working Paper**.
- Duleep, H.O., 1994. Social security and the emigration of immigrants. *Soc. Secur. Bull.* 57, 37.
- Duleep, H.O., Dowhan, D.J., 2002. Insights from longitudinal data on the earnings growth of US foreign-born men. *Demography* 39 (3), 485–506.
- Dustmann, C., 1993. Earnings adjustment of temporary migrants. *J. Popul. Econ.* 6 (2), 153–168.
- Dustmann, C., 1995. Savings behavior of return migrants. *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 115, 511–533.
- Dustmann, C., 2003. Return migration, wage differentials, and the optimal migration duration. *Eur. Econ. Rev.* 47 (2), 353–369.
- Dustmann, C., (with discussants Bentolila, S., Faini, R.), 1996. Return migration: the European experience. *Econ. Pol.* 213–250.
- Dustmann, C., Frattini, T., 2013. The Fiscal Effects of Immigration to the UK, CReAM DP 22/2013.
- Dustmann, C., Glitz, A., 2011. Migration and education. In: *Handbook of the Economics of Education*, vol. 4, pp. 327–439.
- Dustmann, C., Meghir, C., 2005. Wages, experience and seniority. *Rev. Econ. Stud.* 72 (1), 77–108.
- Dustmann, C., Rochina-Barrachina, M.E., 2007. Selection correction in panel data models: An application to the estimation of females' wage equations. *Econometrics J.* 10 (2), 263–293.
- Dustmann, C., Weiss, Y., 2007. Return migration: theory and empirical evidence from the UK. *Br. J. Ind. Relat.* 45 (2), 236–256.
- Dustmann, C., Glitz, A., Vogel, T., 2010. Employment, wages, and the economic cycle: differences between immigrants and natives. *Eur. Econ. Rev.* 54 (1), 1–17.
- Dustmann, C., Fadlon, I., Weiss, Y., 2011. Return migration, human capital accumulation and the brain drain. *J. Dev. Econ.* 95 (1), 58–67.
- Eckstein, Z., Weiss, Y., 2004. On the wage growth of immigrants: Israel, 1990–2000. *J. Eur. Econ. Assoc.* 2 (4), 665–695.
- Edin, P.-A., LaLonde, R.J., Åslund, O., 2000. Emigration of immigrants and measures of immigrant assimilation: evidence from Sweden. *Swed. Econ. Pol. Rev.* 7 (2), 163–204.
- Faini, R., Strøm, S., Venturini, A., Villosio, C., 2009. Are Foreign Migrants More Assimilated than Native Ones? IZA Discussion Paper No. 4639.
- Fertig, M., Schurer, S., 2007. Labour Market Outcomes of Immigrants in Germany: The Importance of Heterogeneity and Attrition Bias. IZA Discussion Paper No. 2915, Institute for the Study of Labor (IZA).
- Gundel, S., Peters, H., 2008. What determines the duration of stay of immigrants in Germany? Evidence from a longitudinal duration analysis. *Int. J. Soc. Econ.* 35 (11), 769–782.
- Heckman, J.J., 1979. Sample selection bias as a specification error. *Econometrica*, 153–161.
- Hu, W.-Y., 2000. Immigrant earnings assimilation: estimates from longitudinal data. *Am. Econ. Rev.* 90 (2), 368–372.
- Jasso, G., Rosenzweig, M.R., 1982. Estimating the emigration rates of legal immigrants using administrative and survey data: The 1971 cohort of immigrants to the United States. *Demography* 19 (3), 279–290.

- Jasso, G., Rosenzweig, M.R., 1990. The New Chosen People: Immigrants in the United States. National Committee for Research.
- Jensen, P., Pedersen, P.J., 2007. To stay or not to stay? Out-migration of immigrants from Denmark. *Int. Migrat.* 45 (5), 87–113.
- Kaushal, N., 2011. Earning trajectories of highly educated immigrants: does place of education matter. *Industrial and Labor Relations Review* 64, 323.
- Kaushal, N., Shang, C., 2013. Earnings Growth of Mexican Immigrants: New versus Traditional Destinations. Working Paper No. 7427, Institute for the Study of Labor (IZA).
