Estimating Immigrant Earnings Profiles when Migrations are Temporary[1]

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

Abstract: The assumption that all migrations are permanent, which pervaded the early microdata-based research on immigrant career profiles, is not supported by the empirical evidence. Rather, many – if not most – migrations appear to be temporary. In this paper, therefore, we illustrate the estimation challenges when migrations are temporary. As in an overwhelming share of the selective out-migration literature, our basic structure assumes that the process that determines out-migration is unrelated to other choices that affect wage growth, such as human capital investment or labour supply decisions, which greatly simplifies the analysis. When the choice of whether and when to out-migrate also affects decisions that determine wage growth, the problem becomes inherently dynamic and requires a more structural approach to estimation, which we briefly discuss.

Keywords: immigration, return migration, assimilation, earnings profile, selection

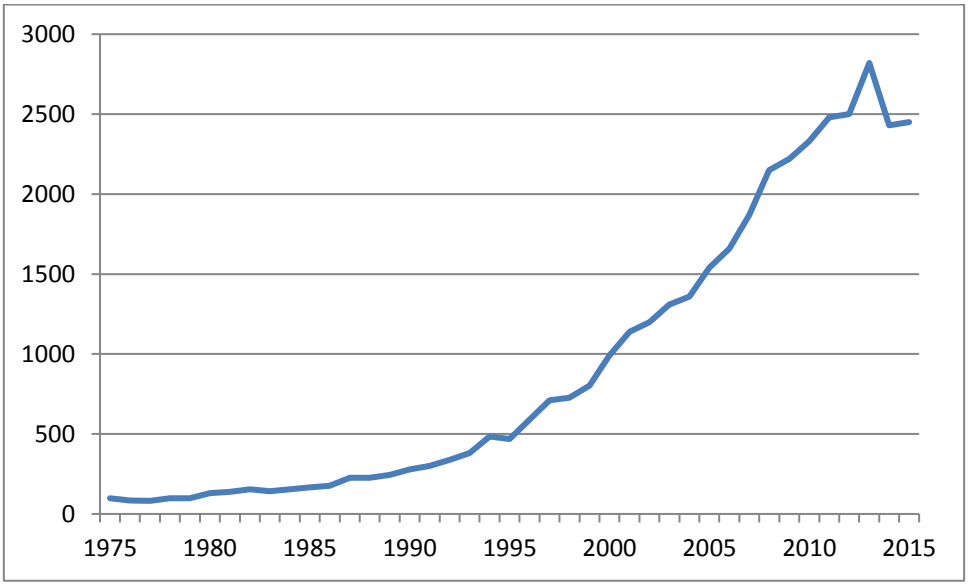
JEL codes: F22, J15, J31, J61, O15

[1] This paper is based on Christian Dustmann's Presidential Address at the EALE/SOLE meeting in Montreal/Canada, June 2015. Christian Dustmann acknowledges funding through the ERC Advanced Grant DMEA and by the DFG (grant no. 1764/0).
[2] Department of Economics and Centre for Research and Analysis of Migration (CReAM), University College London
[3] Department of Economics and Centre for Research and Analysis of Migration (CReAM), University College London

1. Introduction

Ever since Chiswick's (1978) seminal paper on the earnings assimilation of immigrants in the U.S., the estimation of immigrants' earnings and career profiles has been an important and growing area of research in migration economics. For instance, a simple keyword search on Google Scholar using "earnings assimilation immigrants" shows a steep increase from fewer than 100 papers a year on the subject in the late 1970s to over 2,800 in 2013 (see Figure 1). The broad interest in this subject is not surprising: the earnings that immigrants receive in destination countries and the evolution of their earnings paths are an important indicator not only of their own success and performance, but also of their overall contribution to the host countries' economies in terms of GDP growth and tax contributions.

Figure 1: Scholarly Papers on Immigrant Earnings Assimilation



Notes: The graph plots the results by year of a Google keyword search using "earnings assimilation immigrants" (https://scholar.google.co.uk/, accessed on February 29, 2016).

The estimation of *immigrant earnings equations*, however, is far from straightforward because, in contrast to Mincer type earnings equations for workers born and likely to remain

permanently in the same country, they are affected by selection into and out of the underlying population. Such selection is an issue both when immigrants arrive and again if they leave the country before the end of their productive lives. In the first instance, an identification problem arises for the type of cross-sectional analysis used by Chiswick (1978). If the cohort quality of new arrivals changes, this problem may lead to biased estimates of earnings profiles (see Borjas 1985), a risk that researchers counter in a variety of ways (e.g., Baarth, Bratsberg and Raaum, 2004; Bratsberg, Baarth and Raaum, 2006; and see Dustmann, Glitz and Vogel's 2010 assessment of Borjas's assumptions). Today, the availability of better data – often providing longitudinal information on single arrival cohorts – allows authors more flexibility in addressing this issue. The second case, selective out-migration, is largely ignored in most early literature, probably because the type of data available at the time made it impossible to address. This situation is nevertheless at least as problematic as the first type of selection and generally far more difficult to deal with.

In this paper, we discuss problems that arise when estimating immigrants' earnings profiles, and when out-migration occurs. Our analysis is related to several earlier papers of ours on the subject. In Dustmann and Görlach (2015), we address the problems that arise when estimating the earnings profiles of immigrants and when out-migration is not random. One major conjecture that pervades the extant literature is that the migrant decisions to out-migrate are unrelated to decisions that may affect or be affected by career profiles. Although this assumption simplifies analysis considerably by allowing researchers to use off-the-shelf methods to deal with selection and attrition, usually no attempt is made to provide a behavioural model of why some migrants may want to return home and how this choice interacts with other immigrant decisions. In Dustmann and Görlach (2016), this assumption is relaxed. We develop a general dynamic framework for modelling out-migration, and discuss various motives as to why migrants may want to return. We further point out how migrant

return decisions are in fact influenced by and may in turn influence other choices that affect the migrants' career and ultimately their earnings paths in the destination country. Under more general scenarios, the reduced form estimators discussed in Dustmann and Görlach (2015) may not be applicable, particularly when stochastics must be taken into account or if migrants can choose how much to invest in human capital. In such a case, estimation must rely on the type of structural modelling applied in Adda, Dustmann and Görlach (2016).

In Section 2 of this paper, we briefly review the challenges that selective out-migration poses for the estimation of immigrant career profiles, how these challenges can be addressed in a reduced form context, and which assumptions are needed to identify causal parameters. We then discuss how giving up the assumption of independence between out-migration decisions and decisions about career profiles may contaminate the causal interpretation of estimates obtained from reduced form estimators, inducing a need for structural estimation of career profiles. In Section 3, we develop a very simple model of return migration and earnings that allows us to obtain linear equations for earnings profile estimation, which we simulate under two scenarios that differ only in how immigrants incorporate current shocks-to-earnings when making their out-migration decisions. Whereas in the first scenario, these shocks are not observed either at the beginning of each period or when the out-migration decision is made; in the second scenario, the realizations are observed by the agent but not the econometrician. We show that in the first case, the estimators outlined in Section 2 produce consistent parameter estimates of the wage growth of the initial arrival cohort, which does not hold for the second case.

