ECONSTOR
Make Your Publications Visible.

A Service of

ZBW Leibniz-Informationszentrum Wirtschaft
Leibniz Information Centre for Economics

Dustmann, Christian

Working Paper

Selective Outmigration and the Estimation of Immigrants Earnings Profiles

CReAM Discussion Paper Series, No. 02/14

Provided in Cooperation with:
Rockwool Foundation Berlin (RF Berlin)

Suggested Citation: Dustmann, Christian (2014) : Selective Outmigration and the Estimation of Immigrants Earnings Profiles, CReAM Discussion Paper Series, No. 02/14, Centre for Research & Analysis of Migration (CReAM), Department of Economics, University College London, London

This Version is available at:
https://hdl.handle.net/10419/295468

Standard-Nutzungsbedingungen:

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

Terms of use:

Documents in EconStor may be saved and copied for your personal and scholarly purposes.

You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.

If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.

WWW.ECONSTOR.EU

Mitglied der Leibniz-Gemeinschaft

Centre for Research and Analysis of Migration

CReAM

Discussion Paper Series

CDP No 02/14

Selective Outmigration and the Estimation of Immigrants' Earnings Profiles

Christian Dustmann and Joseph-Simon Görlach

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

Selective Outmigration and the Estimation of Immigrants' Earnings Profiles[1]

Christian Dustmann[2] and Joseph-Simon Görlach[3]

January 2014

Abstract - This chapter begins by documenting that temporary migrations are not only very common, but that outmigration of immigrants is selective both in terms of migrants' individual characteristics and their economic outcomes. We then examine the problems that arise when estimating immigrants' earnings profiles when outmigration is selective, and discuss the identifying assumptions needed to answer three different questions on immigrants' earnings careers. We show how better data can help to relax these assumptions, suggest appropriate estimators, and provide an illustration using simulated data. We finally provide an overview of existing papers that use different types of data to address selective outmigration when estimating immigrants' earnings profiles.

Keywords: migrant selection, outmigration, return migration, earnings profiles

[1] Chapter prepared for Chiswick and Miller (eds.), *Handbook on the Economics of Migration*. 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.

[2] CReAM and Department of Economics, University College London, Drayton House, 30 Gordon Street, London WC1H 0AX, UK; e-mail: c.dustmann@ucl.ac.uk

[3] CReAM and Department of Economics, University College London, Drayton House, 30 Gordon Street, London WC1H 0AX, UK; e-mail: joseph-simon.gorlach.09@ucl.ac.uk

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, Boustan and Eriksson, 2012a). 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 outmigration may lead to endogenous selection of the resident immigrant population, so that if earnings profile estimations ignore the possibly selective character of outmigration, 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 outmigration 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 outmigration that is selective.[4]

[FIGURE 1 HERE]

We can illustrate the problem as follows: supposing 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* 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 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 $(\tilde{\mu}_2^c - \tilde{\mu}_1^c)/(t_2 - t_1)$, where $\tilde{\mu}_t^c$ 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 outmigration 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 outmigration is random. In the case of non-randomness – for example, only the least successful leaving the country – the distribution of immigrants

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

residing in the host country would be truncated from below as in the figure. Hence, a simple OLS estimator that ignores this selective outmigration 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 outmigration (Section 3). We also outline the various ways in which to address these issues. We finally provide an example 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 Outmigration

As shown in Figure 2 for a number of major OECD countries, immigration over the last decade has been accompanied by very sizeable outmigration.[5] 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 into the U.S. in the early 1900s subsequently emigrated back to their countries of origin, a figure that actually may have been over twice as large (Bandiera, Rasul, and Viarengo, 2013). Such temporariness continues today: during the 2000–2010 period, almost

[5] See OECD (2013) for the country specific variable definitions.

2.1 million foreign-born residents out-migrated from the U.S. (Bhaskar, Arenas-Germosén, and Dick, 2013), a pattern also characteristic of many other countries. A recent report by the OECD (2008), for instance, estimates outmigration rates after 5 years of residence of 60.4% for immigrants entering Ireland in 1993–1998, 50.4% for immigrants entering Belgium in 1993–1999, 39.9% for immigrants entering the UK in 1992–1998; 39.6% for immigrants entering Norway in 1996–1999, and 28.2% for immigrants entering the Netherlands in 1994–1998.

[FIGURE 2 HERE]

Assessing outmigrations, 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 outmigration by country of origin

The most important question in estimating immigrant career profiles in destination countries is not the pure scale of outmigration but whether it is in any way selective and who the outmigrants are. A study by Dustmann and Weiss (2007), for example, uses British 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 estimate the extent of outmigration 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 outmigration.

Figure 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 5 years, which, if interpreted as emigration, indicates that among those who stayed for at least 1 year, only about 60% of the male and 68% of the female foreign-born immigrants remained in Britain 5 years later.

[FIGURE 3 HERE]

Figure 4 (see also Dustmann and Weiss, 2007, Figure 3) pools male and female immigrants but distinguishes between origin (left panel) and ethnicity (right panel).[6] Overall, it shows a large variation in the emigration propensities across immigrants from different origin countries: the outmigration 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 1 year left within 5 years of arrival), while outmigration for individuals from Asia (mainly India and Pakistan) and Africa is far less pronounced. In fact, there seems to be little indication of any emigration for immigrants from Africa and the Indian subcontinent. Rather, instead of being equally distributed across origin countries, outmigrants seem to come predominantly from English-speaking countries that are economically similar to the UK. Rendall and Ball (2004) report a very similar picture: whereas about 65% of immigrants from Canada and the U.S. emigrate within 5 years, only about 15% of immigrants from the Indian subcontinent do so.

[FIGURE 4 HERE]

[6] Fractions larger than one in the left panel are due to sampling error.

Borjas and Bratsberg (1996), using U.S. data, find similar large variation in outmigration rates across origin countries, with the lowest outmigration rates reported for immigrants from Asia (cf. Dustmann and Weiss , 2007). Specifically, they estimate that only 3.5% of Asian immigrants who arrived in the U.S. 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, find that Europeans have the highest propensity to leave the U.S. and immigrants from Asia the lowest, with western hemisphere immigrants taking an intermediate place. In an earlier paper analyzing outmigration 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 percent, with emigration rates for Koreans and Chinese not exceeding 22 percent (Jasso and Rosenzweig, 1982). Canadian emigration, in contrast, was between 51 and 55 percent, and emigration rates for legal immigrants from Central America, the Caribbean (excluding Cuba), and South America were at least as high as 50 percent and possibly as high as 70 percent.

Patterns reported for Canada are not dissimilar. Beaujot and Rappak (1989) report that outmigration for those arriving in the 1951–1970 period was highest for individuals born in the U.S. (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) report that only 3–6 percent of immigrants from non-English-speaking countries are likely to leave as compared with 20–30 percent of immigrants from English-speaking countries (see also Lukomskyj and Richards, 1986).

Research for Scandinavian countries, on the other hand, 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, Raaum, and Sørlie (2007), using data from the Norwegian population register, estimate that as much as 84% of immigrants from the U.S. leave compared with only 9% from Vietnam. They interpret this finding to mean stark differences in outmigration behavior based on the home country's economic development and distance from Norway. In addition, their administrative data include information on where foreign-born emigrants travel subsequently. They report 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 outmigrants from neighboring Nordic countries returned home. Like Bratsberg et al. (2007), based on 1991–2000 emigrant data from *Statistics Sweden*, Nekby (2006) finds 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 outmigrations observed are not return migrations but migrations that continue to other destination countries. Overall, she finds that outmigration probabilities are highest for immigrants from North America and from Western European countries of origin.

The large outmigration propensity of Nordic migrants in comparison to other groups is confirmed by Jensen and Pedersen (2007) for Denmark, who find that although 80% of Turkish immigrants remained in Denmark for 10 years or more, only 20% of the Nordic immigrants did so. Edin, LaLonde, and Åslund (2000) show similar patterns for Sweden: about 44% of Nordic immigrants had left that country within 5 years of their arrival, a number that is significantly lower for immigrants from non-OECD countries. Klinthäll (2003) reports 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 5, which combines the estimates – drawn from a large number of empirical studies – on the fractions of post-war 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 arrival against the number of years since immigration.[7] 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 outmigrated immigrants on the years since immigration (ysm), we find that the outmigration 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 1). The estimated coefficients and the differences across immigrant groups are illustrated by the fitted lines in Figure 5.

[FIGURE 5 HERE]

[7] For Figure 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 outmigration. 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, Schluter, and Wahba (2013), Bohning (1984), Borjas and Bratsberg (1996), Bratsberg, Raaum, and Sorlie (2007), Edin, LaLonde, and Åslund (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.

	Fraction that has emigrated by region of origin	
	Africa/Asia	Americas/Europe/Oceania
ysm	0.019	0.037
cons	0.075	0.129
N	52	97

Table 1: OLS-coefficients of time since immigration for foreign-born emigration by origin region

2.2 Selection on education

Although a number of papers have examined the relation between educational attainment and outmigration, there is no clear pattern across the literature.[8] For returnees from the U.S. among the Puerto Rican immigrant population, Ramos (1992) reports a positive selection on education, and Zakharenko (2008), working with CPS data, estimates the probability of emigration for immigrants to the U.S. from any destination to be lower for highly educated immigrants. He also shows, 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, reports 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 outmigrate. For example, Jensen and Pedersen (2007) find a positive relation between outmigration probabilities and educational attainment among immigrants in Denmark, while Dustmann (1996) finds that intended migration durations are longer among less educated immigrants in Germany. Nevertheless, he also reports that the probability that immigration is intended to be permanent increases with years of schooling. Also for Germany, Constant and Zimmerman (2011), using information

[8] 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.

on multiple migration spells, show that repeat migration is more likely among male and less educated individuals, whereas Beenstock (1996), using data for Israel, shows that the stays of more highly educated immigrants who arrived in the 1970s are more likely to be temporary. For Italy, Coniglio, De Arcangelis, and Serlenga (2009) confirm that schooling increases the probability of outmigration even among undocumented immigrants. However, Carrion-Flores (2006), using data from the Mexican Migration Project 1982–1999, reports the opposite for Mexican immigrants in the U.S.: she finds a positive effect of high educational attainment on the likelihood of returning to Mexico. Maré, Morten, and Stillman (2007), on the other hand, find for New Zealand that outmigration 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 outmigrants 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 hypothesize 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 is confirmed for linked Finish and Swedish data by Rooth and Saarela (2007), who find significant selection on educational attainment but no evidence of selection on earnings conditional on education. From a sending country's perspective, Pinger (2010) reports 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 outmigration with respect to educational attainment is context dependent.

