Never Breaking the Wall

The old assimilation paradigm confronted to recent econometric techniques *

Domenico Depalo
University of Rome "Tor Vergata"

December 7, 2008

Abstract

In this paper we analyze the process of integration into the European societies of the cohorts of immigrants that reached Western Europe before the mid-1990s using a longitudinal dataset.

The relevant question of the paper concerns the assimilation of immigrants to natives in terms of earnings. However, because not all the population is employed, as recognized in most of the existing literature, in the empirical analysis we model the sample selection mechanism, without sacrificing the panel nature of our dataset, as instead always happens in the literature.

Using a OLS we would conclude that the assimilation process is strong and alive, but, as going to the FE and to an attrition bias FE consistent estimator, results are dramatically changed and we finally reject the assimilation hypothesis in earning. Where the assimilation has some importance is instead in the employment mechanism. Amongst the various robustness checks, of particular relevance is the cure for possible endogeneity between earning and length of stay in the host country, which requires the generalization to instrumental variables of available estimators.

* I would like to thank Lucia Corno, Giuseppe De Luca, Franco Peracchi, Francis Vella, Alessandra Venturini, Klaus Zimmermann and seminar participants at ESPE conference (University College, London) and EEA conference (Bocconi University, Milan). Migration economics came to my mind thanks to Riccardo Faini, who was more a father than a teacher. All errors are mine. Corresponding author: Domenico Depalo, Tor Vergata University, Viale Columbia, 2 00133 Roma, Italy. E-mail: `domenico.depalo@uniroma2.it`.

1 Introduction

Almost all Western European countries are experiencing an increasing inflow of immigrants. The statistics released by the OECD for the period 1994–2001 show that the ratio of immigrants to the total population has grown steadily in all Western European countries considered, except Belgium. This is an important theme because on the one hand these are ageing population societies, and on the other hand opinion polls from the Eurobarometers and the Global Social Survey show that the large majority of natives is opposed to a further increase in migration. European policy makers must therefore strike a delicate balance between economic needs, that would dictate a substantial increase in the number of foreign workers, and political and electoral imperatives.

One common argument against immigration is that labor market outcomes are much worse for immigrants than for natives, fostering natives' suspicion that migrants tend to ride on the welfare state and depress the overall wage level (Borjas, 1995). However it has been extensively argued that after a number of years in the host country the wage level of immigrants is as high as that of natives. In this paper we try to test this hypothesis, taking a completely new approach in the migration literature. Indeed only a fraction of the population works, which raises concerns for the sample selection, as widely recognized in previous studies. Also, nowaday many datasets follow individuals over time, so that a researcher can disentangle the catching up process from the compositional effect (Borjas, 1985) if is interested in a (conditional) comparison of earnings between natives and immigrants. This feature is less appreciated, since studies that control for the selection of the sample systematically sacrifice the panel nature of the data. There exist however relatively new and unexplored econometric techniques that allow for the sample selection while preserving the panel (e.g. Kyriazidou, 1997). To the best of my knowledge, this is the first time that a study of migration takes properly into account the selection bias, while at the same time preserving the panel nature of the data, and because results are somewhat different, the proposed innovation is "informative".

In this paper we use the European Community Household Panel (ECHP), a longitudinal household survey organized and co-ordinated by Eurostat, which covers a wide range of topics, giving comparable information across the member–states of the European Union before the 2004 enlargment (EU-15).

We focus on earnings, distinguishing between natives and immigrants, and conditioning on a set of personal characteristics and a set of labor characteristics. For migrants we also control for the time spent in the host country.

The last will be a crucial indicator that we interpret in terms of assimilation: If, all other things equal, the difference in earning is gone after migrants have spent sufficiently many years in the host country, there is assimilation in terms of earning.

The contribution to the literature of this paper is a "tour" in the assimilation paradigms literature. We start from old econometric techniques, such as Ordinary Least Square on those individuals who actually work, and find a strong assimilation process; Unfortunately, OLS is not very powerful in this case, because does not consider neither the selection bias nor the panel nature. Hence our first stop is a "longitudinal perspective" with Fixed Effect on employed individuals and find that much of the assimilation results are already gone;[1] Though the FE is certainly more appropriate than the OLS, there are other possible improvements controlling for the sample selection. Our last stop is then a more recent technique that considers at once the panel structure and sample selection, thus being consistent for the structural parameter of interest and efficient in the sense that it preserves all the features of the data. Most of the literature has emphasized the differences across models and/or countries to justify the differences in results. Differently, we take here a different perspective to show that almost anything can be said on assimilation depending exclusively on the estimator that we use. With these data we cannot ignore the selectivity of the sample and, once we do that, the assimilation process in the wage equation is at least questionable. The approach is far more informative than simply this, because where assimilation is instead important is in the probability to be employed.

We run several robustness checks, although a reason of concern is the potential endogeneity of the length of stay in the host country, which requires an appropriate technique to be developed, as we do in the last part of the paper. However it is important to stress at the outset that the paper is empirical, so we do not investigate the asymptotic. The complete set of results has a number of policy implications that balance the economic needing and the political imperatives.

The remainder of the paper is organized as follows. After a brief review of the economic assimilation literature in Section 2, we address some econometric issues (Section 3). Section 4 describes the data that we used in Section 5 where we present some summary statistics (Section 5.1) and the results of our regression analyses (Section 5.2). Section 6 is a step towards controlling for endogeneity of years since migration. Finally, Section 7 offers some conclusions.

[1] The literature taking into account the sample selection in labour market outcomes for immigrants is huge (see for example Borjas, 1987, ans Fertig and Schurer, 2007, and the references therein), but sacrifices the longitudinal nature of most dataset, hence we prefer to focus on this "novelty".

2 A Brief Review of the Economic Assimilation Literature

In this paper we focus on the assimilation process in the wage equation, where "assimilation" is the catching up of the wages of migrants with respect to those of natives as going from smaller to larger length of stay in the host country. Nonetheless, it is important to recognize that there are other dimensions, in addition to the labour market outcomes, where migrants' integration is readily amenable to measurement, in particular housing, education, and health. We leave them as topics for future research.

The initial gap between natives' and immigrants' earning is justified from local specific know-how: If some skills are not completely transferable across countries, immigrants are at an earning disadvantage when they arrive in the host country, but, as time spent there increases, immigrants acquire such local knowledge and the earning gap narrows (i.e., assimilation).

The seminal articles on economic assimilation–Chiswick (1978), Borjas (1985) and LaLonde and Topel (1992)–are all based on U.S. data and all use the information included in the U.S. Census, but reaches opposite conclusions that range from over-assimilation to under-assimilation.

Over-assimilation was attributed to the fact that migrant workers are positively selected: That is, they tend to be more entrepreneurial, more talented, and less risk averse (Chiswick, 1978). In fact, there could be a confounding factor underlying this finding. Because the assimilation process evolves over time, having a single cross-section, one cannot distinguish between pure assimilation and compositional effect.

This important point was first raised by Borjas (1985) who documented evidence of under-assimilation in the United States, due to the lower "quality" of the most recent cohorts of immigrants as a consequence of changes in immigration policy, different economic conditions in the destination country, or changes in the composition of the cohorts.

On the empirical side, this pushed the research toward the use of panel data, which are more suited for longitudinal analyses. The merit of following individuals over time is to shed light on the pure assimilation process for those that are in the sample, at the likely cost that the observed immigrants are also the most successfull (i.e. return migration is not random). As a consequence, if we estimate the earning equation on the sub sample of "survivors" without any correction, the estimates will be upward biased. Related to this point, for Germany, Fertig and Schurer (2007) notice that as long as selection is systematically linked to labor market performance, the estimated parameters are inconsistent and, once they correct for the selection process, they basically reject the existence of a concave assimilation profile. In the next section we treat this point.