2. Selective Outmigration and the Estimation of Earnings Equations

In a recent report, the OECD (2008) estimates that, depending on country and time period considered, 20 to 50 percent of immigrants leave the host country within the first five years after arrival. In 2011, foreign-born outflows stood at 21 percent of migrant inflow to Australia; 41 percent, 64 percent, and 76 percent to the UK, Germany and Spain; and 71 percent and 87 percent to Korea and Japan, respectively (OECD, 2013). For the U.S., an estimated 2.1 million foreign-born individuals emigrated between 2000 and 2010 (Bhaskar, Arenas-Germosén and Dick, 2013).

In Figure 2 (taken from Dustmann and Weiss 2007), we display the survivor function for immigrants who arrived in the UK over the 1992– 2002 period and stayed for at least one year. As the figure shows, return migration is quite substantial, with only about 60 percent of an arrival cohort still in the country after five years. A comparison of Figure 2a and b further suggests that survival rates are similar for males and females but quite different for immigrants of different origins.[4]

Because these out-migrations are unlikely to be random, they may select the original arrival cohort along some dimension that is correlated with outcomes, such as earnings. If the parameter of interest is an arrival cohort's earnings growth rate in the host country – measured here by the change in log earnings per year a migrant has been in the country[5] – then selective out-migration will produce biased estimates of this parameter when (log) earnings are regressed on years since migration. This bias will depend on the type of

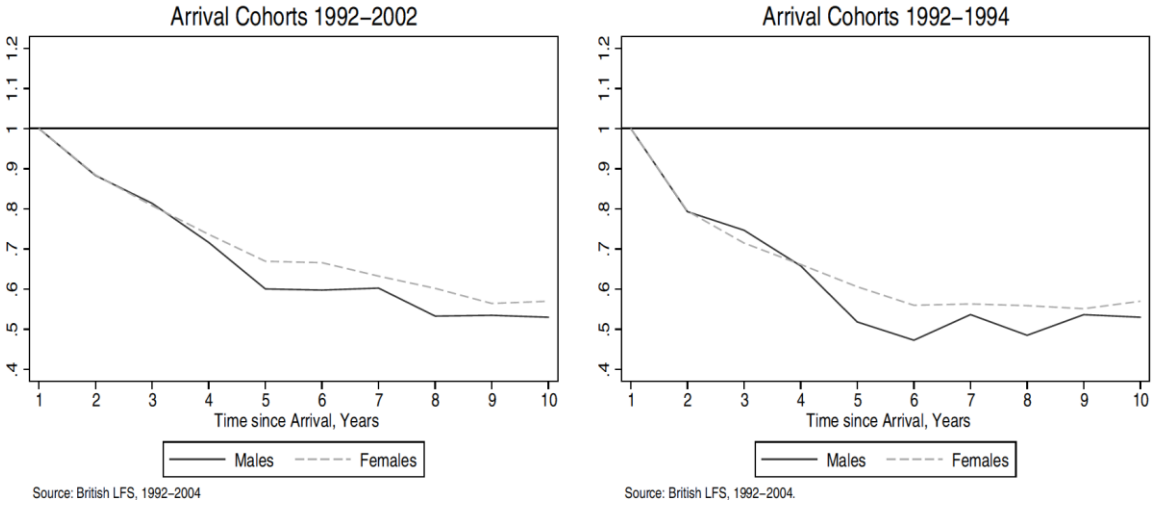
[4] In compiling these figures, we take advantage of the fact that each wave of the British Labour Force Survey is a random sample from the population, so by using information on arrival year, we can compute the survival of every arrival cohort over the 1992–2002 window. Some numbers in the left-hand figure are larger than 1 because of sampling error.

[5] Although typically, we would condition on education and (potential) labour market experience in the origin country while allowing for non-linearity in these profiles; for simplicity, we ignore this aspect in our subsequent discussion.

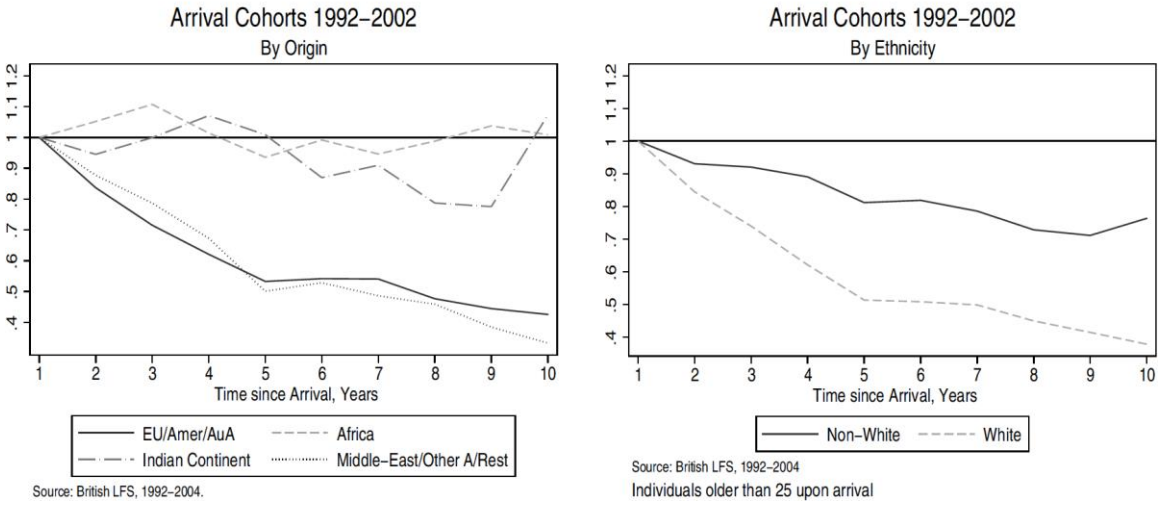
selection. For instance, if migrants with higher unobserved productivity tend to return earlier, the earnings distribution of those who have been in the country longer will be truncated. Hence, instead of pinpointing the increase in mean earnings given time in the country for the original arrival cohort, the regression line will identify the increase in the mean of the truncated earnings distributions, where truncation increases with cohort age.

Figure 2: Survival Rates

Migrant Survival, by Gender.