2.3 Selection on earnings

A number of studies investigate the relation between outmigration and immigrant earnings,[9] an association that is far from straightforward. Dustmann (2003), for instance, points out that changes in earnings may impact 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 outmigration of U.S. immigrants find 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, identifies a negative effect of wages on the probability that immigrants to the U.S. will outmigrate, and Abramitzky, Boustan, and Eriksson (2012a) reach a similar conclusion for U.S. immigrants even in the so-called age of mass migration. Using U.S. census data from 1900, 1910, and 1920, together with the Integrated Public Use Microdata Series (IPUMS) for 1900, they construct a panel of U.S. natives and immigrants from a number of major sending countries who arrived between 1880 and 1900, and find that outmigrants were negatively selected on earnings. In Abramitzky et al. (2012b), on the other hand, the same authors note 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) assume that in the absence of selective outmigration, period effects in earnings regressions for Israeli- and native-born workers in the U.S. should be the same. Using a sample of such individuals from pooled 1980 and 1990 census data, they perform separate earnings regressions for the two groups. If outmigration

[9] 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. (2013) for their results on the effect of unemployment spells on the probability of outmigration.

were random, once years spent in the U.S. 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 U.S.-born workers, a finding that the authors interpret as evidence that Israelis who return from the U.S. are negatively selected among all Israeli immigrants. For the U.S., Reagan and Olsen (2000) also find 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) identifies a negative relation between wages and the probability of return migration by Mexican immigrants in the U.S., but only during the first year of residence: this effect turns positive for immigrants who have remained in the U.S. 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 outmigration 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, finds 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 finds 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 establish a negative association, conditional on the source country, between immigrants' incomes and the likelihood of leaving Sweden within 5 or 10 years after arrival.

Also for Sweden, Nekby (2006), by allowing outmigration rates to change nonlinearly along the earnings distribution, shows 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 shows 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 (2013) also identify a U-shaped relation between incomes and outmigration 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 compute outmigration probabilities for different income groups. They find that such probabilities are U-shaped with respect to income, with the highest probability of leaving among migrants in the lowest income group. They interpret 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, Chiswick, and Paltiel (2010) find no association whatsoever between immigrant earnings in 1983 and the individual still residing in the country in 1995.

2.4 Other characteristics and outmigration

Several studies investigate the relation between outmigration and individual characteristics other than earnings and education, including the possibility that outmigration probabilities increase around retirement age. Waldorf (1995), for instance, reports that among immigrants to Germany, the intention to return within the next 4 years increases close to retirement age,

while Cobb-Clark and Stilman (2008) find among immigrants to Australia that actual outmigration rates are highest at retirement age. For the U.S., 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, Zhang, Bean, and Passel (2006) reveal that outmigration rates are higher for younger immigrants. A positive effect of age on the probability of remaining in the host country is also reported by Bijwaard (2010) for immigrants to the Netherlands, although Edin et al. (2000) find 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) find 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 provide evidence for Denmark that the relation between age at immigration and outmigration differs by immigrant country of origin. That is, like Aydemir and Robinson (2008), they find that for immigrants from industrialized countries, age at immigration is negatively correlated with the probability of outmigration. They also show, 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 first, Sander (2007), in an analysis of the effect of health shocks on emigration decisions among immigrants in Germany, reveals that negative health shocks do indeed affect the probability of outmigration, although the effect appears to depend on whether these shocks are transitory or permanent. That is, although the likelihood of outmigration 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) reports lower conditional return probabilities for Mexican immigrants in the U.S. who are from economically more active regions in Mexico, an outcome he explains by the higher value that savings accumulated in the U.S. have for these migrants on their return. Reyes and Mameesh (2002) also find that economically more active and urban regions in the U.S. attract more permanent migrants.

Naturally, outmigration rates vary by immigrant status: refugees are considerably less likely to leave the host country than economic migrants, and part of the variations in outmigration across immigrants from different countries of origin can be explained by this (e.g., Aydemir and Robinson, 2008; Duleep, 1994; Lundh and Ohlsson, 1994; Klinthäll, 2007). The way in which emigration rates vary among different immigrant groups in the Netherlands is the subject of a study by Bijwaard (2010), who demonstrates that emigration hazards are higher for labor migrants than for individuals who immigrated for family reasons. Outmigration 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 U.S. 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 outmigration 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 outmigration is 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 outmigrate from and remigrate back to the host country in a way that leads to selective outmigration 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 outmigration. In doing so, we assume that immigrants' decisions on human capital investment, labor supply, and job choice are not determined by their migration and outmigration 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.[10] 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 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

[10] 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.

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) outmigration.

Answering Q1 based on repeated cross-sectional data, however, reveals little about whether outmigration 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, outmigration 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 outmigration. 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 outmigration?*

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 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 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 outmigration is exogenous to the process that affects immigrant earnings; in other words, outmigration 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:

(1) $w_{it}^c = \mu_t^c + \varepsilon_{it}^c$,

where w_{it}^c are 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 these 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, $\mu_{it}^c = \overline{\mu_t^c} + x_{it}'b$.

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

$s_{it}^c = \prod_{c<\tau\leq t} \mathbf{1}[z_{i\tau}^c{}'\beta + u_{i\tau}^c > 0]$,

where **1**[A] is an indicator function that takes the value 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 $E(\varepsilon_{it}^c | s_{ic}^c = 1) = 0$ for all *t*. This latter means that the expectation of the error 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 outmigration, 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 outmigration requires considerably more model restrictions than generally acknowledged in the literature on the earnings assimilation of immigrants.

The assumption that outmigration 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 outmigration and remigration patterns (e.g., repeat migration) would require us to model not only the outmigration 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 data sets, 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

$$(2)\ E\left(w_{it_2}^c \middle| \mu_{t_2}^c, s_{it_2}^c = 1\right) - E\left(w_{it_1}^c \middle|, \mu_{t_1}^c s_{it_1}^c = 1\right) = [\mu_{t_2}^c - \mu_{t_1}^c] + [E\left(\varepsilon_{it_2}^c \middle| \mu_{t_2}^c, s_{it_2}^c = 1\right) - E\left(\varepsilon_{it_1}^c \middle| \mu_{t_1}^c, s_{it_1}^c = 1\right)]$$

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 outmigrated. 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 outmigration. When outmigration is "exogenous" and the last two last terms in (2) equal zero, a simple OLS estimation of equation (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 ($E * O$) but not *between* them, then conditioning on country of origin and education in a flexible way will eliminate the selection bias, as now $E(\varepsilon|u, z, E * O) = 0$. Thus, conditional on $E * O$, outmigration is independent of ε (see Cameron and Trivedi, 2005, pp 863ff, or Wooldridge, 2010, pp 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 $var(u_{it}^c) = 1$, and that $E(e_{it}^c|u_{ic}^c, \dots, u_{it}^c) = \sigma_{eu}(t)u_{it}^c$ – 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 $E(\varepsilon_{it}^c|\mu_t^c, s_{it}^c = 1)$: it follows from our assumptions that

(3) $E(\varepsilon_{it}^c|\mu_t^c, s_{it}^c = 1) = E(\varepsilon_{it}^c|s_{it}^c = 1) = E(e_{it}^c|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it}^c > -z_{it}^c{}'\beta) +$

$E(\alpha_i|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it}^c > -z_{it}^c{}'\beta) = \sigma_{eu}(t)E(u_{it}^c|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it}^c > -z_{it}^c{}'\beta) +$

$E(\alpha_i|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it}^c > -z_{it}^c{}'\beta),$

where $\sigma_{eu}(t)$ is the covariance between u_{it}^c and e_{it}^c.[11] Thus, the bias we obtain in estimating $\Delta\mu_{t_1t_2}^c$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 side reflects selective outmigration 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 this latter corresponds to selection being systematically related to the immigrants' unobserved and time invariant productivity, the former reflects selective outmigration 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 $\Delta\mu_{t_1t_2}^c$ without assuming that both α_i and e_{it}^c 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 $\Delta\mu_{t_1t_2}^c$ 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

[11] We maintain the assumption of a unit variance for the residual in the selection equation throughout.

administrative data sources. In these data sets, 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 c 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 data sets do not typically provide longitudinal information on the entire entry cohort; rather, the migrant sample is determined by 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 data sets of that format used in the literature are longitudinal. For example, Lubotsky (2007)[12] links information on immigrants in the CPS sample who are still residing in the U.S. in 1994 to social security earnings records, thereby providing longitudinal information on these migrants back to 1951. By construction, this data set 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 data sets over (repeated) cross- sectional data is that it provides 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

[12] 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.