For a meaningful comparison, economic assimilation is evaluated conditional on a set of characteristics, that we can conveniently divide in three areas: Individual specific, labour market conditions, political environment. Related to the human capital of the immigrants, the set of variables includes education, labour market experience, language proficiency and so on (see for example Adsera and Chiswick, 2007). One obvious criticism with these characteristics concerns their cross-country comparability. For labour market variables, not only the business cycle upon arrival in the destination country, but also the sector of employment is crucial (Rosholm, Scott, and Husted, 2000).

Finally, Pennix, Schoorl and van Praag (1994) notice that also the reasons for immigration change over time. If new waves of immigrants are political refugees or family members joining their kin, the nature of immigration changes from labor migration to residential migration, with negative effect on the economic assimilation hypothesis.

In our analysis we focus more specifically on the first area, mainly because of data availability.

3 Econometric Issues

In the previous section we quickly mentioned the sample selection problem when estimating the wage equation. Here we specifically address the issue. When we study labor market outcomes with longitudinal data, we have the virtue of following individuals over time at the cost of the attrition to the sample, which can be either monotone or non-monotone. How to deal with selection is not obvious, because in reality we face different scenarios, related to the structure of dependency between the missing data mechanism and the variables of interest, each requiring different approach.

In the most favorable case, when the response probability is independent of the outcome of interest and covariates, the missing data mechanism on the outcome variable is said to be missing completely at random (MCAR) and consitent estimates are obtained using the sub sample of individuals that do not attrite. In a less favorable case, missing mechanism depends only on some covariates (X), but not on the outcome of interest (the mechanism is said missing at random, MAR). As a consequence, after conditioning on X, the response probability is independent of the outcome of interest and, even though the subsample of missing values is not random with respect to the outcome variable, it is within subclasses of X. In this case it is still possible to obtain consistent estimates of parameters of interest simply weighting the complete case to adjust for bias (Little, 1995 and Little and Rubin, 2002). There is a third possible, and even worst scenario, when the response probability is directly affected by the outcome variable and the consistency is

not guaranteed neither using the observed outcomes nor through the weighting process. In what follows we treat this case. However there is a huge literature which focuses on non missing at random scenario in the cross-sectional case but relatively few studies on selection in panel data, thus we only consider this latter stream of the literature and refer to Heckman (1979) and Ahn and Powell (1986) for the former.

Among the few methods proposed over time in the panel data case, the most appealing are by Verbeek (1990), Verbeek and Nijman (1992), Wooldridge (1995) and Kyriazidou (1997). Only the last methodology leaves all the unobservable components completely unspecified, whereas the others assume a normal distribution for the unobservable effect and the idiosyncratic errors in the selection process. Because they are nonrobust to distributional mis-specifications, we prefer to avoid the assumption, and we focus only on the method by Kyriazidou (1997).

The object of the study is a panel data model, where the sample selection works through a binary rule, whilst the linear regression contain additive permanent unobservable individual-specific effects:

$$y_{i,t} = d_{i,t}\, y^*_{i,t} = d_{i,t}\,(x^*_{i,t}\,\beta + \alpha^*_i + \epsilon^*_{i,t}) = x_{i,t}\,\beta + \alpha_{i,t} + \epsilon_{i,t} \qquad \text{i=1,...,n; t=1,...,T} \tag{1}$$

$$d_{i,t} = 1\,(w_{i,t}\,\gamma + \eta_i + u^*_{i,t} > 0) \tag{2}$$

where $y_{i,t}$ is the outcome of interest and $d_{i,t}$ is the selection indicator, $x^*_{i,t}$ are the explanatory covariates for the outcome of interest, and $w_{i,t}$ determines the selection; Apart from the error terms, each equation is characterized by individual specific components, namely α^*_i and η_i. Also, the joint distribution $(y^*_{i,t}, x^*_{i,t})$ is observed only when $d_{i,t}$ takes value 1, whilst $(d_{i,t}, w_{i,t})$ is always observed. Furthermore, we are interested in the case when the number of individual, n, goes to infinity, while the time length, T, is fixed.

A closer look at equations 1 and 2 reveals that a linear fixed effect estimator is not appropriate to fully account for the sample selectivity, even though it is a useful tool to have a quick idea of the results, as suggested for example by Verbeek and Nijman (1992) along with other investigations, and in some very special cases is even consistent. In fact, because the α_is in Eq. 1 are a function of time-varying regressors and the unobservable individual effect of the selection equation, first differencing the data as in the linear fixed effect model, does not fully consider the pattern of individual specific components, which changes over time. To circumvent this specific problem we can think of focussing only on the sub sample of individuals observed in (at least) two different time periods, i.e. $d_{i,1} = d_{i,2} = 1$, taking the first difference of the variables in equation 1, which yields

$$E(y_{i,2} - y_{i,1}|Z_i) = (x^*_{i,2} - x^*_{i,1})\,\beta + E(\epsilon^*_{i,2} - \epsilon^*_{i,1}|Z_i)$$

where $Z \equiv \{d_{i,2} = d_{i,1} = 1, x^*_{i,j}, w_{i,j}, \alpha^*_i, \eta_i\}$, for notational convenience and from Eq. 1 $d_{i,2}\,\alpha^*_i = d_{i,1}\,\alpha^*_i \Rightarrow \alpha_{i,2} = \alpha_{i,1}$. In this procedure, the fixed effects from the main equation are purged. Although, in general $E(\epsilon^*_{i,2} - \epsilon^*_{i,1}|Z_i) \neq 0$, so that we face an omitted variable problem and the simple FE is inconsistent (i.e., the longitudinal counterpart to the original formulation in Heckman, 1979). It turns out that a better strategy is to consider equation 1 as

$$y_{i,t} = x_{i,t}\,\beta + \alpha_{i,t} + \lambda_{i,t} + v_{i,t}$$

where $\lambda_{i,t} \equiv E(\epsilon^*_{i,t}|Z_i)$ is selectivity and $v_{i,t} \equiv \epsilon_{i,t} - \lambda_{i,t}$ has expected value equal to zero by construction. This formulation makes clear that the fixed effect estimator on the sub sample of non attriters is not enough to solve the sample selection problem, unless for the very special cases when $\lambda_{i,t} = 0$ (no selectivity) or $\lambda_{i,2} = \lambda_{i,1}$ (fixed selectivity). Thus for a consistent estimate of β, both λ and α must be differenced out.

A convenient procedure is in two steps and parallels that proposed in Ahn and Powell (1993) for the cross sectional case. In the first step the sample selection mechanism is consistently estimated, and used for the construction of weights to be employed in the second step; in the second step the parameters for the outcome of interest are estimated through a weighted regression, giving more weight to observations whose selection bias is lower. A nice intuition for the importance of the weighting process is that "for an individual that is selected into the sample for two time periods, it is reasonable to assume that the magnitude of the selection effect in the main equation will be the same if the oserved variables determining selection remain constant over time. Therefore, time differencing the outcome equation will eliminate not only its unobservable individual effect but also the sample selection effect" (Kyriazidou, 1997, p 1337).

The semiparametric estimator is then a weighted estimator:

$$\hat{\beta}_n = \left[\sum_{i=1}^{n} \frac{1}{T_i - 1} \sum_{s<t} \hat{\Psi_{i,n}}(x_{i,t} - x_{i,s})'(x_{i,t} - x_{i,s})\,d_{i,t}\,d_{i,s}\right]^{-1} \times \left[\sum_{i=1}^{n} \frac{1}{T_i - 1} \sum_{s<t} \hat{\Psi_{i,n}}(x_{i,t} - x_{i,s})'(y_{i,t} - y_{i,s})\,d_{i,t}\,d_{i,s}\right] \tag{3}$$

where

$$\hat{\Psi_{in}} \equiv \frac{1}{h_n} K\left(\frac{(w_{i,t} - w_{i,s})\,\gamma}{h_n}\right) = \frac{1}{h_n} K\left(\frac{W\,\gamma}{h_n}\right)$$

and is normally distruted, with $n^{(r+1)/(2\,(r+1)+1)}\,(\hat{\beta} - \beta) \xrightarrow{d} (0, h^{-1}\,\Sigma^{-1}_{xx}\,\Sigma_{xv}\,\Sigma^{-1}_{xx})$, where r determines the order bias reducing of the weighting function (a Kernel).