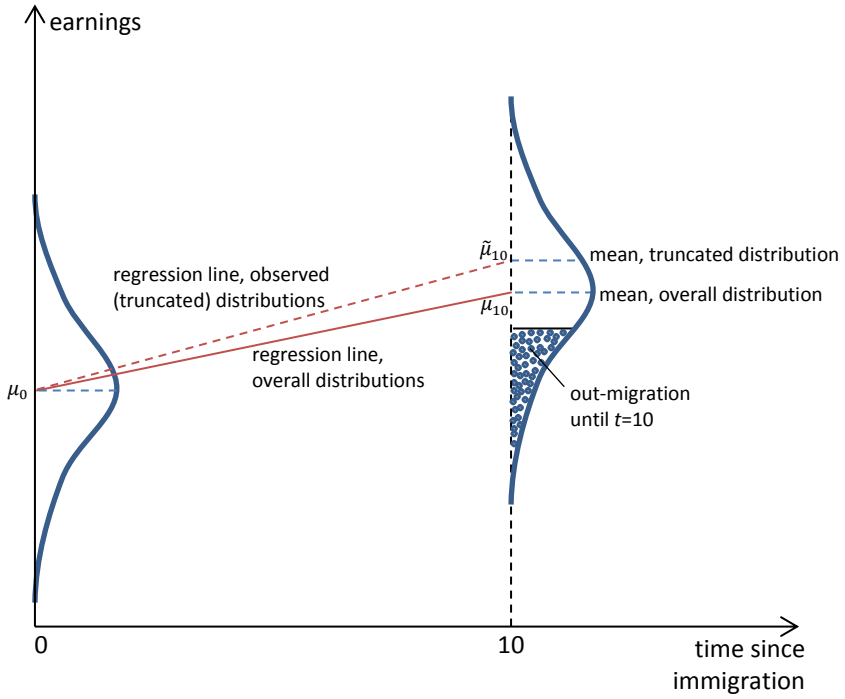
Migrant Survival, by Origin and Ethnicity.



Source: Dustmann and Weiss (2007), based on British Labour Force Survey data, 1992–2004.

In Figure 3, where we have depicted (log) earnings on the vertical axis, and time in the country on the horizontal axis, we illustrate the extreme case of out-migration determined by earnings only, which truncates the remaining earnings distribution. This figure further assumes that immigrants' ranking within the earnings distribution does not change over time, although we relax this assumption in the subsequent discussion. The first graph (left side) represents the earnings distribution for a cohort of migrants who arrive simultaneously just after arrival in the host country. The second then gives the earnings distribution of this same arrival cohort 10 years later. Supposing a data set made up of two repeated cross-sections (e.g., census data), if there is no out-migration or if out-migration is random, a regression of log earnings on time in the country for this arrival cohort will identify the slope of the solid line, which graphs the wage growth of the original arrival cohort. If, however, out-migration is selective, the distribution of earnings in $t = 10$ will not be representative of the earnings distribution of the original arrival cohort. If it is negatively selective, meaning that those with lower earnings potential leave the country earlier, then the distribution of earnings in $t = 10$ will be truncated from below, and the regression line will pass through the mean of the *truncated* distribution, corresponding to the dashed line in Figure 3. Thus, the coefficient obtained from a simple regression using the two cross-sections will be upwardly biased and the potential wage growth of the original arrival cohort overestimated. If, on the other hand, out-migration is positively selective, the estimate of wages growth for this cohort will be downwardly biased. Then, to recover the wage growth of the original arrival cohort, we would need to re-construct the wage distribution in $t = 10$. How challenging this re-construction would be depends on selection type.

Figure 3: Selection Bias under Non-Random Out-Migration



Several studies have attempted to determine the direction of this selection empirically for a variety of major immigration countries. Most studies using Canadian or U.S. data find that those who leave are predominantly drawn from the lower end of the earnings distribution (cf. Borjas, 1989; Hu, 2000; Lubotsky, 2007; Picot and Piraino, 2012). The evidence for Europe on the other hand is more mixed, and results from larger administrative data from Sweden and The Netherlands point to a U-shaped selection pattern of out-migrants, where the lowest and highest earners are most likely to leave (see Nekby, 2006; Bijwaard and Wahba, 2014).

2.1 Research question

Before examining the empirical challenges of estimating the wage growth of a particular immigrant arrival cohort when out-migration is selective, we need to briefly reflect on the questions typically related to this issue. In Figure 2, the solid line, whose slope characterizes wage growth, illustrates the wage growth of an arrival cohort under the *counterfactual*

situation that no member has left the country. Yet why should we be interested in a counterfactual parameter representing a purely hypothetical migrant composition in later periods and an earnings profile that will never be realized. For many, if not most, of the questions typically asked by policy makers, this hypothetical earnings path is not relevant. What they really care about is the earnings position of migrants who are in the country, which is given by the dashed line. If, for instance, we want to predict the tax contributions that immigrants who arrived in t_0 will make between t_0 and t_1, we need to calculate the earnings of migrant populations in the country at any point between t_0 and t_1 as a basis for computing their tax contributions. This sum can be derived from the earnings of all those who remain from the original arrival cohort in the years after arrival, with average growth represented by the dashed line in the figure. Obviously, knowledge of the counterfactual earnings profile of the initial arrival cohort had nobody re-migrated (solid line) would be of little help in addressing this question.

However, whereas any interpretation of the wage growth rate in a behavioural model refers to a parameter that can be related to a single individual, the growth rate of the dashed line is purely descriptive and thus not interpretable within a decision framework like a Mincer wage equation. Furthermore, estimating the wage growth of the initial immigrant arrival cohort allows us to assess whether out-migration is positively or negatively selective. In Figure 3, for instance, the dashed line being above the solid line shows that the selective out-migration is negative. Identifying this direction is of key importance for policy: if the most (least) productive immigrants are the first (last) to leave the country, then the policy maker may want to consider measures to prevent that scenario. The wage growth experienced by individuals, and which determines their choices, thus becomes a key parameter of interest.

2.2 Estimating wage equations when out-migration is selective

In the simple example above, we implicitly assume that out-migration is correlated with unobservable characteristics that are constant for individuals, meaning that selection is related to unobserved individual earnings potential. Under this assumption, earnings *growth* can be identified as long as repeated information is available for the same individual, as is the case in panel data. An early paper by Pischke (1992) implements such an approach, where his comparison of OLS and within-group estimates suggests that guest-workers leaving Germany in the 1980s were slightly positively selected on earnings.

To illustrate, we consider the following simple earnings function:

$$(1)\ y_{it} = \alpha + \gamma\ ysm_{it} + \epsilon_{it},$$

with $\epsilon_{it} = \mu_i + e_{it}$, where y_{it} and ysm_{it} are the log earnings and years since migration, respectively, of individual i in a given entry cohort observed in period t. ϵ_{it} summarizes the unobserved determinants of the individual's earnings, which comprise an individual specific component μ_i and a time varying component e_{it}. How the relation in equation (1) should be estimated to retrieve parameter γ depends on the assumptions made about the relation between ϵ_{it} and the process that governs the out-migration selection.