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 outmigration 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

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

Note the difference between equations (2) and (4). Whereas in (2), the selection indicator refers to the particular year in which the cross section was collected, in (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 data set is longitudinal, it will be a balanced panel of immigrants. No matter the nature of the selection, the expression in (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 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 (3) and assume that $E(e_{it}^c|u_{ic}^c, \dots, u_{iT}^c) = \sigma_{eu}(t)u_{it}^c$, to rewrite $E(\varepsilon_{it}^c|\mu_t^c, s_T^c = 1)$ as

(3′) $E(\varepsilon_{it}^c|\mu_t^c, s_T^c = 1) =$

$E(e_{it}^c|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) + E(\alpha_i|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) =$

$\sigma_{eu}(t)E(u_{it}^c|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta) + E(\alpha_i|u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta),$

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 (1) on the pooled t_1 to T cross-sections with years since migration indicators included would yield the following estimation equation:

(1′) $w_{it}^c = \mu_{t_1}^c + \Delta\mu_{t_1(t_1+1)}^c D[t = t_1 + 1] + \cdots + \Delta\mu_{t_1T}^c D[t = T] + \varepsilon_{it}^c,$

where $D[.]$ is an indicator function, equal to one if its argument is true and zero otherwise. An OLS regression can then be run on (1') to identify wage growth $\Delta\mu_{t_1t_2}^c$ (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 (3') is unequal to zero for $corr(\alpha_i, u_{it}^c) \neq 0$, 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 outmigrants is related not only to the unobservable time constant variables α_i but also to contemporaneous time variant unobservables in the earnings equation,[13] 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 outmigration. In that case, an OLS estimator using stock based data will not

[13] Restricting the correlation in time variant unobservables to contemporaneous realizations between e_{it}^c and u_{it}^c simplifies the exposition, but this correlation could be generalized.

identify $\Delta\mu^c_{t_1t_2}$ (neither will a difference (or fixed effects) estimator in case the data set is longitudinal, see below). To illustrate, we consider the conditional expectation of earnings growth between periods t_1 and t_2:

$$
\begin{aligned}
&E\left(w^c_{it_2} - w^c_{it_1} \middle| \Delta\mu^c_{t_1t_2}, \mathrm{s}^c_T = 1\right) \\
&\qquad = \Delta\mu^c_{t_1t_2} \\
&\qquad + \left[\sigma_{eu}(t_2)E\left(u^c_{it_2} \middle| u^c_{ic} > -z^{c\prime}_{ic}\beta, \dots, u^c_{iT} > -z^{c\prime}_{iT}\beta\right)\right. \\
&\qquad \left. - \sigma_{eu}(t_1)E\left(u^c_{it_1} \middle| u^c_{ic} > -z^{c\prime}_{ic}\beta, \dots, u^c_{iT} > -z^{c\prime}_{iT}\beta\right)\right].
\end{aligned}
$$

The bias term in brackets will only vanish if $\Delta\sigma_{eu}(t) = \sigma_{eu}(t_2) - \sigma_{eu}(t_1) = 0$ and $E\left(u^c_{it_2} - u^c_{it_1} \middle| u^c_{ic} > -z^{c\prime}_{ic}\beta, \dots, u^c_{iT} > -z^{c\prime}_{iT}\beta\right) = 0$, meaning that both the covariance σ_{eu} and (whenever $\sigma_{eu} \neq 0$) the selection threshold $-z^{c\prime}_{it}\beta$ are constant over time. In 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^c_{it}. Hence, if selective outmigration 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 $u^c_{ic} > -z^{c\prime}_{ic}\beta, \dots, u^c_{iT} > -z^{c\prime}_{iT}\beta$ 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 outmigration by comparing estimates from stock based data with those for repeated cross-sectional data for a particular entry cohort, as is done for $\Delta\mu^c_{t_1t_2}$ 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 (3), the bias can be decomposed into two parts: The first part derives from selection on time constant unobservables, which is equal to

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

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 outmigrants 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 outmigrants 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

$$\begin{aligned}\sigma_{eu}(t_2) E\left(u_{it_2}^c \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it_2}^c > -z_{it_2}^c{}'\beta\right) \\ - \sigma_{eu}(t_1) E\left(u_{it_1}^c \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it_1}^c > -z_{it_1}^c{}'\beta\right),\end{aligned}$$

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

$$\begin{aligned}(4')\ & E\left(\alpha_i \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it_2}^c > -z_{it_2}^c{}'\beta\right) - E\left(\alpha_i \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it_1}^c > -z_{it_1}^c{}'\beta\right) \\ & + \sigma_{eu}(t_2) E\left(u_{it_2}^c \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it_2}^c > -z_{it_2}^c{}'\beta\right) \\ & - \sigma_{eu}(t_1) E\left(u_{it_1}^c \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{it_1}^c > -z_{it_1}^c{}'\beta\right) \\ & - \left[\sigma_{eu}(t_2) E\left(u_{it_2}^c \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta\right)\right. \\ & \left. - \sigma_{eu}(t_1) E\left(u_{it_1}^c \middle| u_{ic}^c > -z_{ic}^c{}'\beta, \dots, u_{iT}^c > -z_{iT}^c{}'\beta\right)\right],\end{aligned}$$

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

outmigrate if they experience a positive earnings shock ($\sigma_{eu} < 0$), and that older individuals are more likely to leave ($-z_{it}^{c\prime}\beta$ 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 ($corr(\alpha_i, u_{it}^c) > 0$), 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 data sets, 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 data set 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 provides information on who left the country between $t-1$ and t.[14] Furthermore, if immigrants are observed from the time of arrival, the mean earnings level μ_c^c 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

[14] 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.

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 $\Delta\mu^c_{t_1t_2}$. 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 such is not the case, we can construct such a data set from longitudinal data by discarding all individuals who dropped out of the sample for some periods. As with outmigration, 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 outmigration does not depend on time variant observables that refer to period t_2 and change unpredictably.[15]

To focus on selection correlated with the time variant unobservables e^c_{it}, consider the earnings equation in (1) written in differences,

$(1'')\ \Delta w^c_{it} = \Delta\mu^c_{t_1t_2} + \Delta e^c_{it}.$

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

(5) $s^c_{it} = \mathbf{1}[z^{c\prime}_{it} b_t + v^c_{it} > 0]$.

[15] 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).

Equation (5) can thus be seen as a reduced form selection equation in which all explanatory variables from (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 outmigration, 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 $E\left(\Delta e_{it}^c \middle| \Delta\mu_{t,t-1}^c, z_{it}^c, v_{it}^c, s_{it}^c = 1\right) = r_{\Delta e_t v_t} v_{it}^c$, where $r_{\Delta e_t v_t}$ is the covariance between Δe_t and v_t, the expectation of wage growth in (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

$$E\left(\Delta w_{it}^c \middle| \Delta\mu_{t,t-1}^c, z_{it}^c, s_{it}^c = 1\right) = \Delta\mu_{t,t-1}^c + E\left(\Delta e_{it}^c \middle| \Delta\mu_{t,t-1}^c, z_{it}^c, s_{it}^c = 1\right) = \Delta\mu_{t,t-1}^c + r_{\Delta e_t v_t} \lambda(z_{it}^{c\prime} b_t),$$

where the last equality follows from the normality assumption and $\lambda(z_{it}^{c\prime} b_t) = \phi(z_{it}^{c\prime} b_t)/\Phi(z_{it}^{c\prime} b_t)$ 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, pp 837f) suggests to estimate 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 $D[t = c+1]\, \hat{\lambda}_{it}, \ldots, D[t = T]\, \hat{\lambda}_{it}$, where the $D[t = \tau]$ are dummy variables, with $D[t = \tau] = 1$ if $t = \tau$. Doing so yields the following equation, which can be estimated using pooled OLS:

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

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 outmigration 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. This latter can then 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 2 summarizes the discussion of identification of immigrants' earnings profiles in terms of the simple model above.

	Selection on time constant unobservables	**Selection on both time constant and time varying unobservables**
Unrestricted repeated cross sections:		
$\boldsymbol{\mu_c}$	not identified	not identified
$\boldsymbol{\Delta\mu_{t_1t_2}}$	not identified	not identified
Stock based samples:		
$\boldsymbol{\mu_c}$	not identified	not identified
$\boldsymbol{\Delta\mu_{t_1t_2}}$	identified	not identified
Longitudinal data:		
$\boldsymbol{\mu_c}$	identified if panel starts at $t = c$	identified if panel starts at $t = c$
$\boldsymbol{\Delta\mu_{t_1t_2}}$	identified	identified

Table 2: Identification of immigrant earnings profiles under selective outmigration

It should be noted that in order to focus on the key issues related to selective outmigration, 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.[16] Otherwise, their predetermination would lead to an endogenous regressor bias in the difference 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.

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; however, there is increasing evidence that this assumption may be violated (see, e.g., Borjas and Bratsberg, 1996; Barth, Bratsberg, and Raaum, 2004; Dustmann, Glitz, and Vogel, 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.[17] 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 outmigration given different assumptions about the correlation between unobservables in the earnings and selection equations, when

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

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

these are not sufficiently accounted for. We us 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

$$(6)\ w_{it} = \mu_t + \alpha_i + e_{it} \equiv \mu_o + \gamma ysm_{it} + \alpha_i + e_{it},$$

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 side of the equation. Supposing also that the selection rule for an immigrant remaining in the host country is given by

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

then outmigration is an absorbing state (i.e. once outmigrated, an individual will not reappear in the data set at a later point in time).[18]

Assuming first that u_{it} is correlated only with α_i and not with e_{it}[19] (i.e., selection is on individual-specific time constant unobservables in the earnings equation), then data generated in our simulation is as shown in Figure 6, where the black dots represent the observed log earnings of immigrants still residing in the host country and the grey dots the immigrants from the original arrival cohort who outmigrated. Given our assumptions about the nature of

[18] 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 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 outmigrate.

[19] We assume that $corr(\alpha, u) = 0.7$.

selection (with selection into staying being positively correlated with unobserved productivity α_i), outmigration is negatively selective.

[FIGURE 6 HERE]

Table 3 lists the results of using the OLS estimator for equation (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 (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.[20] 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.

	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**
cons	2	2.277921	2.341263	2.438709	2.450425	-
		(0.0010001)	(0.0028334)	(0.0016322)	(0.0012547)	
ysm	0.02	0.0278496	0.021146	0.0199223	0.0200594	0.0199664
		(0.0000591)	(0.0008531)	(0.0001797)	(0.0000845)	(0.0000534)

Table 3: Selection on time-constant unobservables only

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 higher Ts).

[20] 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.

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 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}[21] 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

$$(7)\ E(w_{it}|s_{iT}=1) = \mu_t + E(\alpha_i|s_{iT}=1) + E(e_{it}|s_{iT}=1)$$
$$= \mu_t + E(\alpha_i|u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta)$$
$$+ E(e_{it}|u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta),$$

and given our normality assumption, OLS identifies

$$\Delta\mu_{t_1t_2} + \sigma_{eu}(t_2)\, E\left(u_{it_2}\middle|u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta\right) -$$
$$\sigma_{eu}(t_1)\, E\left(u_{it_1}\middle|u_{ic} > -z_{ic}'\beta, \dots, u_{iT} > -z_{iT}'\beta\right).$$

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

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

probability) or the correlation between outmigration 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 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 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 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.