Notice that, in constrast to the parametric cross-sectional case, with this semi-parametric panel-data estimator we need at least one exclusion restriction even from a technical perspective, because,

among the others, one important assumption to obtain consistent estimates is that the matrix $E(\Delta x' \Delta x|W = 0)$ is finite and nonsingular. Without exclusion restrictions, when W equals zero Δx is also zero and the matrix is singular.

An important final issue for an evaluation of our empirical conclusions concerns the relative performances of the OLS, the FE without any correction and the FE plus sample selection correction. We replicate the same simulation as in Kyriazidou (1997) and notice that in the case of positive correlation between outcome of interest and probability of being in the sample, the selection bias fixed effect consistent estimator (AB-FE) is always the best estimator compared to the OLS and FE, in terms of mean and median bias and in terms of RMSE. The OLS estimates return completely misleading results because they always overestimate the value of the true β, whereas the FE does substantially better, though not as well as the AB-FE estimator. Fixing the time period and letting the number of individual increasing, the estimator increases its performance.

These results cast legitimate doubts on results obtained with OLS estimators in panel data plus attrition. In order to evaluate the assimilation process in Western Europe during 1990s, it will be important to understand how much of the results is driven by the estimator and how much is pure assimilation. This question is answered in the rest of the paper.

4 The Data

This section describes the data set that we use (Section 4.1) and discusses our definition of immigrant (Section 4.2). In this paper we use the ECHP which gives comparable information across the member-states of the European Union before the 2004 enlargment. Since we want to compare natives to immigrants, we give our operational definition of this latter group, which is not obvious given the data that we have.

4.1 Brief Description of the ECHP

The ECHP is a multi-country longitudinal household survey based on a standardized questionnaire. The survey involves annual interviews of a representative sample of households and individuals in each country for eight years from 1994 to 2001. In the first wave (1994), a sample of almost 130,000 people aged 16+ years was interviewed in the then 12 Member States of the European Union. In this section we review only some key features of the ECHP and refer to Peracchi (2002) for additional details.

The target population of the ECHP consists of people living in private households, defined on

"common living arrangements", throughout the national territory of each country.

Within each country, the original sample of households and persons is followed over time at annual intervals unless they drop out. To reflect the demographic changes in the population and to keep the panel cross-sectionally representative of the population, new households formed by at least one sample person are added across waves. For the same reasons, when people reach age 16 are interviewed.

In fact, a critical aspect of the dataset at hand is the deterioration of representativeness over time. Because the ECHP does not employ refreshment sample, its cross-sectional representativeness tends to deteriorate over time due to both non-random sample attrition and to the presence of demographic changes arising from the arrival of new immigrants.

In an attempt to limit this drawback, the ECHP User Data Base provides a set of weights for sample households and persons, that are intended to reflect the structure of the population, taking into account the sample design and the individual characteristics. Weights for longitudinal analyses in wave t are obtained from wave $t-1$ and are adjusted for the response probability in the current period. In Section 3 we made clear that adjustment classes should be chosen so that respondents in weighting class j are a random sample of the sampled units (that is, the data are MCAR within adjustment class j). Since weights can be applied in practice only when the set of variables observed for both respondents and non respondents is limited and has predictive power for response probabilities (Little and Rubin, 2002), it is important to know which covariates are used to explain the dichotomus variable of the response probability. In the ECHP such covariates are age, gender, number of economically active individuals within the household, year of the arrival of the household in the current address, household type, main source of income and the split off of the household. Clearly, whether weights are "good" or not depends in many cases on the outcome of interest. See, for example, Vandecasteele and Debels (2007) for a (positive) judgment of the ECHP weights when the outcome of interest is poverty, and indirectly income.

4.2 Definition of Immigrant Status

Three questions in the ECHP questionnaire provide information on immigrant status for foreign individuals. The first is "previous foreign country of residence before coming to the present country", the second is "foreign country of birth", and the third is "citizenship".[2]

[2] For the first two variables three versions are available with different level of detail. There is however a trade-off between data availability (by country) and levels of detail (by area of origin). Citizenship has a rather coarse classification in only four categories (see De Giorgi and Pellizzari, 2005).

In what follows, we define foreign immigrant based on country of birth for two important reasons: First the adeherence to the international standard definition of international migration, and second the robustness to naturalisation and return migration, which affect the variables regarding citizenship and last foreign country of residence, respectively. Furthermore, there is a huge correspondence between foreign country of birth and last foreign country of residence.

Natives represent the great majority in all countries, whereas the percentage of immigrants depends on the country we consider. The fraction of immigrants in the sample ranges from a minimum of about 2% in Italy and Spain to a maximum of about 8% in Austria, Belgium, and France. In the remaining countries, they represent less than 5% of the total population. Another source of variability is the composition of immigrants by area of origin. Though in general a very high percentage of immigrants comes from other EU-15 countries, we observe a different composition depending on specific country under study, namely in France and Portugal the majority of immigrants come from Africa, in Spain from America, and in Austria from other European countries.

This is our sample, but for the relevance of the paper we would like to have a flavour of how close it is with respect to the official statistics.

We compare our data to those released by the OECD in Table 1.[3] For comparison puroposes, it is worth noting that the OECD data are somewhat heterogeneous because "data on the flows and the stock of migrants and related issues ... are derived from a wide variety of sources and the nature of these sources varies across countries. This makes the application of standardised definitions difficult and hence particular attention needs to be paid to the characteristics of the data, especially in the context of international comparisons" (OECD 2005, pag. 1).

The upper panel of the table shows the data collected by the ECHP, whereas the lower panel shows the data from the OCED. The two different sources of data are quite close for the first wave of the sample, in 1994, except possibly Ireland and Portugal. However, as going from the first to the last wave, the differences between the two data sources tend to be large, and the fraction of non-natives in the ECHP tends to be much smaller than the fraction of foreigners in the OECD data.

The reason for this phenomenon is twofold. One is the lack of refreshment samples in the ECHP, one is the differentiate rate of panel attrition between natives and immigrants. Notice that even in the most favorable case of no attrition, because of the lack of refreshment samples in the ECHP,

[3] Data have been downloaded from the OECD web site `http://www.oecd.org/topicstatsportal/0,2647,en_2825_494553_1_1_1`

the survey cannot capture the trend towards an increasing inflow of immigrants in all countries considered. The higher attrition rates for immigrants leads to a downward trend in the sample fraction of immigrants on the total population and exacerbates the problem.

Table 2 shows one-year percentage attrition rates by country, wave, and immigrant status, with attrition rates defined as the ratio of the number of people lost to the sample between waves j and $j+1$ to the number of people included in the sample in wave j. Attrition rates differ considerably, not only by country and immigrant status, but also over time. In particular, we observe the highest attrition rates in Denmark and Ireland, and the lowest attrition rates in Italy (except for the last wave), Portugal and Spain. In some countries (mainly Belgium, Ireland, and Italy), attrition rates tend actually to increase over time (for a study on attrition in the ECHP see Nicoletti and Peracchi, 2002).

We also estimated a regression for the number of years an individual is observed in the sample and, consistent with the existing literature, found that the "fidelity" to the questionnaire is higher for women and employed workers, but further increases with age, number of children and social relation, overall when asked in the local community; For immigrants is increasing also in the years since migration. Number of visits and minutes for the interview have instead a negative impact as the age when higher education was completed. These figures add some more skepticism to assimilation results where the attrition to the sample is not considered.