We denote an out-migration event in any year t by the indicator variable $s_{it} = 0$. Because data sets do not typically follow individuals across international borders, it is impossible to identify repeat migration by the same individual. We thus consider only first-time migrants whom we assume to be permanently lost from the sample if they choose to return home. In this case, out-migration is an absorbing state, implying that $s_{it} = s_{i0} \cdot s_{i1} \cdot \ldots \cdot s_{it-1} \cdot s_{it}$. It then follows that the mean of the earnings distribution conditional on the individual's years and continued presence in the country is given by

$$(2)\ E[y_{it}|ysm_{it}, s_{it} = 1] = \alpha + \gamma\ ysm_{it} + E[\mu_i + e_{it}|ysm_{it}, s_{it} = 1].$$

Obviously, if $E[\mu_i + e_{it}|ysm_{it}, s_{it} = 1] = 0$, the OLS estimator will produce an unbiased estimate of γ. This will be the case when out-migration is random. Hence, in Figure 2, the earnings distribution of those still in the country at $t = 10$ will be the same as the earnings distribution for the original arrival cohort had nobody out-migrated. If that assumption is violated, however, meaning that $E[\epsilon_{it}|ysm_{it}, s_{it} = 1] \neq 0$, OLS will yield a biased estimate of γ.

How to solve this problem depends on selection type. When selection occurs only on time-constant unobservable components, μ_i, $E[\epsilon_{it}|ysm_{it}, s_{it} = 1] \neq 0$ but $E[\epsilon_{it} - \epsilon_{it-1}|ysm_{it}, s_{it} = s_{it-1} = 1] = 0$. Hence, the selection terms can be eliminated through simple differencing without imposing any further assumptions on the selection process, meaning that a difference estimator can be used to obtain an unbiased estimate of γ. Note that in this case, although the slope parameter γ is identified, the initial earnings α are not.

Another way to identify earnings growth in the absence of panel data is that proposed by Hu (2000) and Lubotsky (2007), who use stock-sampled data that include all (or a random sample of) individuals who survive in the host country at least until period $\bar{t}$, as well as multiple random samples of the surviving cohort in the years before $\bar{t}$. These samples need not be longitudinal and can be repeated cross-sections, meaning that individuals need not be identifiable in different waves. Such data may be generated, for example, when immigrant samples are linked to administrative data, allowing re-construction of earlier observations for all those in the sample in period $\bar{t}$. Using different samples of U.S. immigrants, both authors highlight that (negatively) selective out-migration biases OLS estimates from pooled cross-sections upward.

In this latter case, expression (2) becomes

$$(2')\ E[y_{it}|ysm_{it}, s_{i\bar{t}} = 1] = \alpha + \gamma\ ysm_{it} + E[\mu_i + e_{it}|ysm_{it}, s_{i\bar{t}} = 1].$$

As before, if $E[\epsilon_{it}|ysm_{it}, s_{i\bar{t}} = 1] \neq 0$, OLS will produce biased estimates. If, however, selection is indeed only on time-constant unobservables μ_i (i.e., $E[e_{it}|ysm_{it}, s_{i\bar{t}} = 1] = 0$), then $E[\epsilon_{it}|ysm_{it}, s_{i\bar{t}} = 1] = E[\epsilon_{i\tau}|ysm_{i\tau}, s_{i\bar{t}} = 1]$ for $t \neq \tau$. In this case, because the sample is restricted to individuals who stay in the country at least until period $\bar{t}$, simple OLS estimation of (1) will produce unbiased estimates of parameter γ. For this sub-sample, the selection term $E[\mu_i|ysm_{it}, s_{i\bar{t}} = 1]$ is constant in each wave t of the sample. Therefore, when (1) is estimated using these data, the selection term is absorbed into the intercept, and simple OLS estimation will produce unbiased estimates of γ but not α.

When selection is on time-variant unobservables e_{it}, however, $E[e_{it}|ysm_{it}, s_{it} = 1]$ and $E[e_{it}|ysm_{it}, s_{i\bar{t}} = 1]$ change over time. In this case, neither the difference or FE estimator for panel data nor the OLS estimator for stock-sampled data will produce unbiased estimates of γ. Rather, deriving a consistent estimate requires that the selection process be modelled explicitly, as described in Dustmann and Görlach 2015 (see also Dustmann and Rochina-Barrachina (2007) for estimators that account for correlated unobserved individual heterogeneity in selection and wage equations).

3. A Model of Out-migration and Earnings

We illustrate the points discussed above by developing and then simulating a model of return migration in which return generates a correlation between the unobservables that simultaneously determine out-migration and wage growth. Although simple, our model clearly demonstrates the challenge that selective out-migration poses for econometric estimation of immigrant earnings profiles. We first consider a scenario in which returns to skills α_1 are higher in an individual's country of origin o, while the general productivity level

α_0 is higher in the foreign destination country d (cf. the Chiquiar and Hanson, 2005, scenario for Mexico-U.S. migration) to which an individual may choose to migrate for a certain period of time.

In this model, log earnings in location $l \in \{o, d\}$ are given by

$$y_{it}^l = \alpha_0^l + \alpha_1^l S_{it} + e_{it}^l,$$

where S_{it} denotes the (log of) accumulated skills, α_1^l is the return to skills in location l (meaning that $\exp(\alpha_1^l S_{it})$ is i's human capital in period t in country l), α_0^l is the log rent on human capital in country l, and $e_{it}^l \sim N\left(0, \sigma_{e^l}^2\right)$ is a transitory shock to earnings. Assuming for simplicity that skills accumulate at a constant rate $(\exp(\theta) - 1)$, given some initial endowment $S_{i0} \sim N(\mu_S, \sigma_S^2)$, the log level of skills accumulated is $S_{it} = S_{it-1} + \theta$. When the rental rate to human capital in the destination country is higher than in the origin country ($\alpha_0^d > \alpha_0^o$) but returns to skills are higher in the latter ($\alpha_1^o > \alpha_1^d$), individuals may migrate temporarily from o to d to benefit from a high α_0^d when their skills are low but a high skill price α_1^o in the origin country when they have increased.

Because we are focusing here on immigrant earnings, we consider a population of individuals who have chosen to emigrate at the beginning of their working lives. If after migrating, immigrants can choose their own locations (and other aspects like consumption) in each period and do so to maximize expected life-time utility, then a migrant's life cycle problem at age a_{it} (i.e., with $T - a_{it}$ remaining until the end of life) can be written as

$$V(\Omega_{it}) = \max_{c,l} \sum_{t=0}^{T} \beta^t E[u(c_{it})],$$

$$s.t. A_{it+1} \leq (1 + r)A_{it} + y_{it}^l(S_{it}) - c_{it},$$

where the vector $\Omega_{it} = \{a_{it}, l_{it-1}, A_{it}, S_{it}\}$ collects the relevant state variables age (a_{it}), location (l_{it-1}), stock of financial assets (A_{it}) and level of skills (S_{it}). In such a simple framework, when returns to assets are equal across locations and individuals are assumed not to be credit constrained, consumption is perfectly smoothed. Moreover, as long as skills are accumulated at the same rate in the two locations, emigration and return decisions do not affect life time earnings beyond the current period. Location choices are thus reduced to a simple maximization of the current wage.[6]