	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**
cons	2	2.352842	2.436231	2.485823	2.48313	-	-
		(0.0008919)	(0.0024729)	(0.0015218)	(0.0011942)		
ysm	0.02	0.0242654	0.0086049	0.016204	0.0182887	0.0184484	0.0209794
		(0.000055)	(0.0007664)	(0.0001715)	(0.0000818)	(0.0000528)	(0.001225)

Table 4: Selection on time-constant and time-varying unobservables

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 3), the bias consists now of two opposing parts: first, the same upward bias resulting from selection on time constant unobservables, as in Table 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 4 than in Table 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 2 and 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 (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).

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 outmigration 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 outmigration is an absorbing state means that $s_{it} = 1$ always implies $s_{it-1} = 1$, to identify the slope parameter $\Delta\mu^{c}_{t_1 t_2}$, 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 from 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 purpose, 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:

$$(8)\ \Delta w_{it} = \gamma \Delta ysm_{it} + r_t \frac{\phi(\hat{b}_0 + \hat{b}_1 w_{it-2})}{\Phi(\hat{b}_0 + \hat{b}_1 w_{it-2})} + \Delta e_{it}.$$

By estimating this model, we obtain the results given in the last column of Table 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 (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.[22]

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 outmigration driven by skill endowment and accumulation, we extend the work of Borjas and Bratsberg (1996) to produce a simple model of temporary migration.[23] In this model, log earnings in the origin country *o* and destination country *d* take the following form:[24]

$$(9)\ w_l = \mu_l + \varepsilon_l,$$

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

[22] 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.

[23] See Dustmann, Fadlon, and Weiss (2011) for a generalization of this model to multiple skills and a dynamic setting and Dustmann and Glitz (2012) for a simplified version.

[24] 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 consider a lower than expected return in the host country as a reason for return migration.

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 $corr(\varepsilon_d, \varepsilon_o) = 1$. 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

$$(10)\ w_{do} = \tau(\mu_d + \varepsilon_d) + (1 - \tau)(\mu_o + \varepsilon_o + \tau k),$$

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 (9) and (10) into the conditions above.

In Figure 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 re-migrants 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.

[FIGURE 7 HERE]

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 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 nonmigrants 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

$Cov(\varepsilon_d - \varepsilon_0, \varepsilon_0) = (\pi - 1)$. Consider now the mean earnings of those who have emigrated to the destination country and who are observed there in period *t*:

$$E(w_d|\tau > t) = \mu_d + \pi E(\varepsilon_o|\tau > t) =$$

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

where $q(t) = \frac{k(2t-1)-[\mu_d-\mu_o]}{\pi-1}$, 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 in the home country who migrates to the destination country, μ_d, and a term that reflects the selection of migrants from the overall population:

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

Hence, the mean entry wage of the original arrival cohort (denoted as μ_c^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 nonmigrants 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 outmigration truncates the skill distribution. This case is the one shown in Figure 1, where selective outmigration 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., outmigration is negatively selective), the wage growth between periods t_1 and $t_2 > t_1$ obtained from repeated cross-sectional data is

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

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 reexamine these same issues using stock sampled data, then

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

where $w_d(\tau)$ are 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 Outmigration 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 U.S. 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 U.S. 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 U.S. and differences in immigrants' initial earnings positions after arrival. Moreover, whereas the early literature implicitly assumes that these cohort differences equal zero, Borjas (1985) stresses 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 address this problem by employing repeated cross-sections, which allow to address this problem, under some assumptions.

As we have already demonstrated, similar issues arise when immigrant outmigration is selective. Even with repeated cross-sectional data, assimilation estimates can still be biased when selective outmigration drives what seem to be changes in cohort quality (see, e.g., Chiswick, 1986). Several early studies take advantage of longitudinal data to address this problem. Borjas (1989), for example, uses longitudinal information from the 1972–1978

Survey of Natural and Social Scientists and Engineers (which is based on immigrant listings in the 1970 U.S. census) to analyze the earnings paths of a particular immigrant group. He finds that among foreign-born scientists and engineers in the sample, attrition – which is assumed to be largely driven by emigration from the U.S. – is more likely for individuals with less favorable economic outcomes. Because this data set includes earnings information dating back to 1969, it allows him to perform separate estimations for the 1969–1971 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 is lower for immigrants who leave the sample later. In subsequent work, Pischke (1992) addresses 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) estimate log wage regressions using different samples of Mexicans residing in the U.S. and Mexicans who returned to their native country. Based on their comparison of estimates from single versus repeated cross-sectional data, however, they conclude 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 U.S. Hu (2000), for example, compares 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 outmigration. This interpretation, however, relies on the assumption that the selection of outmigrants either occurs on time constant unobservables only, or that selection does not vary systematically over time (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 1950–1959 and 1965–1969 immigrant cohorts catching up with U.S.-born workers within 10 years of arrival, the longitudinal data indicate a decline in earnings residuals by time spent in the U.S. for this population both in levels and relative to the U.S.-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 U.S.-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 (1931–1941), separating the effects of age at immigration or pre-migration experience and changes in average immigrant cohort earnings is difficult.

Lubotsky (2007) addresses this difficulty by using a broader sample, based on 1951–1997 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 compares the estimates of the time since immigration effects with those obtained from the 1970–1990 cross-sectional census data. Consistent with the hypothesis of negatively selective outmigration, the wage profiles obtained from the longitudinal data are flatter.[25] Like Hu, however, Lubotsky (2007) uses 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 estimates his earnings equation in levels and only includes cohort- rather than

[25] 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 U.S. 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.

individual-specific indicators. Consistent with the greater likelihood that it is low-earning immigrants who will leave, the stock based longitudinal data indicates a less pronounced deterioration in cohort earnings on entry to the U.S. In an earlier study that uses part of the same stock based longitudinal data as Lubotsky (2007), Duleep and Dowhan (2002) analyze immigrant assimilation in the U.S. 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 outmigrants 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. (2012a) construct a panel of U.S. residents from 1900-1920 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 outmigrants 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), report 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 U.S. 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 data sets 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 is taken in an early paper by Pischke (1992), who estimates 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) use 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 5 years. In line with the U.S. literature, they also find 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 demonstrate that if outmigration is not taken into account, earnings assimilation by OECD immigrants is overestimated by about 90 percent, a figure largely in line with Lubotsky (2007). Arai (2000), however, criticizes 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) analyzes immigrant assimilation in Finland using longitudinal data that follow immigrants from the time of their arrival. To address selective emigration, he compares OLS estimates of immigrant earnings growth based on the whole sample with estimates based only on immigrants that stay for at most 5 years. Although he does show that short-term migrants experience no earnings growth, he acknowledges 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 estimates 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 outmigrants 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 outmigrants 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 $\frac{d\sigma_{eu}}{dt} < 0$). 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, Bratsberg, and Raaum (2012), on the other hand, in their analysis of the role of native-immigrant differences in job mobility, find 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 5 years after arrival (cf. Sarvimäki , 2011). Nevertheless, their findings permit no conclusion that outmigration is random. Skuterud and

Su (2009), using a rotating panel of immigrants and natives in Canada, find little difference between OLS and fixed effect estimates of immigrant wage assimilation. They reconcile this observation with the contrary findings for stock based samples from the U.S. (e.g., Hu, 2000; Lubotsky, 2007) in two ways: First, they argue that in the Canadian case, emigrant selection may be less clear because many more able immigrants may move onward to the U.S. Second, they point out that if emigration is correlated with heterogeneity in wage growth rather than wage levels, then it is possible that $E(\alpha_i|u_{ic} > -z_{ic}'\beta, \ldots, u_{it} > -z_{it}'\beta) = 0$, 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, Hanel, and McVicar (2012) analyze immigrant earnings in Australia; however, their very brief discussion paper does 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) use the German Soco-Economic Panel (GSOEP). As a panel, this data set enables explicit modeling of outmigration. Specifically, these authors formulate 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 include family and country of origin characteristics as explanatory variables in the selection equation. They only find 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 assume that the covariance between unobservables in the earnings and selection equations is constant over time, although they do allow it to be immigrant cohort specific. More important, they include selection correction terms in the levels rather than the differenced earnings equation. This assumes that selective outmigration 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 outmigration is an absorbing state then such an estimating equation would not appropriately correct for selective outmigration because it ignores that 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, Strom, Venturini, and Villosio (2009), who use a similar framework for Italy.

Biavaschi (2013), working with U.S. data, challenges the view that emigration is negatively selective. She formulates a selection model and then, using data from the 2000 U.S. and Mexican censuses, semi-parametrically estimates the counterfactual density of the wage

residuals of Mexican immigrants in the U.S. had there been no outmigration. Biavaschi uses an identification at infinity argument to recover this counterfactual distribution if only one cross-section of data is available. She finds 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 U.S. would have higher wages. Assuming that outmigration is not correlated with unpredictably changing variables that affect wages, an alternative method of accounting for selective outmigration is proposed by Kim (2007). Specifically, he applies 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 data sets. Like Hu (2000), he finds 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 outmigration is taken into account.

The problem of selective outmigration is recognized well beyond studies of immigrant earnings assimilation. Kaushal's (2011) analysis of the returns to U.S. versus overseas education, for example, shows that, based on longitudinal data from the National Survey of College Graduates, U.S.-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 their study, however, whether foreign-born individuals acquire their education in the U.S. 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) reemphasizes the need to use longitudinal data when emigration is selective. Using a similar longitudinal stock based sample of immigrants to the U.S. as in Lubotsky (2007), he argues that not only estimates of wage growth will be inconsistent if outmigration 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 outmigration on empirical estimates have been conducted by Bratsberg, Raaum, and Røed (2010), who analyze immigrant employment in Norway, by Matos (2011) for immigrant career paths in Portugal, and Kaushal and Shang (2013) for wage assimilation among U.S. immigrants in different destination areas.