5 Empirical Analysis

The ultimate goal of this section is to test the economic assimilation hypothesis in Western European countries, controlling for the sample selection bias and preserving all the characteristics of the available dataset, with particular attention to the panel nature.

We first describe the data and show some relevant descriptive statistics (Section 5.1)[4] to give a rough idea of the data at hand, then we move to the core of the paper, the empirical analysis (Section 5.2).

5.1 Selection of the Sample and Descriptive Statistics

5.1.1 Sample Selection

Given the outcome of interest, we focus attention on the working age population, which we conventionally define as being in the range 20–64. The resulting sample is of 87,799 individuals (459,419

[4] Detailed descriptive statistics throughout the section are available from the author if not presented.

obs.) starting from an initial number of 113,549 (610,107 obs.). To be consistent with the definition of immigrants that we gave, we focus on a restricted number of countries, namely Austria, Belgium, Denmark, France, Ireland, Italy, Portugal and Spain. Each individual is observed from a minimum of 1 year to a maximum of eight years; Quite interesting, the frequency of individuals observed for 8 years is the highest, and droup outs after only one interview is the second highest share. This is consistent with the documented evidence on survey data.

5.1.2 Relevant Descriptive Statistics

Labour market outcomes are generally better for men than for women and for natives than for immigrants. For employment status, the employment rate differentials by gender tend to be higher than by immigrant status. At the same time, current monthly earnings are always higher for men than for women, irrespective of immigrant status. The relative difference between men and women earnings varies by country, but is generally around 20%, or 150 euros and larger for natives. Natives tend to have higher mean earnings than immigrants, but the differences are 5–10 percent, left aside Portugal. The variability of earnings (measured by either the standard deviation or the interquartile range) tends to be higher for men than for women and for immigrants than for natives.

Cross-country variability is substantial. Portugal has the lowest earnings of all countries considered, no matter the subsample of interest (men/women, natives/immigrants). In Belgium, France, and Ireland, the mean of current monthly earnings of a male worker is above 1300 euros; in Austria and Denmark it is 50–100 euros lower; while in Spain and Italy, it is about 150–200 euros lower. However, when we consider earnings, conditioning on key covariates is crucial, and therefore these unconditional statistics could also be "misleading", because some important information remain concealed. As noted by Borjas and Mincer (1976), schooling and (potential) experience alone have a sizable explanatory power. It turns out that there is a increasing returns both on education and on experience, though differentiated across countries (Table 3). Such differences greatly enrich our description and reveal that, although the education-experience profile is much different across countries, in general the return on education is higher than return on experience. Finally, for some countries, even the return on labor market experience is very different by educational level and in particular is somehow higher for the highest level of education. Clearly, in these cases the earning gap between lower and higher educational levels increases over time.

By immigrant status, fixing education and experience, earnings for immigrants are lower for each combination.

Thus relevant covariates are education and labour market experience, that are closely related to skill. Indeed on the one hand, less skilled workers tend to qualify for assistance programs, potentially increasing the total fiscal costs of immigration and, on the other hand, under certain conditions, regarding capital owners, sostituibility between native and immigrant workers and elasticity of labor demand curve, there is a positive immigration surplus, i.e. "increasing national income accruing to natives", as long as the skill composition of migrants is different from that of natives (Borjas, 1995).

In our data, people with highest education is generally the lowest share in the sample, and intermediate educational levels are observed for one third of the population in each country, left aside natives in Portugal and Spain.

Some differences emerge in the comparison of ability of countries to attract highly skilled immigrants: Belgium, Denmark and Spain are most sucessfull in attracting immigrants with high eduational grade, whereas Austria, Italy and Portugal are the least successfull. This should be added to the fact that the latter group of countries have also the lowest share of most educated natives. From this perspective, the success of Spain should instead be emphasized because here the difference between the share of most educated individuals in natives and immigrants is the greatest. Finally least educated natives is a very high share in Italy, Portugal and Spain both in absolute terms and relative to the same category for immigrants.

As largely expected, men have more labor market experience than women and, in general, natives than immigrants. Important exceptions are Belgium and France where the opposite holds, a mirror of the older age of migrants with respect to natives in both countries.

To conclude the whole section it is of interest that once the education-experience profile is considered, there is still an earning-gap between immigrants and natives. Is it attributable to transferability of skills? If so, the longer is the period spent in a country, the higher is the observed earnings, otherwise the assimilation in the earning process would be more a myth than a truth. We try to answer this question in the following section.

5.2 Empirical Analysis

This section presents empirical results for our analysis of earnings. We carry on three different analyses, namely the OLS on those individuals who are employed, the Fixed Effects estimator, again for employed individuals and finally the Attrition Bias consistent estimator. To the best of my knowledge, this is the first time that attrition bias is properly taken into account when studying

migration and earnings, without sacrificing the panel nature of the data. Though many studies recognize the sample selectivity in the labour market, which might be even more severe in studies of migration, none of these preserve the panel nature of many available dataset. Among papers on selection, see Heckman (1979) and subsequent literature for labor market studies, Borjas (1987) and Fertig and Schurer (2007) for studies on migration. On the other hand, some authors preserve the panel structure of the dataset, but in general has been done more as a robustness check than as an independent approach (see Adsera and Chiswick, 2007, for example) and always the huge cost has been the proper treatment of the sample selection. One likely reason is the belief that "FE eliminates most forms of unobserved heterogeneity. Although certain form of selection bias are elimiated by FE estimators, other forms of selection bias and heterogeneity will not be eliminated" (Vella, 1998, p 156). It must be emphasized that here we do not face this trade-off between panel and selection.

The object of the analysis is the economic assimilation in terms of earnings. Assimilation is achieved when migrants reach the same wage level of natives, after they have spent sufficiently many years in the host country. We stick to this definition and, as a measure of earnings, we use log monthly earnings, which has some advantages over other competitors like hourly wage (see Borjas and Mincer, 1976).[5] We stick to the existing literature also for conditioning covariates. For all regressions the overall explicative power of our models measured from the R^2 is quite poor. The main reason is that we pool all the countries without any control for country fixed effect, which can not be estimated in the FE models. Adding such country fixed effect enormously increases the goodness of fit but adds nothing to our qualitative conclusions, whereas complicates the comparison across methodologies. The alternative of one regression for each country is instead confronted to the problem of small within group variation in some FE cases for immigrants and thus is discarded. Finer specifications are available from the author upon request.

5.2.1 Simple Least Square Estimation

Our first analysis is a weighted linear regression for the conditional mean of log monthly earnings on employed individuals. It should be borne in mind that the estimated $\hat{\beta}$ is consistent for β only if the missing mechanism depends on some covariates X but no on Y. The reference model is

$$E(Y|X) = \alpha + \beta X, \tag{4}$$

[5] Self-employed are sometimes excluded from such analyses, which we did as a robustness check, with no difference with respect to our final results.

where X is the set of covariates which contains the years of labor market experience and its square, dummies for schooling attainments, a dummy for not having a spouse, the number of children below age 6. When we focus on migrants, we also control for the length of stay in the host country throughout a polynomial of second degree. In this case we constrained the coefficients on labor market experience to be equal across immigrant status; This constraint is useful for a better comparability across estimation methods, because nor the simple FE nor the attrition bias consistent estimator are able to identify simultaneously the coefficients for length of stay and labor market experience. Below we treat this point more extensively. The intercept of each model throughout the section corresponds to the mean of log monthly earnings for the reference person: An individual with 20 years of labor market experience, basic education only, with a spouse but not children. For the model fitted to the sub samples of immigrants, the intercept refers to 10 years since migration. In all the cases, significance levels are based on estimated asymptotic standard errors that are robust to heteroskedasticity and to clustering arising from the panel structure of the data.