Such an endogenous return decision implies the following (log) earnings function in the destination country:

$$(3)\ y_{it}^{d} = \alpha_0^d + \alpha_1^d S_{i0} + \alpha_1^d \theta ysm_{it} + e_{it}^d \equiv \alpha_0^d + \gamma\, ysm_{it} + \mu_i + e_{it}^d,$$

where we have re-written $\alpha_1^d S_{i0}$ as an individual fixed effect μ_i, and the structural parameter $\alpha_1^d \theta$ as γ. At any t, earnings in the destination country are only observed for a sub-set of the initial arrival cohort, those who have not yet returned home. In this equation, the parameter of interest is $\alpha_1^d \theta$ just as in the many studies since Chiswick's (1978) that seek to estimate the earnings equations of international migrants.[7] By combining the returns to skills, α_1^d, and rate of skill accumulation, θ, in the destination country, this parameter captures the contribution of (log) skill growth to human capital. As above, we consider only the case in which all migrants are first-time migrants and are lost from the sample if they choose to return (meaning that return is an absorbing state).

[6] Our argument that a correct estimation of immigrant earnings profiles requires a modelling of migrants' return decisions carries over to more complicated models and to other motives for why individuals may choose to migrate only temporarily (see Dustmann and Görlach, 2016, for an extensive discussion of such cases within a life cycle framework).

[7] This group includes articles by Carliner (1980), Long (1980), Borjas (1985), LaLonde and Topel (1992), Pischke (1992), Dustmann (1993), Lindstrom and Massey (1994), Edin, LaLonde and Åslund (2000), Hu (2000), Duleep and Dowhan (2002), Fertig and Schurer (2007), Lubotsky (2007), Skuterud and Su (2009), Sarvimäki (2011), Picot and Piraino (2012), and Abramitzky, Boustan and Eriksson (2014).

The sources of selection bias in the estimation of earnings processes now depend on the timing assumed in the model. We consider two scenarios. First, the location choices in each period are made based on expected future earnings, and transitory shocks to earnings e_{it}^{l} are realized only after a location l has been chosen. Second, rather than the return migration decision being based on expected earnings, the potential earnings level in the destination country (including the transitory shock e_{it}^{d}) is known in each period before the location choice is made and is the basis for this decision. These two scenarios differ only in how the transitory income shock information is used in the period decision problem, a variable that, as we now demonstrate, has important implications for estimating immigrant earnings profiles. Whereas in the first scenario, estimators of the type discussed in Section 2, in conjunction with longitudinal or stock-sampled data, identify the earnings growth of individuals in the original arrival cohort, they cannot do so in the second scenario, in which return decisions are also based on the current transitory shock to earnings.

3.1 The return decision

3.1.1 Return decision based on expected earnings

In the first scenario, in which return decisions are based on expected earnings and transitory shocks e_{it}^{l} are only realized after a location l has been chosen, the selection rule can be written as

$$
\begin{aligned}
s_{it} &= \prod_{\tau \le t} \mathbf{1}\left[E\left[y_{i\tau}^{d} - y_{i\tau}^{o} | S_{i0}, ysm_{i\tau}\right] > 0\right] \\
&= \prod_{\tau \le t} \mathbf{1}\left[\alpha_0^d - \alpha_0^o + \left(\alpha_1^d - \alpha_1^o\right) S_{i0} + \left(\alpha_1^d - \alpha_1^o\right) \theta ysm_{i\tau} > 0\right] \\
&\equiv \prod_{\tau \le t} \mathbf{1}[\bar{v} + v_i + \delta\, ysm_{i\tau}) > 0],
\end{aligned}
$$

where the structural parameters are summarized as $\bar{\nu} = \alpha_0^d - \alpha_0^o$, $\nu_i = \left(\alpha_1^d - \alpha_1^o\right)S_{i0}$ and $\delta = \left(\alpha_1^d - \alpha_1^o\right)\theta$. Based on equation (3), the observed earnings growth of immigrants still residing in the destination country is

$$E\left[y_{it_2}^d | ysm_{it_2}, s_{it_2} = 1\right] - E\left[y_{it_1}^d | ysm_{it_1}, s_{it_1} = 1\right]$$
$$= \gamma + E\left[\mu_i | ysm_{it_2}, s_{it_2} = 1\right] - E\left[\mu_i | ysm_{it_1}, s_{it_1} = 1\right]$$

for $ysm_{it_2} - ysm_{it_1} = 1$. In such a scenario, where $\alpha_1^d < \alpha_1^o$, immigrants who are more skilled on arrival (with higher S_{i0}, and thus lower $\nu_i = \left(\alpha_1^d - \alpha_1^o\right)S_{i0}$) will choose to return earlier because they need less time to reach a sufficiently high level of skills to make return to the origin country worthwhile. The selective return migration on the time-constant unobservable determinants ($\mu_i = \alpha_1^d S_{i0}$) of individual earnings is therefore positive, meaning that $E\left[\mu_i | ysm_{it_2}, s_{it_2} = 1\right] - E\left[\mu_i | ysm_{it_1}, s_{it_1} = 1\right] < 0$ biases simple OLS estimates of γ downward. The structural earnings growth parameter γ can, however, be recovered by regression on a differenced earnings equation that eliminates the time-constant unobserved effect μ_i. It can also be identified using pooled estimation in levels for a sub-sample of immigrants known to stay until some predetermined time period $\bar{t}$ (see Hu, 2000, and Lubotsky, 2007, for stock samples of U.S. immigrants).

3.1.2 Return decision based on realized earnings

In the second scenario, rather than basing return decisions on expected earnings, the immigrants know their potential earnings level in the destination country (including the transitory shock e_{it}^d), and can thus choose their location based on this earnings realization. The selection rule for staying in the destination country then becomes

$$s_{it} = \prod_{\tau \le t} \mathbf{1}\left[y_{i\tau}^{d} - E[y_{i\tau}^{o} | S_{i0}, ysm_{i\tau}] > 0\right]$$

$$= \prod_{\tau \le t} \mathbf{1}\left[\alpha_0^d - \alpha_0^o + \left(\alpha_1^d - \alpha_1^o\right) S_{i0} + \left(\alpha_1^d - \alpha_1^o\right) \theta ysm_{i\tau} + e_{i\tau}^{d}) > 0\right]$$

$$\equiv \prod_{\tau \le t} \mathbf{1}\left[\bar{v} + v_i + \delta\, ysm_{i\tau} + e_{i\tau}^{d}) > 0\right].$$