5 Conclusions

In this chapter, we address selective outmigration – 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 outmigration 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 outmigration, 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 outmigrates. 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 outmigrants. However, given more general assumptions, data are needed that allow modeling of the outmigration selection process; for example, longitudinal data that is 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 outmigration. We also demonstrate the impossibility of making any general statement about the direction of selective outmigration, unless longitudinal data is available that allows modeling outmigration, and that the direction and magnitude of selection may differ both across countries and across immigrant groups within the same country. Unfortunately, this 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 outmigration.

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 is available that allows assessment of selective outmigration, the key assumption that we made here is that the process that governs outmigration 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.

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." NBER Working Paper No. 18011 (2012a).

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "Europe's Tired, Poor, Huddled Masses: Self-Selection and Economic Outcomes in the Age of Mass Migration." *American Economic Review* 2012b, 102(5): 1832–1856.

Ahmed, Bashir, and J. Gregory Robinson. “Estimates of Emigration of the Foreign-Born Population: 1980–1990.” U.S. Census Bureau Population Division Working Paper No. 9 (1994).

Alders, Maarten, and Han Nicolaas. “One in Three Immigrants Leave within Six Years.” *Statistics Netherlands* online, 20 January 2003.

Arai, Mahmood. "Comments on Edin, LaLonde and Åslund: Emigration of Immigrants and Measures of Immigrant Assimilation: Evidence from Sweden." *Swedish Economic Policy Review* 7 (2000): 205-211.

Aydemir, Abdurrahman, and Chris Robinson. "Global labour markets, return, and onward migration." *Canadian Journal of Economics/Revue canadienne d'économique* 41, no. 4 (2008): 1285-1311.

Bandiera, Oriana, Imran Rasul, and Martina Viarengo. "The Making of Modern America: Migratory Flows in the Age of Mass Migration." *Journal of Development Economics* 102 (2013): 23-47.

Barth, Erling, Bernt Bratsberg, and Oddbjørn Raaum. "Identifying earnings assimilation of immigrants under changing macroeconomic conditions." *The Scandinavian Journal of Economics* 106, no. 1 (2004): 1-22.

Barth, Erling, Bernt Bratsberg, and Oddbjørn Raaum. "Immigrant wage profiles within and between establishments." *Labour Economics* 19, no. 4 (2012): 541-556.

Beaujot, Roderic, and J. Peter Rappak. "The Link between immigration and emigration in Canada, 1945-1986." *Canadian Studies in Population*, vol. 16, no. 2 (1989): 201-216.

Beenstock, Michael. "Failure to absorb: remigration by immigrants into Israel." *International Migration Review*, vol. 30, no. 4 (1996): 950-978.

Beenstock, Michael, Barry R. Chiswick, and Ari Paltiel. "Testing the immigrant assimilation hypothesis with longitudinal data." *Review of Economics of the Household* 8, no. 1 (2010): 7-27.

Beggs, John J., and Bruce Chapman. "Male immigrant wage and unemployment experience in Australia." In *Immigration, Trade and the Labor Market*, pp. 369-384. University of Chicago Press, 1991.

Bellemare, Charles. "A life-cycle model of outmigration and economic assimilation of immigrants in Germany." *European Economic Review* 51, no. 3 (2007): 553-576.

Bhagwati, Jagdish N., and Richard A. Brecher. "National welfare in an open economy in the presence of foreign-owned factors of production." *Journal of International Economics* 10, no. 1 (1980): 103-115.

Bhaskar, Renuka, Belkinés Arenas-Germosén, and Christopher Dick. "Demographic Analysis 2010: Sensitivity Analysis of the Foreign-Born Migration Component." (2013).

Biavaschi, Costanza. *Recovering the Counterfactual Wage Distribution with Selective Return Migration*. Working Paper, 2013.

Bijwaard, Govert E. "Labour market status and migration dynamics." *IZA Discussion Paper* no. 4530 (2009).

Bijwaard, Govert E. "Immigrant migration dynamics model for The Netherlands." *Journal of Population Economics* 23, no. 4 (2010): 1213-1247.

Bijwaard, Govert E., Christian Schluter, and Jackline Wahba. "The impact of labour market dynamics on the return-migration of immigrants." *Review of Economics and Statistics* (forthcoming 2013).

Bijwaard, Govert E., and Jackline Wahba. "Do High-Income or Low-Income Immigrants Leave Faster?" *NORFACE Migration Discussion Paper* no. 2013-13.

Böhning, Wolf-Rüdiger. "*Studies in international labour migration*". London: Macmillan, 1984.

Borjas, George J. "Assimilation, changes in cohort quality, and the earnings of immigrants." *Journal of Labor Economics*, vol. 3, no. 4 (1985): 463-489.

Borjas, George J. "Immigrant and emigrant earnings: a longitudinal study." *Economic inquiry* 27, no. 1 (1989): 21-37.

Borjas, George J., and Bernt Bratsberg. "Who Leaves? The Outmigration of the Foreign-Born." The *Review of Economics and Statistics* 78, no. 1 (1996): 165-176.

Bratsberg, Bernt, Oddbjørn Raaum, and Kjetil Sørlie. "Foreign-born Migration to and from Norway." *International Migration, Economic Development and Policy* (2007): 259-291.

Bratsberg, Bernt, Oddbjørn Raaum, and Knut Røed. "When minority labor migrants meet the welfare state." *Journal of Labor Economics* 28, no. 3 (2010): 633-676.

Cameron, A. Colin and Pravin K. Trivedi. *Microeconometrics: methods and applications*. Cambridge university press, 2005.

Carliner, Geoffrey. "Wages, earnings and hours of first, second, and third generation American males." *Economic Inquiry* 18, no. 1 (1980): 87-102.

Carrión-Flores, Carmen E. "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*. 2006.

Chiswick, Barry "The effect of Americanization on the earnings of foreign-born men." *The Journal of Political Economy*, vol. 86, no. 5 (1978): 897-921.

Chiswick, Barry R. "Is the new immigration less skilled than the old?" *Journal of Labor Economics* (1986): 168-192.

Chiswick, Barry R., and Paul W. Miller. "Negative Assimilation of Immigrants: A Special Case, The." *Indus. & Lab. Rel. Rev.* 64 (2010): 502.

Cobb-Clark, Deborah, Barbara Hanel, and Duncan McVicar. "Immigrant wage and employment assimilation: A comparison of methods." (2012).

Cobb-Clark, Deborah, and Steven Stillman. "Emigration and the age profile of retirement among immigrants." *IZA Discussion Paper* no. 3874 (2008).

Cohen, Yinon and Yitchak Haberfeld. “Self-Selection and Return Migration: Israeli-Born Jews Returning Home from the United States during the 1980s.” *Population Studies* 55, no. 1 (2001).

Coniglio, Nicola D., Giuseppe De Arcangelis, and Laura Serlenga. "Intentions to return of clandestine migrants: The perverse effect of illegality on skills." *Review of Development Economics* 13, no. 4 (2009): 641-657.

Constant, Amelie, and Douglas S. Massey. "Return migration by German guestworkers: Neoclassical versus new economic theories." *International Migration* 40, no. 4 (2002): 5-38.

Constant, Amelie, and Douglas S. Massey. "Self-selection, earnings, and out-migration: A longitudinal study of immigrants to Germany." *Journal of Population Economics* 16, no. 4 (2003): 631-653.

Constant, Amelie F., and Klaus F. Zimmermann. "Circular and repeat migration: counts of exits and years away from the host country." *Population Research and Policy Review* 30, no. 4 (2011): 495-515.

Cortes, Kalena E. "Are refugees different from economic immigrants? Some empirical evidence on the heterogeneity of immigrant groups in the United States." *Review of Economics and Statistics* 86, no. 2 (2004): 465-480.

de Coulon, Augustin, and Matloob Piracha. "Self-selection and the performance of return migrants: the source country perspective." *Journal of Population Economics* 18, no. 4 (2005): 779-807.

Duleep, Harriet Orcutt. "Social security and the emigration of immigrants." *Social Security Bulletin* 57 (1994): 37.

Dustmann, Christian. "Earnings adjustment of temporary migrants." *Journal of Population Economics* 6, no. 2 (1993): 153-168.

Dustmann, Christian. "Savings behavior of return migrants." *Zeitschrift für Wirtschafts- und Sozialwissenschaften* 115 (1995): 511-533.

Dustmann, Christian. "Return migration, wage differentials, and the optimal migration duration." *European Economic Review* 47, no. 2 (2003): 353-369.

Dustmann, Christian (with discussants Samuel Bentolila, and Riccardo Faini). "Return migration: the European experience." *Economic Policy* (1996): 213-250.

Dustmann, Christian, Itzhak Fadlon, and Yoram Weiss. "Return migration, human capital accumulation and the brain drain." *Journal of Development Economics* 95, no. 1 (2011): 58-67.

Dustmann, Christian, and Albrecht Glitz. "Migration and education." *Handbook of the Economics of Education* 4 (2011): 327-439.

Dustmann, Christian, Albrecht Glitz, and Thorsten Vogel. "Employment, wages, and the economic cycle: Differences between immigrants and natives." *European Economic Review* 54, no. 1 (2010): 1-17.

Dustmann, Christian, and María Engracia Rochina-Barrachina. "Selection correction in panel data models: An application to the estimation of females' wage equations." *The Econometrics Journal* 10, no. 2 (2007): 263-293.

Dustmann, Christian, and Yoram Weiss. "Return migration: theory and empirical evidence from the UK." *British Journal of Industrial Relations* 45, no. 2 (2007): 236-256.

Dustmann, Christian, and Tommaso Frattini. “The Fiscal Effects of Immigration to the UK”. CReAM DP 22/2013.

Eckstein, Zvi, and Yoram Weiss. "On the wage growth of immigrants: Israel, 1990–2000." *Journal of the European Economic Association* 2, no. 4 (2004): 665-695.

Edin, Per-Anders, Robert J. LaLonde, and Olof Åslund. "Emigration of immigrants and measures of immigrant assimilation: Evidence from Sweden." *Swedish Economic Policy Review* 7, no. 2 (2000): 163-204.