Moreover we run a regression where control for the year and the country where interviews took place, through a set of dummies, but since in general the results of main covariates do not change, we prefer this simpler specification which has the virtue of a easier comparability across different methodologies, at the cost of a much smaller explained variance.[6]

From Table 4 the raw difference in earnings, by gender, amounts to about 40% (about 30% for natives and 50% for non EU-15 immigrants) which is higher than the difference by immigrant status (aside from those from EU-15 countries, the difference is about 10% for men and 25% for women). We estimated a inverse U-shape profile for labor market experience, with a different profile by gender because the ceiling is achieved faster, but at a lower level, for women. Not having a spouse impacts negatively on native men's earnings, but for women the effect of the marital status is quite different whether we consider natives, immigrants from EU-15 countries or immigrants from non EU-15 countries and in particular the first (native) loose, the second (EU-15) are unaffected and the last (non EU-15) are better off from not having a spouse. The effect of the highest educational level is stronger for women (return on education is 30% higher for women in the sample of immigrants) and for immigrants, even though for men the differences are quite small across immigrant status, whereas intermediate levels have low explicative power, at least for immigrants and, quite strange, negative for natives. In fact, controlling for years and country of the interview, this result vanishes,

[6] The R^2 when controlling for dummies is always above 80%, comparable with other studies on the same dataset (for example Adsera and Chiswick, 2007).

which is the only important qualitative difference when controlling for these fixed effects. The number of children depresses the wage levels and the impact is higher in the sample of women and immigrants. Finally, in general, full year, full time and self-employed workers have higher earnings than otherwise similar workers.

The assimilation hypothesis is evaluated thanks to a second degree polynomial in Years-Since-Migration. Except for women from EU-15, as the time spent in the host country increases, the earning increases as well and after about 15–20 years the difference in earnings between natives and immigrants is gone, coherently with a bunch of studies such as Chiswick (1978) and Adsera and Chiswick (2007), among others. More precisely: There is evidence that the assimilation process for immigrants is strong and alive.

5.2.2 Panel Data Fixed Effects Estimation

When we move to fixed effect estimator in Table 5 we estimate

$$E(D\,Y|X) = D\,X\beta, \tag{5}$$

where D is a "de-mean" operator. We adopt the very same model that we used for the WLS analysis. The estimated $\hat{\beta}$ will be consistent for β even if the missing mechanism is non-random (i.e.,$\lambda_{i,t} = 0$), as long as selectivity is fixed over time ($\lambda_{i,1} = \lambda_{i,2}$ in Section 3). For this reason the FE estimator is strongly suggested at least as an exploratory exercise.

When we study the assimilation process in the sample of migrants we need to impose a restriction either on the labor market experience profile, or on the assimilation process itself. This is so because we have a polyomial of second degree in labor market experience and Years Since Migration but cannot simultaneously estimate their linear terms. Assume that we observe all individuals each year, so that we have no other problem except on covariates; In this case, we would have the two linear terms taking only value 1, which is the increment of those variables as going from one period to the next, i.e. there is multicollinearity by construction. Since the object of the study is the assimilation process, and since the labor market experience profile can be constrained on the basis of the sample of natives, we prefer to impose that both natives and immigrants have the same gain from job experience. This assumption cannot be avoid, and the economic justification is on two different grounds: One is that gains in productivity from job experience (translated in higher earning) should be equal across immigrant status, all other things equal and if there is no discrimination; Second, the existing estimates on the same subject point toward a very small, if any, difference across different immigrant status related to the experience profile, from a practical perspective (for the

same data, Adsera and Chiswick (2007) estimate similar numbers across immigrants status and gender).[7]

The results are somehow different from those obtained from the WLS regression on the subsample of workers and, at least for immgrants, many covariates that were strongly significant before have low explicative power now. The experience profile is much more concave than in the previous case, not having a spouse has a negative impact only in the sample of native men, but not in the other sub samples of interest and schooling has a positive impact only when the highest educational level is achieved, while going from the lowest to the intermediate level significantly increases earnings only for natives. Also, differently from the previous case, the effects are stronger for men than for women.

For immigrants, the assimilation results for those from EU-15 countries is completely gone once we account for the individual specific effect, whereas for those from non EU-15 countries we estimated a different concavity. The chatching up phenomenon is now at least questionable for women, and weaker for men.

Because one usual justification for FE without correction is that most of the selection bias is implicitly controlled for in the individual specific effect, it is certainly true that individual unobservable components determining selection explain much of what the WLS improperly estimates as assimilation. What is interesting now is whether something more can be added to this fairly pessimistic view. One promising development is introducing the mechanism that determines the selection rule, as we do in the next part.

5.2.3 Selection in the Panel Data Context

In Table 6 we corrected for the sample selection bias while preserving, at the same time, the panel nature of the data. The reference model is Eq. 1 and 2, where the selection mechanism determines whether the individual is employed or not, and, conditional on being employed in at least two waves, we test the assimilation hypothesis for earning. Here the discussion proceeds on two different levels, namely the selection equation and the equation of interest.

With respect to the selection mechanism, it is useful a brief discussion on the variables that we used, because for a correct identification of the structural parameter of interest is required an exclusion restriction (Section 3). For exclusion restriction we exploit various modules of the

[7] Instead of a polynomial of second degree, we could have a set of dummies. However the ECHP run only 8 years, and therefore the share of individuals that "move across dummy categories" of length of stay, thus providing useful information for the estimation, would be small and thus uninformative, unless each dummy is of very short length. In the limit this would be the same of having a polynomial of (at least) second degree, which therefore is preferred.

whole questionnaire: One is about social relations, one is about household characteristics, one is about labour market conditions. For social relations three questions are the membership to any club, whether the individual talk to neighbours and whether the individual meets friends or relatives outside household; for household characteristics we control for the household size and the number of active workers in the household; for labour market conditions we consider the amount of unemployment benefits. To be good instruments, these variables should affect the employment status, but not the wage level. We think that socialization and household characteristics satisfy this condition, only after controlling for (a proxy of) reservation wage. Indeed, the solution for the maximization problem for an entrepreneur who faces a perfect competition market and a two inputs production function (labor and other goods), is using the factor of productivity as much as its marginal productivity equals its price. Though this simple framework can be extended in several directions, the result about the price (wage)/productivity relation holds. Therefore, a way to reject our exclusion restrictions should be that individuals change their labor market productivity through their socialization. Even though to some extent there is certainly a positive spillover from socialization, this spillover should be higher when individuals have similar jobs but in general is hard to think of a doctor improving her productivity because she has a social relation with a carpenter: Here we do not know the personal characteristics of the people with whom individuals socialize and indeed the question refers to social relations with people from any background. From this viewpoint, socialization (mostly through talking) and networks are a strong channel to come in the labor market (searching process), but not for improving labor market productivity, aside from few exceptions. At the same time it could well be that socialization is a way to increase the finding rate of jobs and eventually it would increase the reservation wage. Hence, to make socialization a credible exclusion restriction, it is important to control for unemployment benefit as a raw proxy of reservation wage, under the rationale that the higher the benefit, the higher the reservation wage, the lower the probability to be employed. Unemployment benefits are a lower bound for wages, so it is worth stressing that they are important for the interpretation of socialization rather than as exclusion restriction itself. Social capital is an important individual good, but probably in many cases is only one dimension of a more complicated decision process that in fact involves also collective decisions. Hence, as one more recent stream of the literature considers the decisions at household level rather than at individual level, we control for the household size and the number of active workers in the household. Heuristically, we also run a regression with these covariates added in the main equation, but they are largely insignificant.

In the equation of interest, Eq. 1 in Section 3, we control for the same set of covariates that we used so far with other methods. Results are presented in Table 6; We first discuss the selection equation, then the earning equation.