Immigrants who in a non-stochastic setting would choose to return in a given period may postpone this return if they undergo a sufficiently large positive earnings shock e_{it}^{d}, while others may choose to return prematurely if faced with a very negative host country earnings shock. Hence, $E\left[e_{it_2}^{d} | ysm_{it_2}, s_{it_2} = 1\right] > E\left[e_{it_1}^{d} | ysm_{it_1}, s_{it_1} = 1\right]$, and the selection of returnees on time-variant unobservables is negative. Again from equation (3), the earnings growth observed for immigrants who choose to stay becomes

$$E\left[y_{it_2}^{d} | ysm_{it_2}, s_{it_2} = 1\right] - E\left[y_{it_1}^{d} | ysm_{it_1}, s_{it_1} = 1\right]$$

$$= \gamma + E\left[\mu_i | ysm_{it_2}, s_{it_2} = 1\right] - E\left[\mu_i | ysm_{it_1}, s_{it_1} = 1\right]$$

$$+ E\left[e_{it_2}^{d} | ysm_{it_2}, s_{it_2} = 1\right] - E\left[e_{it_1}^{d} | ysm_{it_1}, s_{it_1} = 1\right].$$

Under this condition, positive (negative) shocks e_{it}^{d} to earnings will prolong (shorten) a stay in the destination country so that $E\left[e_{it_2}^{d} | ysm_{it_2}, s_{it_2} = 1\right] - E\left[e_{it_1}^{d} | ysm_{it_1}, s_{it_1} = 1\right]$ will bias estimates of γ upward both in a regression on levels or on a differenced earnings equation.

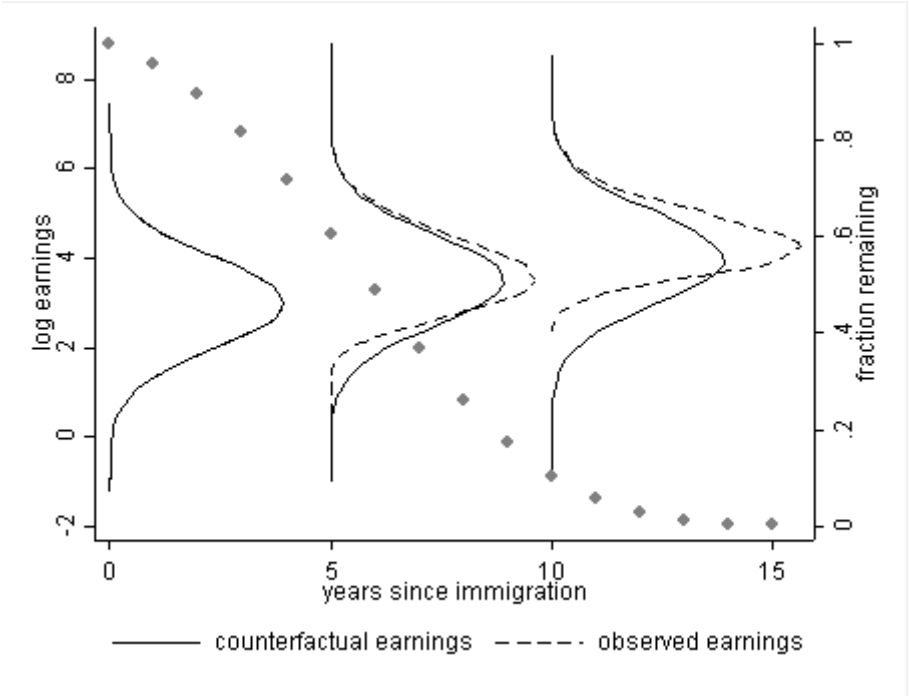
The above selection patterns, derived from a simple life cycle model, illustrate that selection on different unobserved earnings determinants can work in opposite directions, with an ambiguous overall effect. This ambiguity implies that – if return decisions are based on earnings realizations and not earnings expectations – a comparison of estimates obtained by an OLS estimator versus a fixed effects estimator of earnings slopes, or from unrestricted

pooled cross-sectional data versus stock-sampled data will be informative about selection on time-constant individual wage effects only and not about selection on time-variant components. Moreover, whereas the earnings growth of an initial immigrant cohort from which nobody out-migrated, γ, can be recovered in the former case by differencing out individual fixed effects or using stock-sampled data, a correction for selection on non-constant unobservables requires a more structural approach and explicit modelling of the selection process.

3.2 Numerical Example

We illustrate the possible biases from selective out-migration in estimators commonly employed in the literature by using the simple structural model outlined above to simulate immigrant earnings paths under the two scenarios. First, in Figure 4, we illustrate a case where, when the return decision is based on realized earnings, a sample of immigrants remaining in the country becomes increasingly positively selected as a disproportionately large share leaves from the lower part of the cohort's earnings distribution. The solid curves graph the log earnings densities of the original immigrant cohort in the counterfactual situation that nobody out-migrated (left scale). The dashed lines represent the actual observed densities of the selected remaining sample, which shift increasingly upwards since they contain immigrants with on average higher earnings realizations. The figure also shows the fraction of the original immigrant cohort remaining in the destination country as time passes (right scale, gray dots).

Figure 4: Selection and Out-Migration



Notes: This figure, based on a simulated sample of 100,000 individuals, illustrates the scenario in which return migration is affected by time-variant earnings determinants (Section 3.1.2). Here, $\alpha_0^{\mathrm{d}} = 3$, $\alpha_0^{\mathrm{o}} = \alpha_0^{\mathrm{d}}/3$, $\alpha_1^{\mathrm{d}} = 0.1$, $\alpha_1^{\mathrm{o}} = 3\alpha_1^{\mathrm{d}}$, $\theta^{\mathrm{d}} = \theta^{\mathrm{o}} = 1$, $\mu_{\mathrm{S}} = 0$, $\sigma_{\mathrm{S}}^2 = 3$, and $\sigma_{\mathrm{d}}^2 = \sigma_{\mathrm{o}}^2 = 1$.

We then use different estimators to calculate immigrant earnings equations using data created under the different scenarios while assuming that the true log earnings growth for each year of residence in the destination country is 0.1. Table 1 list estimates for a simulated sample of 100,000 individuals, and illustrates the bias of estimators commonly used in the literature.[8] In panel (a), in which return migration decisions are based on expected earnings, the OLS estimator underestimates the growth of earnings by more than 15 percent, the result of the positively selective out-migration discussed above. On the other hand, when the estimation is based on stock-sampled data, we obtain the true earnings growth parameter, though we underestimate the initial earnings position of the entire arrival cohort. Nevertheless, the

[8] For this simulation, we set T = 20, $\alpha_0^{\mathrm{d}} = 3$, $\alpha_0^o = \alpha_0^d/3$, $\alpha_1^d = 0.1$, $\alpha_1^o = 3\alpha_1^d$, $\theta^d = \theta^o = 1$, $\mu_S = 0$, $\sigma_S^2 = 3$, and $\sigma_d^2 = \sigma_o^2 = 1$.

estimated parameter $\bar{\mu}$ does measure the initial earnings position of those immigrants that remain in the destination country for at least10 years, whose earnings are about 0.14 log points lower than those of the overall arrival cohort because return migration is positively selective.