Faini, Riccardo, Steinar Strøm, Alessandra Venturini, and Claudia Villosio. *Are foreign migrants more assimilated than native ones?*. No. 4639. IZA Discussion Papers, 2009.

Fertig, Michael, and Stefanie Schurer. "Labour market outcomes of immigrants in Germany: The importance of heterogeneity and attrition bias." (2007).

Gundel, Sebastian, and Heiko Peters. "What determines the duration of stay of immigrants in Germany? Evidence from a longitudinal duration analysis." *International Journal of Social Economics* 35, no. 11 (2008): 769-782.

Heckman, James J. "Sample selection bias as a specification error." *Econometrica: Journal of the econometric society* (1979): 153-161.

Hu, Wei-Yin. "Immigrant earnings assimilation: estimates from longitudinal data." *The American Economic Review* 90, no. 2 (2000): 368-372.

International Organization for Migration, 2011: “World Migration Report 2011”.

Izquierdo, Mario, Aitor Lacuesta, and Raquel Vegas. "Assimilation of immigrants in Spain: a longitudinal analysis." *Labour Economics* 16, no. 6 (2009): 669-678.

Jasso, Guillermina, and Mark R. Rosenzweig. “Estimating the Emigration Rates of Legal Immigrants Using Administrative and Survey Data: The 1971 Cohort of Immigrants to the United States.” *Demography* 19, no. 3 (1982): 279-290.

Jasso, Guillermina, and Mark R. Rosenzweig. *The new chosen people: Immigrants in the United States*. National Committee for Research, 1990.

Jensen, Peter, and Peder J. Pedersen. "To Stay or Not to Stay? Out-Migration of Immigrants from Denmark." *International Migration* 45, no. 5 (2007): 87-113.

Kaushal, Neeraj. "Earning trajectories of highly educated immigrants: Does place of education matter." *Indus. & Lab. Rel. Rev.* 64 (2010): 323.

Kaushal, Neeraj, and Ce Shang. *Earnings Growth of Mexican Immigrants: New versus Traditional Destinations*. No. 7427. Institute for the Study of Labor (IZA), 2013.

Kerr, Sari Pekkala, and William R. Kerr. *Economic impacts of immigration: A survey*. No. w16736. National Bureau of Economic Research, 2011.

Kim, Seik. *Economic Assimilation of Foreign Born Workers in the United States: An Overlapping Rotating Panel Analysis*. No. UWEC-2008-19. 2009.

Kırdar, Murat G. "Labor market outcomes, savings accumulation, and return migration." *Labour Economics* 16, no. 4 (2009): 418-428.

Kirwan, Frank, and Frank Harrigan. "Swedish-finnish return migration, extent, timing, and information flows." *Demography* 23, no. 3 (1986): 313-327.

Klein, Roger W., and Richard H. Spady. "An efficient semiparametric estimator for binary response models." *Econometrica: Journal of the Econometric Society* (1993): 387-421.

Klinthäll, Martin. *Return Migration from Sweden 1968-1996: A longitudinal analysis*. Almqvist & Wiksell International, 2003.

Klinthäll, Martin. "Immigration, integration and return migration." International Symposium on International Migration and Development, Turin (2006), pp. 28-30.

Klinthäll, Martin. "Refugee return migration: return migration from Sweden to Chile, Iran and Poland 1973–1996." *Journal of Refugee Studies* 20, no. 4 (2007): 579-598.

Kyriazidou, Ekaterini. "Estimation of a panel data sample selection model." *Econometrica: Journal of the Econometric Society* (1997): 1335-1364.

LaLonde, Robert J., and Robert H. Topel. "The assimilation of immigrants in the US labor market." In *Immigration and the workforce: Economic consequences for the United States and source areas*, pp. 67-92. University of Chicago Press, 1992.

Lam, Kit-Chun. “Outmigration of Foreign-Born Members in Canada.” *The Canadian Journal of Economics/Revue canadienne d’Economique*, 27, no. 2 (1994): 352-270.

Lindstrom, David P. "Economic opportunity in Mexico and return migration from the United States." *Demography* 33, no. 3 (1996): 357-374.

Lindstrom, David P., and Douglas S. Massey. "Selective emigration, cohort quality, and models of immigrant assimilation." *Social Science Research* 23, no. 4 (1994): 315-349.

Long, James E. "The effect of Americanization on earnings: some evidence for women." *The Journal of Political Economy* 88, no. 3 (1980): 620-629.

Longva, Pål. "Out-migration of immigrants: implications for assimilation analysis." *Department of Economics, University of Oslo, Memorandum* no. 4/2001.

Lubotsky, Darren. "Chutes or ladders? A longitudinal analysis of immigrant earnings." *Journal of Political Economy* 115, no. 5 (2007): 820-867.

Lubotsky, Darren. "The Effect of Changes in the US Wage Structure on Recent Immigrants' Earnings." *The Review of Economics and Statistics* 93, no. 1 (2011): 59-71.

Lukomskyj, Oleh, and Peter Richards. "Return Migration from Australia A Case Study." *International Migration* 24, no. 3 (1986): 603-632.

Lundh, Christer, and Rolf Ohlsson. "Immigration and economic change." In *Population, Economy, and Welfare in Sweden*, pp. 87-107. Springer Berlin Heidelberg, 1994.

Maré, David C., Melanie Morten, and Steven Stillman. "Settlement patterns and the geographic mobility of recent migrants to New Zealand." *New Zealand Economic Papers* 41, no. 2 (2007): 163-195.

Martins, Maria Fraga O. "Parametric and semiparametric estimation of sample selection models: an empirical application to the female labour force in Portugal." *Journal of Applied Econometrics* 16, no. 1 (2001): 23-39.

Massey, Douglas. "Understanding Mexican Migration to the United States." *American Journal of Sociology* 92, no. 6 (1987): 1372-1403.

de Matos, Ana Damas. *The careers of immigrants*. Working Paper, 2011.

Melenberg, Bertrand, and Arthur Van Soest. "Parametric and semi-parametric modelling of vacation expenditures." *Journal of Applied Econometrics* 11, no. 1 (1996): 59-76.

Michalowski, Margaret. "Foreign-born Canadian Emigrants and Their Characteristics,(1981-1986)." *International Migration Review*, vol. 25, no. 1 (1991): 28-59.

Nekby, Lena. "The emigration of immigrants, return vs onward migration: evidence from Sweden." *Journal of Population Economics* 19, no. 2 (2006): 197-226.

Newey, Whitney K. "Two-step series estimation of sample selection models." *The Econometrics Journal* 12, no. s1 (2009): S217-S229.

Organisation for Economic Cooperation and Development. *International Migration Outlook 2008*. OECD Publishing(2008).

Organisation for Economic Cooperation and Development. *International Migration Outlook 2013*. OECD Publishing (2013).

Picot, Garnett, and Patrizio Piraino. "Immigrant Earnings Growth: Selection Bias or Real Progress?." *Statistics Canada Analytical Studies Branch Research Paper Series* 340 (2012).

Pinger, Pia. "Come back or stay? Spend here or there? Return and remittances: The case of Moldova." *International Migration* 48, no. 5 (2010): 142-173.

Piore, Michael. *Birds of Passage: Migrant Labor and Industrial Societies*. Cambridge University Press, 1979.

Pischke, Jörn-Steffen. *Assimilation and the Earnings of Guestworkers in Germany*. No. 92-17. ZEW Discussion Papers, 1992.

Preston, Ian. ”The Effect of Immigration on Public Finances” CReAM DP 23/2013.

Ramos, Fernando. "Out-migration and return migration of Puerto Ricans." In *Immigration and the workforce: Economic consequences for the United States and source areas*, pp. 49-66. University of Chicago Press, 1992.

Reagan, Patricia B., and Randall J. Olsen. "You can go home again: Evidence from longitudinal data." *Demography* 37, no. 3 (2000): 339-350.

Rendall, Michael S., and Deborah J. Ball. "Immigration, emigration and the ageing of the overseas-born population in the United Kingdom." *Population Trends* (2004): 18-27.

Reyes, Belinda I. *Dynamics of immigration: return migration to western Mexico*. Public Policy Institute of California, 1997.

Reyes, Belinda I. "Immigrant Trip Duration: The Case of Immigrants from Western Mexico." *International Migration Review* 35, no. 4 (2001): 1185-1204.

Reyes, Belinda I. "Changes in trip duration for Mexican immigrants to the United States." *Population Research and Policy Review* 23, no. 3 (2004): 235-257.

Reyes, Belinda I., and Laura Mameesh. "Why Does Immigrant Trip Duration Vary Across US Destinations?." *Social science quarterly* 83, no. 2 (2002): 580-593.

Rooth, Dan-Olof, and Jan Saarela. "Selection in migration and return migration: Evidence from micro data." *Economics letters* 94, no. 1 (2007): 90-95.

Sander, Monika. "Return Migration and the 'Healthy Immigrant Effect'." *SOEPpapers on Multidisciplinary Panel Data Research* no. 60 (2007).

Sarvimäki, Matti. "Assimilation to a Welfare State: Labor Market Performance and Use of Social Benefits by Immigrants to Finland." *The Scandinavian Journal of Economics* 113, no. 3 (2011): 665-688.

Shorland, Philippa. "People on the move: A study of migrant movement patterns to and from New Zealand." New Zealand Department of Labour (2006).

Skuterud, Mikal, and Mingcui Su. "*Immigrant Wage Assimilation and the Return to Foreign and Host-Country Sources of Human Capital*." CLSRN Working Paper No. 30, 2009.

Van Hook, Jennifer, and Weiwei Zhang. "Who stays? Who goes? Selective emigration among the foreign-born." *Population research and policy review* 30, no. 1 (2011): 1-24.

Van Hook, Jennifer, Weiwei Zhang, Frank D. Bean, and Jeffrey S. Passel. "Foreign-born emigration: A new approach and estimates based on matched CPS files." *Demography* 43, no. 2 (2006): 361-382.