For employment condition, we estimate a inverse U-shape profile for age, which is steeper for men than for women. Not having a spouse increases the employment probability for women, particularly if from EU, but has negative (for natives) or no effect for men. Educational attainment has a positive and strong impact for natives and non EU-15 women, but only when the highest schooling degree is achieved, because as going from the lowest to the intermediate degree there is no difference (except for native women). Finally, having a higher number of children decreases the probability of being employed, but only in the sample of women, with the highest impact in the sample of EU-15 immigrants.

It is worth emphasizing that, the assimilation process in the employment status is effective, except in the sample of non EU-15 men, for whom the age is more important than length of stay.

For later purposes, notice that many of the covariates used for the earning process in other methods are significant and go in the same direction for employment probability with this approach.

Before the wage equation is studied, we should spend some words on the variables that we used as exclusion restrictions. In general they conform to prior expectations. It is clear that talking to neighbours has a positive impact on the probability of going from non employment to the employment status. The effect is clear in the sample of natives, men and women. For immigrants the variable is significant in the sample of women in the EU-15 immigrants and men in the sample of non EU-15 immigrants. Meeting people is significantly positive, at 10% confidence level, in the sample of native women and, surprisingly, negative for EU-15 women. The membership to clubs has no effect on the probability of changing the labor status. The tests for the joint significance of covariates suggest that the complete set of covariates has a low explicative power for immigrants, as we reject the null hypothesis of parameters jointly equal to zero only for the sample of women coming from a EU-15 country. However we emphasize that we need only one exclusion restriction to obtain consistent estimates, thus we can safely rely on the socialization in the local community that has the desired effect and is generally significant. We take this as evidence that some type of social relations help finding a job, even after controlling for (a proxy of) reservation wage. Moreover, the differentiated patterns by immigrant status suggests a line of research that accounts for the role of network, but unfortunately we can not go any further with these data.

To individual characteristics we add a set of household characteristics that are exogenous with

respect to wage level. It turns out that these explain the employment process only for native women, for whom we etimated the employment probability to be negatively related with household size but positively with number of employed individuals. In general, though, household characteristics have low explicative power, most probably because their variation over time is small.

When we focus on the outcome of interest, the model we have estimated so far is in general unable to give a complete picture of the earning process. Bearing this in mind, we find that wage increases with experience, with point estimate not very different across gender and quite close to the FE estimator. Not having a spouse has a negative impact for native women, otherwise, as in the case of FE estimator, seems not to be very important to explain the earning equation. Fulltime, fullyear and self-employed workers have finally higher earnings than similar workers. For immigrants what makes a difference is the self-employment status of EU-15 women and non EU-15 men.

Most important, the assimilation hypothesis in the earning process is not significant nor jointly, nor by single variable, thus no immigrant group's earning overcomes natives' earning in the long run and for sure not in 15–20 years since migration as in the "OLS perspective". More precisely, the assimilation process has some positive effect in the employment status but not so in the earning process.

This is a very important distinction, because contrarily to the pessimistic view from the FE approach here we reconcile the two opposite results achieved in the literature by identifying a slightly different form of assimilation, namely in the employment process but not in the wage equation.

Because over the time considered the schooling attainment is fix for many individuals, we also estimated the same regression by educational level at first wave. Results are confirmed in this robustness check. Interesting differences are found for the selection mechanism, where immigrant women with the highest educational grade have a lower employment probability if they do not have a spouse but are less affected from the number of children than women with lower educational grade; For the wage equation there is virtually no difference, except the linear term in the experience profile. Another important check was interaction of years since migration and cohort of arrival to check for the difference quality of immigrant (Borjas, 1985), but in general there are no relevant difference with respect to the qualitative results just explained.[8]

With these results at hand, it is interesting to notice that what really changes with respect to

[8] Other less interesting robustness check that left results unchanged were about the pooled natives-immigrants sample and country specific analysis.

FE model is the significance levels of point estimates, not the point estimates themselves. The explanation for this finding is addressed by Heckman (1979, p 155), that "a symptom of selection bias is that variable that do not belong in the true structural equation may appear to be statistically significant determinants of Y when regressions are fit on the selected samples", which is exactly the case here, as noted above.

6 Is Years Since Migration Really Exogenous?

The preceding discussion is evidence against assimilation in the wage equation, obtained using standard model specification but fixing major problems with established practice. This was necessary for comparison purposes, but it is possible to do better. Indeed, estimated coefficients are consistent as long as the covariates are truly exogenous, but if endogeneity comes in consistency breaks down.

In a study of migration it must be considered the case that YSM is endogenous with respect to earnings: It is so, for example, if immigrants have preference for consumption in the home country rather than in the host country (Hill, 1987 and Dustmann, 2006). The existing literature provides determinants of time spent in the host country, that unaffect wages, i.e. valid instruments, like for example cost of travelling or intention to return at the time of arrival.

In this section we investigate how endogeneity affects our conclusions and re-estimated the model. The first natural question is about the importance of endogeneity bias in a panel sample selection estimation. From Eq. 3 we can define $\Psi_{i,n}^{1/2}\left(X_{i,t}-X_{i,s}\right)=\tilde{X}$ and $\Psi_{i,n}^{1/2}\left(Y_{i,t}-Y_{i,s}\right)=\tilde{Y}$, so that the weighted least square estimator is

$$\hat{\beta}_{IV}=(\tilde{X}'\tilde{X})^{-1}(\tilde{X}'\tilde{X})\,\beta+(\tilde{X}'\tilde{X})^{-1}(\tilde{X}'U) \tag{6}$$

which is biased if $\tilde{X}'U\neq 0$, as in the classical least square estimator. If an approapriate set of instruments exists, a consistent estimator is give by

$$\hat{\beta}_{IV}=(\tilde{X}'\,H_Z\,\tilde{X})^{-1}(\tilde{X}'\,H_Z\,\tilde{Y})$$

where H_Z is the projection matrix from first stage of instrument Z and other covariates on endogenous variable X, and standard errors must be accomodated for first stage regression.

This new estimator generalizes the methodology in Kyriazidou (1997), so it is instructive to understand its performances, although the asymptotic theory is beyond the scope of the paper. In Table 7 we repeat the same Montecarlo simulation run by Kyriazidou for 8 periods (the ECHP

length), except that the covariance $cov(Y, X) \neq 0$. As expected from the theoretical bias, if X is truly exogenous, a weighted least square performs better, but even for moderate endogeneity its bias is substantial no matter how large is the sample size. The IV approach is very good although precision increases with the number of individuals.[9]

In our study, for the possible endogeneity of the length of stay in the host country, we instrument YSM with the number of individuals moved in the household. This is a valid instrument because is a costly investment in the host country that unveils (long-run) intentions of permanent migration: In this case the relationship would be positive. It is asked only from wave 2, so the new sample size is reduced. For the same reasons as in Section 5.2, these instruments, while important for the YSM, are not related to wage and so are valid instruments. From the economic theory the ideal instrument would be the wage differential between the two countries but the ECHP does not provide us with such information, thus we view this as the best we can do. Based on the F-statistic from first stage regression it seems that the explanatory power is rather good. To summarize the results from this stage, years since migrations depends positively on the number of individuals moved in the sample and social relations, overall talking, but negatively on meeting (it might well be that they meet people in their home countries). To save space, in Table 8 we report only wage equation and it is clear that all the coefficients are almost equal to those without endogeneity is allowed, thus a Hausman test is not very useful in this particular case.

Most important, the new results are consistent with previous conclusions that the assimilation in the wage equation is at least questionable, whereas in the probability to be employed is effective.