		(a) Selection on time-constant effects			(b) Selection on time-variant effects		
	True parameter values	OLS, all observation	OLS, stayed for 10 or more years	FE estimates	OLS, all observations	OLS, stayed for 10 or more years	FE estimates
$\bar{\mu}$	3	3.0557	2.8624	-	3.0303	2.7524	-
		(0.0022)	(0.0031)		(0.0023)	(0.0062)	
γ	0.1	0.0838	0.0998	0.0999	0.1401	0.1632	0.1646
		(0.0003)	(0.0005)	(0.0004)	(0.0004)	(0.0010)	(0.0005)

Table 1: Selection Bias in Standard Estimators

Column 3 of Table 1 lists the results based on the FE estimator, which yields outcomes very close to the true earnings growth parameter. Of course, in this case the initial level of log earnings is not identified directly from the regression, although it could be backed out as the mean of the predicted fixed effects. Thus, if return migration is selective, and if this selection is based only on individual-specific productivity (as would be the case in a structural model of the type above and when return decisions are based on expected earnings), the availability of both panel or longitudinal data and stock-sampled data allows to obtain unbiased and consistent estimates of the earnings growth parameter. Comparing these latter with the estimates obtained by simple OLS then allows identification of the selective out-migration's direction.

Panel (b) of the table illustrates the case in which the return decision is based on realized earnings, meaning that selection depends not only on individual constant unobservables but also on time-variant shocks to earnings that are correlated with the out-migration decision. In this case, both OLS estimation of a stock sample and FE estimation produce an overestimate of the true wage growth parameter. In this scenario, therefore, the positive selection resulting from the early achievement of earnings parity by highly productive immigrants is overcompensated by the negative selection induced by positive earnings shocks to host country earnings, which lead individuals to remain longer in the destination country. In this scenario, as the table makes clear, none of the estimators produce an unbiased estimate of earnings growth and the direction of selection cannot be inferred through comparison of either of these estimators with the OLS estimator. Rather, consistent estimation of the growth parameter requires that the selection process be specified. Estimators that correct for sample selection can then be used to recover this structural parameter. A precondition is that longitudinal data are available that contain information on leavers prior to their departure. Identification further requires that some of this information can be used to predict future out-migration, while being unrelated to time-variant unobserved wage determinants at the time the out-migration decision is made (see Dustmann and Gorlach 2015 for more detail) .

4. Out-migration Choice and Other Decisions

Until now, we have not considered the possible effects of the decision whether and when to leave the country on other migrant decisions. In our simple example, skills accumulate automatically and any time-variant unobserved earnings determinants are treated as exogenous to the migration choice. In this setting, assumptions about how individuals figure stochastic shocks to earnings into the return decision have important implications for estimator choice. However, the assumption that the decisions of individuals whether and

when to leave the host country do not affect decisions that determine outcomes one may be interested in modelling, such as earnings, consumption or savings, is very strong. If, for instance, migrants can choose whether and how much to invest in human capital, a model might either allow for labour supply choices – meaning that individuals affect the build-up of their earnings potential by accumulating experience (cf. Eckstein and Wolpin 1989) – or permit individuals to actively invest in human capital (cf. Ben Porath 1967). In such models, both the optimal return time and the individual's investment decisions and labour supply choices are interdependent. A longer expected period in the host country leads to more investment in host country-specific human capital and thus steeper earnings profiles.

Adding in individual heterogeneity results in complex relations between decisions that influence the earnings paths and out-migration decision, which in turn affects selection. In such cases, the estimation of immigrant earnings profiles requires a more structural approach that specifies how these choices are determined by the fraction of a lifespan that immigrants expect to spend in the destination country (see Dustmann and Görlach, 2016, for a possible framework). Such an approach should formulate not only the earnings function but also a selection process that specifies how the determinants of time-variant earnings affect location choices (see Adda, Dustmann and Görlach, 2016, for a discussion of how to model earnings profiles when individuals make choices about human capital investment and return migration).

5. Discussion and Conclusions

This re-examination of how to assess immigrant earnings profiles emphasizes that, whereas straightforward Mincer earnings equations are suitable for worker populations whose composition does not change systematically or selectively (other than through non-

employment or non-participation), an original arrival cohort of migrants is likely to change through out-migration. If such out-migration is selective, it may change the distribution of earnings of the remaining migrant population, so that mean or quantile estimators will provide biased estimates of wage growth for (percentiles of) the original arrival cohort. Admittedly, in many instances, analysts may be interested in the evolution of moments of the *remaining* immigrant population rather than in the counterfactual distribution of the original arrival cohort. Yet when interest is focussed on the latter, these counterfactual distributions must be re-constructed. Micro-econometrics offers an extensive tool set for handling such re-construction, as discussed in more detail in Dustmann and Görlach (2015).

In the simplest possible case, when selective out-migration occurs along the distribution of unobserved but constant within-individual productivity, difference or fixed effects estimators can be used to generate unbiased estimates of the earnings growth of the original arrival cohort as long as longitudinal data are available. Stock-sampled data (representative of a population that has survived until a particular point in time) can also be useful for addressing this problem even when only cross-sectional. When selective out-migration results from unobserved and time-variant shocks, however, these estimators cannot produce unbiased earnings growth estimates, and re-constructing the counterfactual earnings distributions requires modelling of the selection process. This involves the additional identification requirement of finding variables that determine out-migration but not earnings growth, which may be challenging because individuals who out-migrate drop out of the sample. This requirement makes the econometric problem similar to the panel attrition problem but different from the problem of simultaneously estimating earning profiles and non-participation, as found in the female labour supply literature.

In the second part of this paper, we illustrate the estimation challenge by setting up a very simple decision model in which individuals consider the return decision in each period based

on future earnings. In this simple structure, wages grow mechanically with time on the labour market and the price of human capital is higher in the host country; however, the return to skills is higher in the immigration country, which may trigger a return once the skill stock (accumulated with each year in the host country) is sufficiently high.

We study the implications of this simple model under two scenarios: In the first, the return decision is made *before* the period-specific random shock to earnings is observed and is based on *expected* earnings in the home and host country. In the second, the decision is made *after* the period-specific shock is known and is based on the *realized* earnings in the respective period. We demonstrate that in the first case, unbiased estimates of wage growth are obtainable using a simple difference estimator; however, in the second, this estimator does not produce unbiased estimates of the underlying growth parameter. Rather, the dependence of out-migration on time-variant shocks requires explicit modelling of the selection process. We further demonstrate that, when estimates are based on real data in a reduced form context, their interpretation greatly depends on the assumptions made about the underlying decision process that determines out-migration.

As in an overwhelming share of the selective out-migration literature, our basic structure assumes that the process that determines out-migration is unrelated to other choices that affect wage growth, such as human capital investment or labour supply decisions, which greatly simplifies the analysis. We follow this approach in our simulated example by assuming that wages grow exogenously, i.e. we do not allow for human capital investment. If the research design does allow for active human capital investment or labour supply choices, the analysis becomes more complex because the decision of when to out-migrate also affects wage growth. Thus, the problem becomes inherently dynamic and requires a more structural approach to estimation, such as the dynamic framework proposed by Dustmann and Görlach (2016) and applied by Adda, Dustmann and Görlach (2016).