Venturini, Alessandra, and Claudia Villosio. "Labour-market assimilation of foreign workers in Italy." *Oxford Review of Economic Policy* 24, no. 3 (2008): 517-541.

Waldorf, Brigitte. "Determinants of International Return Migration Intentions." *The Professional Geographer* 47, no. 2 (1995): 125-136.

Willis, Robert J. "Wage determinants: A survey and reinterpretation of human capital earnings functions." *Handbook of Labor Economics* 1 (1986): 525-602.

Wooldridge, Jeffrey M. "Selection corrections for panel data models under conditional mean independence assumptions." *Journal of Econometrics* 68, no. 1 (1995): 115-132.

Wooldridge, Jeffrey M. *Econometric analysis of cross section and panel data*. The MIT press, 2nd edition, 2010.

Yahirun, Jenjira. "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, Roman. "Return Migration: An Empirical Investigation." *MPRA Paper* No. 13755, 2008.

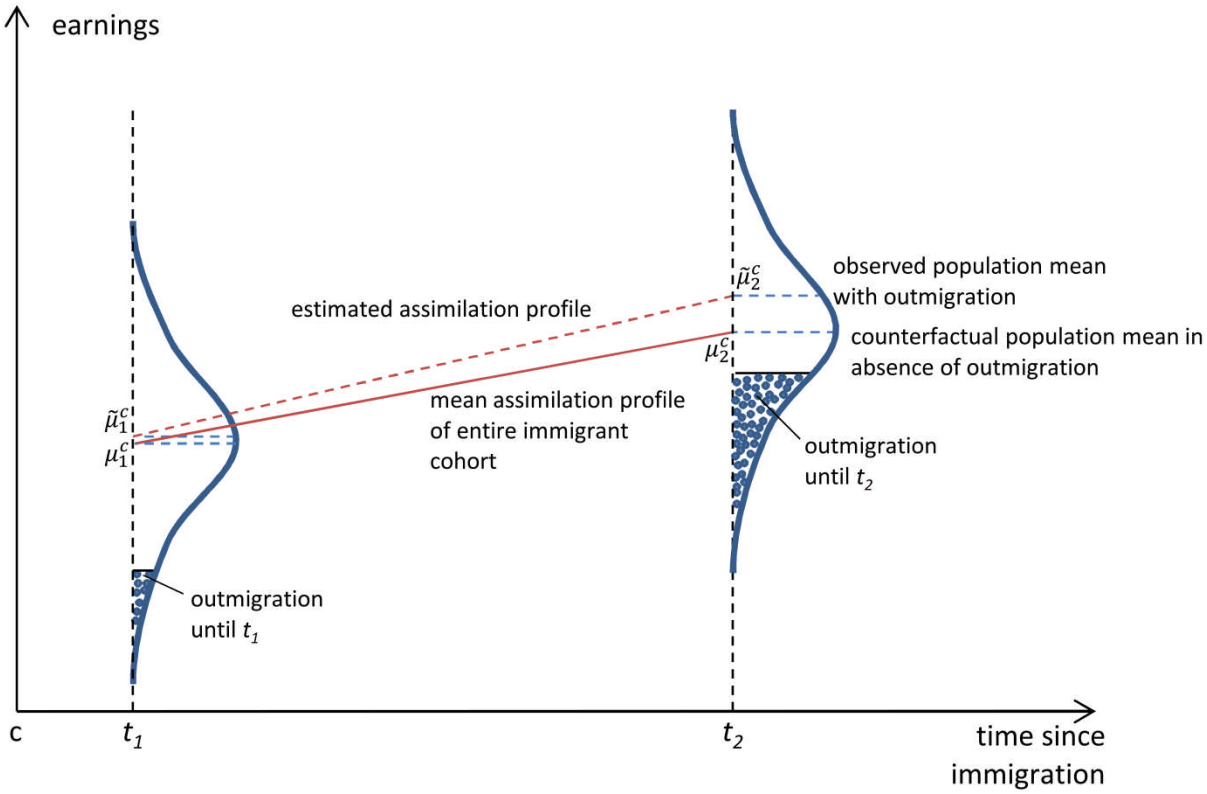
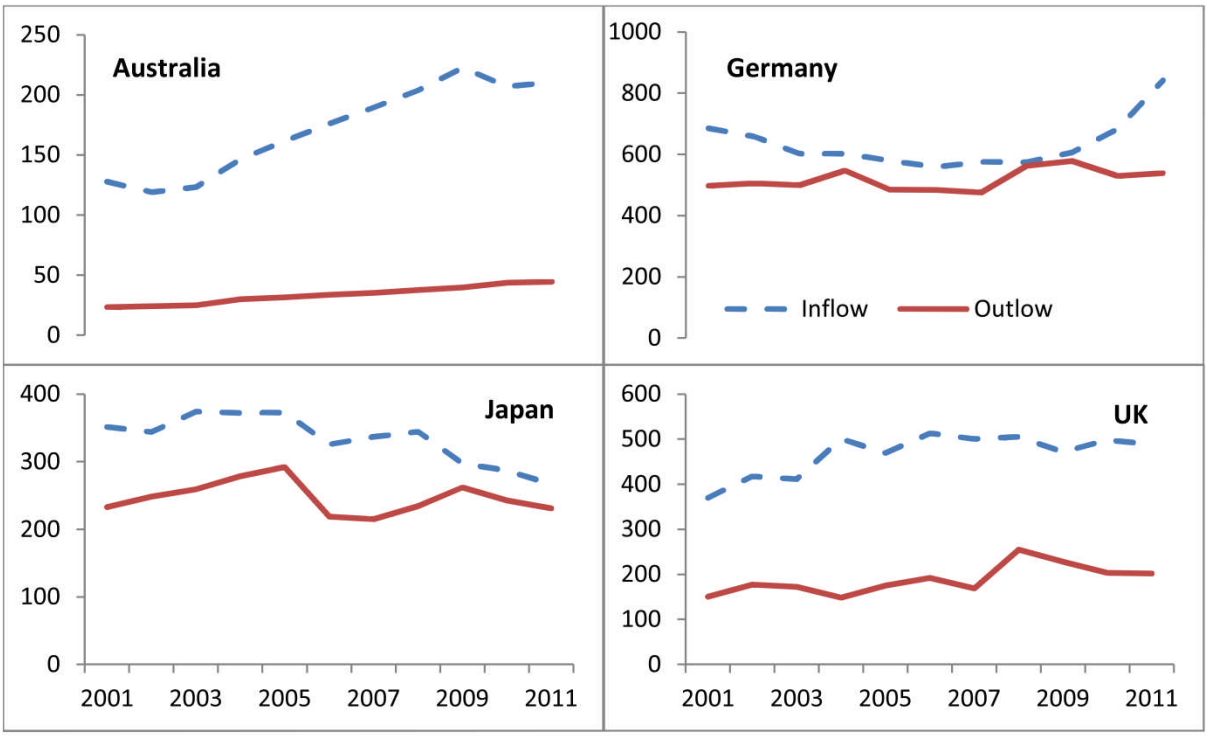


Figure 1: Biased estimation of earnings profiles when outmigration is selective



Source: OECD International Migration Outlook 2013

Figure 2: Immigrant in- and outflows in thousands for selected OECD countries

Migrant Survival, by Gender.

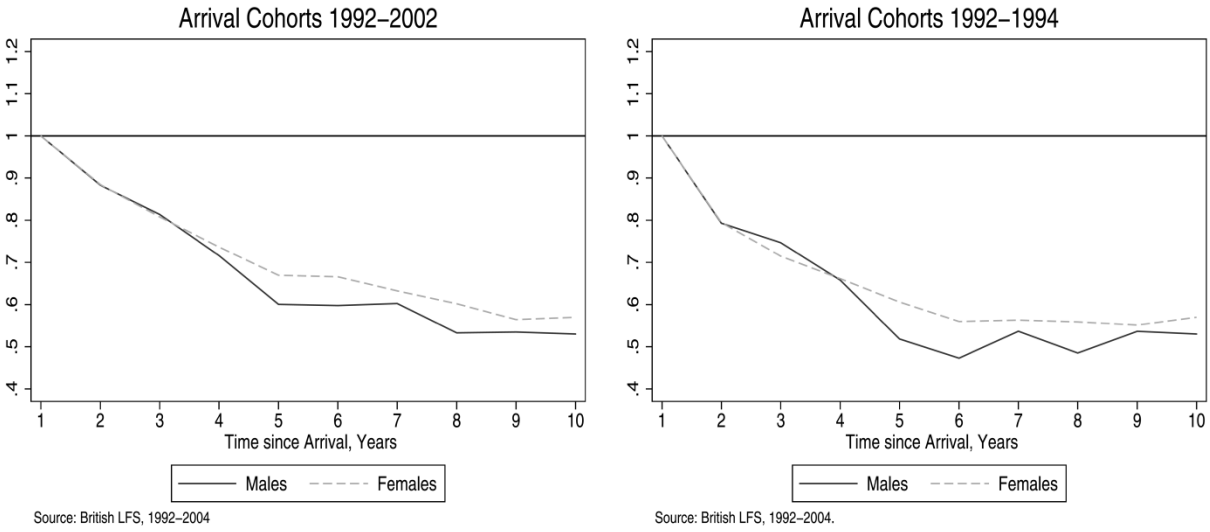


Figure 3: Immigrant survival rates by gender, from Dustmann and Weiss (2007)