- Kerr, S.P., Kerr, W.R., 2011. Economic Impacts of Immigration: A Survey. No. w16736, National Bureau of Economic Research.
- Kim, S., 2009. Economic Assimilation of Foreign Born Workers in the United States: An Overlapping Rotating Panel Analysis, No. UWEC-2008-19.
- Kirdar, M.G., 2009. Labor market outcomes, savings accumulation, and return migration. *Lab. Econ.* 16 (4), 418–428.
- Kirwan, F., Harrigan, F., 1986. Swedish–Finnish return migration, extent, timing, and information flows. *Demography* 23 (3), 313–327.
- Klein, R.W., Spady, R.H., 1993. An efficient semiparametric estimator for binary response models. *Econometrica*, 387–421.
- Klinthäll, M., 2003. Return Migration from Sweden 1968–1996: A Longitudinal Analysis. Almqvist & Wiksell International.
- Klinthäll, M., 2006. Immigration, integration and return migration. In: *International Symposium on International Migration and Development*, Turin, pp. 28–30.
- Klinthäll, M., 2007. Refugee return migration: return migration from Sweden to Chile, Iran and Poland 1973–1996. *J. Refug. Stud.* 20 (4), 579–598.
- Kyriazidou, E., 1997. Estimation of a panel data sample selection model. *Econometrica*, 1335–1364.
- LaLonde, R.J., Topel, R.H., 1992. The assimilation of immigrants in the US labor market. In: *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*. University of Chicago Press, pp. 67–92.
- Lam, K.-C., 1994. Outmigration of foreign-born members in Canada. *Can. J. Econ.* 27 (2), 352–370.
- Lindstrom, D.P., 1996. Economic opportunity in Mexico and return migration from the United States. *Demography* 33 (3), 357–374.
- Lindstrom, D.P., Massey, D.S., 1994. Selective emigration, cohort quality, and models of immigrant assimilation. *Soc. Sci. Res.* 23 (4), 315–349.
- Long, J.E., 1980. The effect of Americanization on earnings: some evidence for women. *J. Polit. Econ.* 88 (3), 620–629.
- Longva, P., 2001. Out-Migration of Immigrants: Implications for Assimilation Analysis. Department of Economics, University of Oslo, Memorandum no. 4/2001.
- Lubotsky, D., 2007. Chutes or ladders? A longitudinal analysis of immigrant earnings. *J. Polit. Econ.* 115 (5), 820–867.
- Lubotsky, D., 2011. The effect of changes in the US wage structure on recent immigrants' earnings. *Rev. Econ. Stat.* 93 (1), 59–71.
- Lukomskyj, O., Richards, P., 1986. Return migration from Australia: a case study. *Int. Migrat.* 24 (3), 603–632.
- Lundh, C., Ohlsson, R., 1994. Immigration and economic change. In: *Population, Economy, and Welfare in Sweden*. Springer, Berlin, pp. 87–107.
- Maré, D.C., Morten, M., Stillman, S., 2007. Settlement patterns and the geographic mobility of recent migrants to New Zealand. *New Zeal. Econ. Paper* 41 (2), 163–195.
- Martins, M.F.O., 2001. Parametric and semiparametric estimation of sample selection models: an empirical application to the female labour force in Portugal. *J. Appl. Econometrics* 16 (1), 23–39.
- Massey, D., 1987. Understanding Mexican migration to the United States. *Am. J. Sociol.* 92 (6), 1372–1403.
- Melenberg, B., Van Soest, A., 1996. Parametric and semi-parametric modelling of vacation expenditures. *J. Appl. Econometrics* 11 (1), 59–76.
- Michalowski, M., 1991. Foreign-born Canadian emigrants and their characteristics (1981–1986). *Int. Migrat. Rev.* 25 (1), 28–59.
- Nekby, L., 2006. The emigration of immigrants, return vs onward migration: evidence from Sweden. *J. Popul. Econ.* 19 (2), 197–226.

- Newey, W.K., 2009. Two-step series estimation of sample selection models. *Econometrica* J. 12 (s1), S217–S229.
- Organisation for Economic Cooperation and Development, 2008. *International Migration Outlook 2008*. OECD Publishing.