To sum up, the method that we use to jointly estimate the selection mechansim for the employment status and the earning process reaveals an interesting phenomenon. The WLS regression on positive earnings is consistent with an assimilation process in earning, which is already questionable as we control for individual FE charcteristics and is completely gone when the selection mechanism is properly accounted for, either with or without correction for endogeneity in the years since migration. However we would be on the wrong way, should we think that assimilation in Europe is an artifact of the economic modelling or the weak econometric methods, because as going from less to more years spent in the host country the probability of being employed workers raises fast with respect to the new arrived immigrants. This invokes a tight definition of the assimilation process, a practice rarely followed by economists.

This conclusions has also an important policy implication, because, with respect to concerns

[9] Notice that we allow for endogeneity only in the main equation, whereas the methodologies by Amemyia (1974) and Foster (1997) can accommodate for endogeneity in the first stage also.

of immigrants riding on the welfare state, governments should be worried with the short run immigration and should instead favour permanent migration in response to ageing societies, because long run immigrants would contribute to the welfare state as employed workers. For what matters the earning process, labour market experience is by far the most important determinant. Here the policy implication to avoid lower wage levels is that "high quality" migration should be favoured (this statement should be further investigated on the basis of the "quality spread" between natives and immigrants, as in Borjas, 1987, and Borjas and Bratsberg, 1996).

These implications result because we reject the over-asimilation paradigm in the earning process but, at the same time, talking of under-assimilation forget a important side of the story which is instead achieved with the methods proposed by Kyriazidou (1997) and its generalization for the employment status.

7 Conclusion

Migration in Europe is steadily increasing over time. Unfortunately there is little consensus among researchers about the assimilation process of migrants in the host country. In this paper we focus on assimilation in one important dimension of human life, the labour market, studying the catching up of earnings of migrants with respect to natives in Europe, using data from the ECHP. Though the issue has been the subject of many previous studies, the contribution to the literature of this paper is the joint modelling of the selection mechanism which determines the probability of being in the sample of employed workers and the wage equation, without sacrificing the panel structure as always happens.

We tried a step toward the solution of the dilemma of assimilation in the earning process by substantially departing from the usual literature. Though much of the literature employs a WLS to test the hypothesis thus claiming that the process is strong and alive, we used first the FE panel data model to partially correct the selection bias and then an attrition bias consistent estimator, while preserving the panel structure of our data. The results are dramatically affected by this correction and eventually we reject the assimilation in the earning. However we would be on the wrong way, should we think that assimilation in Europe is an artifact of the economic modelling, because as going from less to more years spent in the host country the probability of being in the active labor force as employed workers raises fast with respect to the new arrived immigrants. These conclusions hold even when we consider the possible endogeneity between years since migration and earning. From this perspective the assimilation of earning is a trick, but assimilation in the labor

market is partially true.

References

Ahn H. and Powell J. (1993), "Semiparmetric Estimation of Censored Selection Models with a Nonparametric Selection Mechansim", *Journal of Econometrics*, 58: 3–29.

Adsera A., Chiswick B. (2007), "Are There Gender and Country of Origin Differences in Immigrant Labor Market Outcomes across European Destinations?", *Journal of Population Economics*, 20: 495–526.

Amemiya T. (1974), "The Nonlinear Two-Stage Least-Squares Estimator", *Journal of Econometrics*, 2:105–110.

Borjas G.J. (1985), "Assimilation, Changes in Cohort Quality, and Earnings of Immigrants ", *Journal of Labor Economics*, 4: 463–489.

Borjas G.J. (1987), "Self-Selection and the Earning of Immigrants", *American Economic Review*, 77: 531–553.

Borjas G.J. (1995), "The Economic Benefits from Immigration", *The Journal of Economic Perspectives*, 9: 3–22.

Borjas G.J., Bratsberg B. (1996), "Who leaves? The Outmigration of the Foreign-Born", *The Review of Economics and Statistics*, 78: 165–176.

Borjas G.J., Mincer J. (1976), "The distribution of earnings profiles in longitudinal data", NBER working papers, No.143.

Chiswick B. (1978), "The Effect of Americanisation on the Earnings of Foreign-born Men", *Journal of Political Economy*, 86: 897–921.

Chiswick B. (1991), "Speaking, Reading, and Earnings among Low-skilled Immigrants", *Journal of Labor Economics*, 9: 149–170.

De Giorgi G., Pellizzari M. (2005), "Welfare Magnets within Europe and the Cost of Harmonized Social Assistance", unpublished.

Dustmann C. (1997), "Return Migration, Uncertainty and Precautionary Saving", *Journal of Development Economics*, 52: 295 316.

Dustmann C. (1997), "Return Migration - The European Experience", *Economic Policy*, 11: 213 250.

Fertig M. and Schurer S. (2007), "Labor Market Performance of Immigrants in Germany Revisited: The Importance of Omitted Variable and Selection Bias", *IZA Discussion Paper*, No 2915.

Foster M.E. (1997), "Instrumental Variables for Logistic Regression: An Illustration", *Social Science Research*, 26: 487-504.

Kyriazidou E. (1997), "Estimation of a Panel Data Sample Selection Model", *Econometrica*, 65: 1335–1364.

Heckman J. (1979), "Sample Selection Bias as a Specification Error", *Econometrica*, 47: 153–160.

Hill J. (1987), "Immigrant Decisions Concerning Duration of Stay and Migration Frequency", *Journal of Development Economics*, 25: 221-234.

LaLonde R., Topel R. (1992), "The Assimilation of Immigrants in the U.S. Labor Market" In Borjas G., Freeman R. (eds.), *Immigration and the Workforce*, University of Chicago Press, Chicago: 67–92.

Little R. (1985), "Modeling the Drop-Out Mechanism in Repeated-Measures Studies", *Journal of the American Statistical Association*, 90: 1112-1121.

Little R. and Rubin D. (2002), *Statistical Analysis with Missing Data*: Wiley, New York.

Nicoletti C., Peracchi F. (2002), "A Cross-Country Comparison of Survey Partecipation in the ECHP", *ISER Working Papers*, No 32.

OECD (2005), "Statistical Annex" [available at `http://www.oecd.org/dataoecd/24/29/34643131.DOC`].

Penninx R., Schoorl J., van Praag C. (1994), "The impact of International Migration on Receiving Countries: The Case of The Netherlands", Netherlands Interdisciplinary Demographic Institute (NIDI) Report n. 37, The Hague.

Peracchi F. (2002), "The European Community Household Panel: A Review", *Empirical Economics*, 27: 63–90.

Rosholm M., Scott K., Husted L. (2000), "The Times They Are A-Changin, Organisational Change and Immigrant Employment Opportunities in Scandinavia", Centre for Labour Market and Social Research (CLS) Working Paper 7, Aarhus.

Vandecasteele L. and Debels A. (2007), "Attrition in Panel Data: the Effectiveness of Weighting", *European Sociological Review*, 23: 81-97.

Vella F. (1998), "Estimating Models with Sample Selection Bias: A Survey", *Journal of Human Resources*, 33:88-126.

Verbeek M. (1990), "On the Estimation of a Fixed Effects Model with Selectivity Bias", *Economics Letters*, 34: 267–270.

Verbeek M. and Nijman T. (1992), "Testing for Selectivity Bias in Panel Data Models", *International Economic Review*, 33: 681–703.

Wooldridge J. (1995), "Selection corrections for panel data models under conditional man independence assumptions", *Journal of Econometrics*, 68: 115–132.

Table 1: Fraction of immigrants by country and wave (percentage relative frequencies).