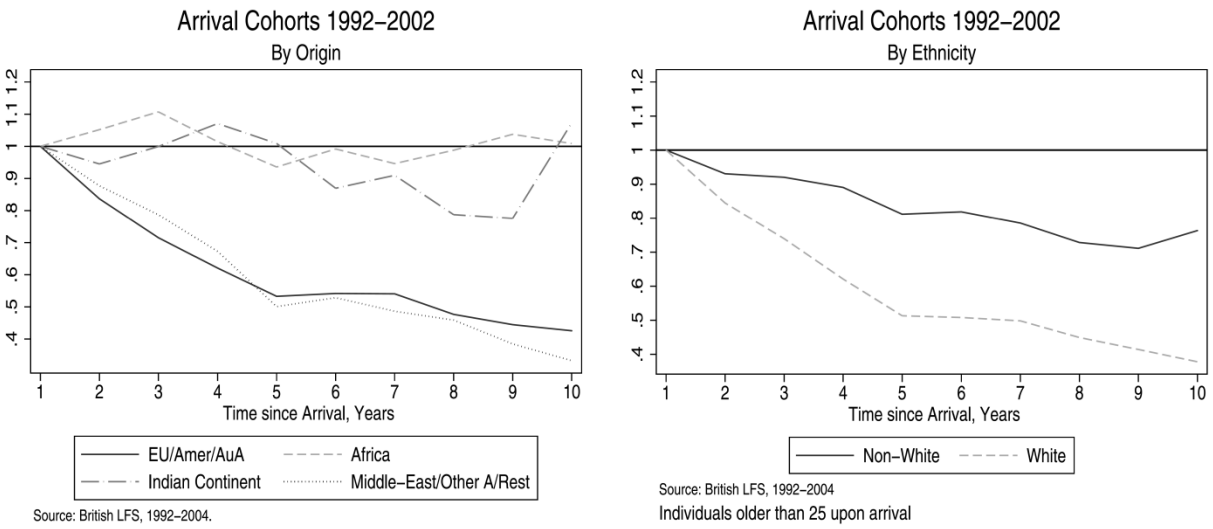


Figure 4: Immigrant survival rates by origin and ethnicity, from Dustmann and Weiss (2007)

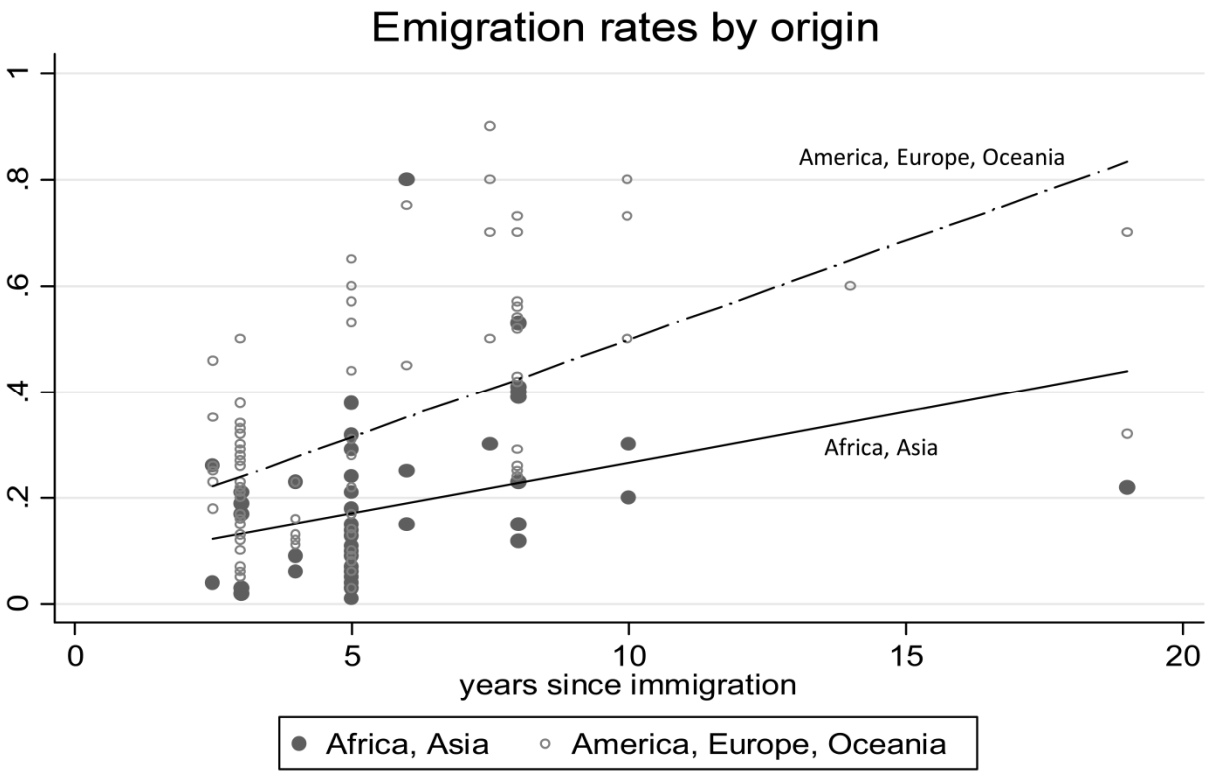


Figure 5: Foreign-born emigration by origin region

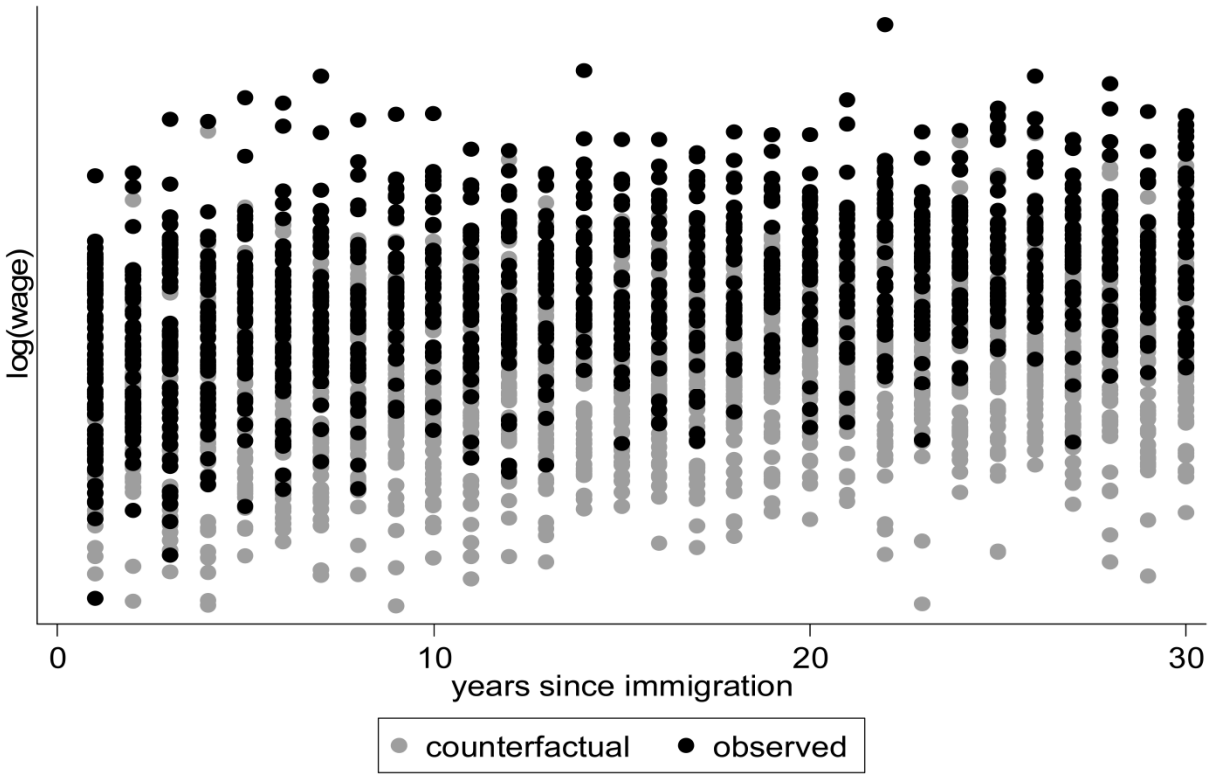


Figure 6: Observed and counterfactual log-earnings

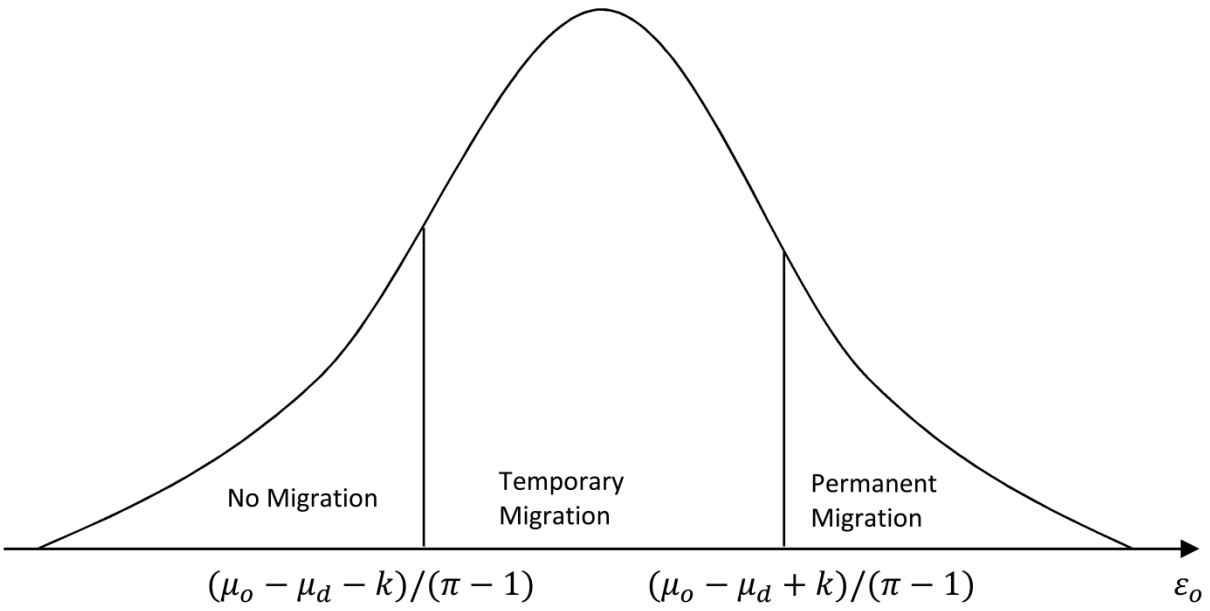


Figure 7: Selection of emigrants and re-migrants under higher returns to skills in the host country