- Organisation for Economic Cooperation and Development, 2013. *International Migration Outlook 2013*. OECD Publishing.
- Picot, G., Piraino, P., 2012. Immigrant Earnings Growth: Selection Bias or Real Progress? Statistics Canada Analytical Studies Branch Research Paper Series 340.
- Pinger, P., 2010. Come back or stay? Spend here or there? Return and remittances: the case of Moldova. *Int. Migrat.* 48 (5), 142–173.
- Piore, M., 1979. *Birds of Passage: Migrant Labor and Industrial Societies*. Cambridge University Press.
- Pischke, J.-S., 1992. Assimilation and the Earnings of Guestworkers in Germany, ZEW Discussion Paper No. 92-17.
- Preston, I., 2013. The Effect of Immigration on Public Finances, CReAM DP 23/2013.
- Ramos, F., 1992. Out-migration and return migration of Puerto Ricans. In: *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*. University of Chicago Press, pp. 49–66.
- Reagan, P.B., Olsen, R.J., 2000. You can go home again: evidence from longitudinal data. *Demography* 37 (3), 339–350.
- Rendall, M.S., Ball, D.J., 2004. Immigration, emigration and the ageing of the overseas-born population in the United Kingdom. *Popul. Trends*, 18–27.
- Reyes, B.I., 1997. *Dynamics of Immigration: Return Migration to Western Mexico*. Public Policy Institute of California.
- Reyes, B.I., 2001. Immigrant trip duration: the case of immigrants from western Mexico. *Int. Migrat. Rev.* 35 (4), 1185–1204.
- Reyes, B.I., 2004. Changes in trip duration for Mexican immigrants to the United States. *Popul. Res. Pol. Rev.* 23 (3), 235–257.
- Reyes, B.I., Mameesh, L., 2002. Why does immigrant trip duration vary across US destinations? *Soc. Sci. Q.* 83 (2), 580–593.
- Rooth, D.-O., Saarela, J., 2007. Selection in migration and return migration: evidence from micro data. *Econ. Lett.* 94 (1), 90–95.
- Sander, M., 2007. Return Migration and the 'Healthy Immigrant Effect', SOEP Papers on Multidisciplinary Panel Data Research No. 60.
- Sarvimäki, M., 2011. Assimilation to a welfare state: labor market performance and use of social benefits by immigrants to Finland. *Scand. J. Econ.* 113 (3), 665–688.
- Shorland, P., 2006. *People on the Move: A Study of Migrant Movement Patterns To and From New Zealand*. New Zealand Department of Labour.
- Skuterud, M., Su, M., 2009. Immigrant Wage Assimilation and the Return to Foreign and Host-Country Sources of Human Capital, CLSRN Working Paper No. 30.
- Van Hook, J., Zhang, W., 2011. Who stays? Who goes? Selective emigration among the foreign-born. *Popul. Res. Pol. Rev.* 30 (1), 1–24.
- Van Hook, J., Zhang, W., Bean, F.D., Passel, J.S., 2006. Foreign-born emigration: a new approach and estimates based on matched CPS files. *Demography* 43 (2), 361–382.
- Venturini, A., Villosio, C., 2008. Labour-market assimilation of foreign workers in Italy. *Oxf. Rev. Econ. Pol.* 24 (3), 517–541.
- Waldorf, B., 1995. Determinants of international return migration intentions. *Prof. Geogr.* 47 (2), 125–136.
- Willis, R.J., 1986. Wage determinants: a survey and reinterpretation of human capital earnings functions. In: *Handbook of Labor Economics*. vol. 1, pp. 525–602.
- Wooldridge, J.M., 1995. Selection corrections for panel data models under conditional mean independence assumptions. *J. Econometrics* 68 (1), 115–132.
- Wooldridge, J.M., 2010. *Econometric Analysis of Cross Section and Panel Data*, second ed. MIT Press.
- Yahirun, J., 2009. Take Me 'Home': Determinants of Return Migration Among Germany's Elderly Immigrants, California Center for Population Research On-Line Working Paper Series 2009-019.
- Zakharenko, R., 2008. Return Migration: An Empirical Investigation, MPRA Paper No. 13755.

This page intentionally left blank