References

Abramitzky, Ran, Leah Platt Boustan, and Katherine Eriksson. "A Nation of Immigrants: Assimilation and Economic Outcomes in the Age of Mass Migration." *Journal of Political Economy* 122.3 (2014): 467-506.

Adda, Jérôme, Christian Dustmann, and Joseph-Simon Görlach. "The Dynamics of Return Migration, Human Capital Accumulation, and Wage Assimilation." Unpublished manuscript (2016).

Barth, Erling, Bernt Bratsberg, and Oddbjørn Raaum. "Identifying Earnings Assimilation of Immigrants under Changing Macroeconomic Conditions." *Scandinavian Journal of Economics* 106.1 (2004): 1-22.

Ben-Porath, Yoram. "The Production of Human Capital and the Life Cycle of Earnings." *Journal of Political Economy* (1967): 352-365.

Bhaskar, Renuka, Belkinés Arenas-Germosén, and Christopher Dick. "Demographic Analysis 2010: Sensitivity Analysis of the Foreign-Born Migration Component." U.S. Census Bureau Population Division Working Paper no. 98, (2013).

Bijwaard, Govert E., and Jackline Wahba. "Do high-income or low-income immigrants leave faster?" *Journal of Development Economics* 108 (2014): 54-68.

Borjas, George J. "Immigrant and Emigrant Earnings: A Longitudinal Study." *Economic Inquiry* 27.1 (1989): 21-37.

Borjas, George J. "Assimilation, Changes in Cohort Quality, and the Earnings of Immigrants." *Journal of Labor Economics* 3.4 (1985): 463-489.

Bratsberg, Bernt, Erling Barth, and Oddbjørn Raaum. “Local Unemployment and the Relative Wages of Immigrants: Evidence from the Current Population Surveys.” *Review of Economics and Statistics* 88.2 (2006): 243-263.

Carliner, Geoffrey. "Wages, Earnings and Hours of First, Second, and Third Generation American Males." *Economic Inquiry* 18, no. 1 (1980): 87-102.

Chiquiar, Daniel, and Gordon H. Hanson. “International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States.” *Journal of Political Economy* 113.2 (2005): 239-281.

Chiswick, Barry R. “The Effect of Americanization on the Earnings of Foreign-born Men.” *Journal of Political Economy* 86.5 (1978): 897-921.

Duleep, Harriet O., and Daniel J. Dowhan. “Insights from Longitudinal Data on the Earnings Growth of U.S. Foreign-Born Men.” *Demography* 39.3 (2002): 485-506.

Dustmann, Christian. “Earnings Adjustment of Temporary Migrants.” *Journal of Population Economics* 6.2 (1993): 153-168.

Dustmann, Christian, Albrecht Glitz, and Thorsten Vogel. “Employment, Wages, and the Economic Cycle: Differences between Immigrants and Natives.” *European Economic Review* 54.1 (2010): 1-17.

Dustmann, Christian, and Joseph-Simon Görlach. “Selective Outmigration and the Estimation of Immigrants’ Earnings Profiles.” In: Barry R. Chiswick and Paul W. Miller (editors), *Handbook of the Economics of International Migration* 1A, North Holland (2015): 489-533.

Dustmann, Christian, and Joseph-Simon Görlach. “The Economics of Temporary Migrations.” *Journal of Economic Literature* 54.1 (2016): 98-136.

Dustmann, Christian, and María E. Rochina-Barrachina. "Selection Correction in Panel Data Models: An Application to the Estimation of Females' Wage Equations." *Econometrics Journal* 10.2 (2007): 263-293.

Dustmann, Christian, and Yoram Weiss. "Return Migration: Theory and Empirical Eevidence from the UK." *British Journal of Industrial Relations* 45.2 (2007): 236-256.

Eckstein, Zvi, and Kenneth I. Wolpin. "Dynamic Labour Force Participation of Married Women and Endogenous Work Experience." *Review of Economic Studies* 56.3 (1989): 375-390.

Edin, Per-Anders, Robert J. LaLonde, and Olof Åslund. "Emigration of Immigrants and Measures of Immigrant Assimilation: Evidence from Sweden." *Swedish Economic Policy Review* 7.2 (2000): 163-204.

Faini, Riccardo, Steinar Strøm, Alessandra Venturini, and Claudia Villosio. "Are Foreign Migrants More Assimilated than Native Ones?" IZA Discussion Paper No. 4639 (2009).

Fertig, Michael, and Stefanie Schurer. "Earnings Assimilation of Immigrants in Germany: The Importance of Heterogeneity and Attrition Bias." SOEP Paper No. 30 (2007).

Hu, Wei-Yin. "Immigrant Earnings Assimilation: Estimates from Longitudinal Data." *American Economic Review* 90.2 (2000): 368-372.

LaLonde, Robert J., and Robert H. Topel. "The Assimilation of Immigrants in the U.S. Labor Market." In *Immigration and the Workforce: Economic Consequences for the United States and Source Areas*, pp. 67-92. University of Chicago Press, 1992.

Lindstrom, David P., and Douglas S. Massey. "Selective Emigration, Cohort Quality, and Models of Immigrant Assimilation." *Social Science Research* 23.4 (1994): 315-349.

Long, James E. “The Effect of Americanization on Earnings: Some Evidence for Women.” *Journal of Political Economy* 88.3 (1980): 620-629.

Lubotsky, Darren. “Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings.” *Journal of Political Economy* 115.5 (2007): 820-867.

Nekby, Lena. “The Emigration of Immigrants, Return vs Onward Migration: Evidence from Sweden.” *Journal of Population Economics* 19.2 (2006): 197-226.

Organisation for Economic Co-operation and Development. *International Migration Outlook 2008*. OECD Publishing, Washington, DC and Paris, 2008.

Organisation for Economic Co-operation and Development. *International Migration Outlook 2013*. OECD Publishing, Washington, DC and Paris, 2013.

Picot, Garnett, and Patrizio Piraino. “Immigrant Earnings Growth: Selection Bias or Real Progress?” *Canadian Journal of Economics/Revue canadienne d’économique* 46.4 (2013): 1510-1536).

Pischke, Jörn-Steffen. “Assimilation and the Earnings of Guestworkers in Germany”. ZEW Discussion Paper No. 92-17 (1992).

Sarvimäki, Matti. “Assimilation to a Welfare State: Labor Market Performance and Use of Social Benefits by Immigrants to Finland.” *Scandinavian Journal of Economics* 113.3 (2011): 665-688.

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.

Venturini, Alessandra, and Claudia Villosio. “Labour Market Assimilation of Foreign Workers in Italy.” *Oxford Review of Economic Policy* 24.3 (2008): 517-541.