	Year							
Country	1994	1995	1996	1997	1998	1999	2000	2001
ECHP data								
Austria	.	8.3	7.6	6.7	6.7	6.3	6.4	6.2
Belgium	9.0	8.6	8.3	8.2	7.9	7.6	7.4	6.7
Denmark	4.0	3.9	3.7	3.4	3.3	2.8	2.9	2.9
France	8.7	8.0	7.4	6.9	6.6	6.3	5.8	5.6
Ireland	4.6	4.7	4.4	4.2	4.0	3.7	3.6	3.5
Italy	1.7	1.8	1.8	1.7	1.6	1.5	1.5	1.4
Portugal	2.7	2.7	2.5	2.4	2.2	2.3	2.3	2.1
Spain	1.9	1.7	1.7	1.5	1.5	1.5	1.5	1.5
OECD data								
Austria	8.90	9.00	9.04	9.08	9.13	9.20	9.34	9.40
Belgium	9.10	9.00	8.97	8.86	8.70	8.80	8.40	8.21
Denamrk	3.80	4.20	4.70	4.70	4.80	4.88	4.80	5.00
France						5.60		
Ireland	2.69	2.69	3.20	3.13	3.00	3.20	3.30	3.90
Italy	1.60	1.70	2.00	2.10	2.10	2.17	2.43	2.36
Portugal	1.58	1.70	1.70	1.76	1.78	1.90	2.08	2.17
Spain	1.18	1.27	1.37	1.60	1.83	2.00	2.20	2.74

Table 2: One year attrition rates by year and immigrant status (percentage).

	Immigrant	Year						
Country	status	1995	1996	1997	1998	1999	2000	2001
Austria	Non imm.	.	10.7	10.0	11.6	10.2	11.8	8.1
Austria	Imm.	.	20.3	20.1	14.5	13.2	9.9	9.7
Belgium	Non imm.	9.9	9.7	11.2	11.5	11.2	10.5	13.0
Belgium	Imm.	10.6	10.6	9.4	13.3	12.2	13.1	19.1
Denmark	Non imm.	13.0	15.3	15.8	16.9	12.6	10.7	8.5
Denmark	Imm.	22.5	23.7	23.4	20.9	22.5	13.5	7.1
France	Non imm.	11.1	7.8	12.0	10.9	10.0	10.4	8.3
France	Imm.	15.3	9.4	14.6	11.7	9.5	9.8	6.5
Ireland	Non imm.	19.9	17.1	13.5	12.8	17.1	22.0	14.7
Ireland	Imm.	18.4	23.2	16.3	16.6	22.0	22.3	16.5
Italy	Non imm.	6.2	5.5	10.5	10.0	8.1	9.5	11.6
Italy	Imm.	7.6	7.4	12.5	13.9	10.3	9.0	17.4
Portugal	Non imm.	5.6	7.4	6.9	8.3	7.1	7.0	6.0
Portugal	Imm.	9.9	12.7	9.1	17.1	6.3	6.1	13.3
Spain	Non imm.	14.8	11.2	13.0	12.5	11.7	12.1	9.2
Spain	Imm.	19.2	15.8	20.8	13.5	16.0	12.6	12.4

Table 3: Descriptive statistics for current monthly earnings by country, gender and immigrant status (all waves). Various educational levels.

		Potential Experience				
Country	Educ.	(0,10]	(10,20]	(20,30]	(30,40]	40+
Natives						
Austria	U	1315.5	1605.5	1905.3	2192.3	2392.3
Austria	I	994.8	1104.9	1209.7	1322.7	1463.2
Austria	L	828.0	891.6	913.5	943.1	980.8
Belgium	U	1177.9	1421.3	1633.2	1719.3	1509.4
Belgium	I	899.0	1080.6	1268.1	1316.6	1272.4
Belgium	L	811.3	1006.0	1022.5	1146.8	927.7
Denmark	U	1221.5	1403.4	1441.7	1472.1	1414.2
Denmark	I	868.8	1218.5	1268.2	1227.1	1120.0
Denmark	L	766.5	1048.9	1112.1	1049.3	914.4
France	U	1346.1	1947.0	2376.7	2606.5	2797.6
France	I	889.1	1146.4	1336.2	1476.4	1304.1
France	L	885.1	1062.1	1152.2	1186.9	1062.7
Ireland	U	1244.7	1891.9	2125.4	2202.1	1983.6
Ireland	I	949.1	1244.4	1343.7	1457.4	1311.3
Ireland	L	837.2	1033.4	1087.0	1046.0	1098.6
Italy	U	1199.0	1497.7	1733.8	1911.5	1440.8
Italy	I	904.2	1122.8	1280.5	1369.9	1421.7
Italy	L	800.0	949.2	1014.0	1012.0	920.9
Portugal	U	1106.5	1557.6	1783.6	1927.5	2055.1
Portugal	I	626.2	897.0	1061.5	1138.8	964.7
Portugal	L	474.0	572.6	641.8	629.1	543.2
Spain	U	1105.3	1592.5	2018.0	2118.2	1984.3
Spain	I	778.5	1138.1	1424.4	1561.0	1530.1
Spain	L	722.7	872.2	989.8	1044.1	977.4
Immigrants						
Austria	U	1007.1	1198.0	1272.2	1482.0	.
Austria	I	901.4	1052.8	1101.4	1218.2	1482.5
Austria	L	817.9	856.1	964.0	953.8	669.1
Belgium	U	1272.7	1858.7	1846.5	1585.2	727.0
Belgium	I	962.1	1166.1	1211.4	1290.1	1005.1
Belgium	L	773.3	852.8	1014.2	967.6	765.5
Denmark	U	1116.6	1440.4	1334.0	1458.7	1317.2
Denmark	I	779.0	1140.7	1140.5	1124.3	620.6
Denmark	L	814.4	959.6	975.1	739.4	819.5
France	U	1574.2	1937.7	2037.9	2609.7	2759.6
France	I	812.3	1047.0	1394.0	1414.8	875.3
France	L	695.6	934.0	1011.8	1189.2	1045.4
Ireland	U	1228.3	1809.2	1996.4	1986.8	1195.4
Ireland	I	895.8	1067.2	1434.8	2199.1	644.3
Ireland	L	627.4	943.7	780.5	693.4	1207.3
Italy	U	1106.3	1272.8	1789.8	3217.5	.
Italy	I	834.1	1082.0	1114.0	1422.1	.
Italy	L	683.6	891.7	1000.3	1076.1	873.1
Portugal	U	1120.3	1326.9	1510.3	2095.4	.
Portugal	I	635.6	824.2	1205.7	2117.8	1115.2
Portugal	L	517.7	671.2	642.4	787.6	647.5
Spain	U	1007.8	1352.6	2018.5	1671.8	1670.1
Spain	I	710.2	867.4	1392.9	1904.9	590.0
Spain	L	684.1	853.0	931.6	835.2	502.8

Note: U= Upper level of education; I= Intermediate; L= Lower.

Table 4: Estimated coefficients of linear model for current monthly earnings for employed by immigrant status and gender (***, ** and * respectively denote an observed significance level below 1%, between 1 and 5% and between 5 and 10%). Ordinary Least Square.

	Natives		EU-15 imm		Non EU-15 imm.	
	Men	Women	Men	Women	Men	Women
Exper.	0.017***	0.007***				
Exper.^{2}	-0.001***	-0.001***				
Not spouse	-0.227***	-0.101***	-0.020	-0.053	0.035	0.167*
High ed.	0.497***	0.568***	0.534***	0.963***	0.624***	0.938***
Interm. ed	-0.525***	-0.627***	-0.024	0.055	0.096	0.082
Children	-0.057***	-0.068***	-0.192***	-0.263***	-0.065**	-0.066
Full-year	0.124***	0.129***	0.139***	-0.026	0.153***	0.166***
Full-time	0.357***	0.499***	0.974***	1.099***	0.096	0.354***
Self em.	0.214***	0.177***	0.855***	0.065	0.317***	0.154
YSM			-0.049***	0.010	0.021***	0.043***
YSM^2			0.001***	-0.000	-0.001***	-0.002***
Constant	11.806***	11.476***	10.345***	10.396***	11.695***	11.225***
Obs.	120078	84865	1958	1779	2660	2026
R-squared	0.063	0.083	0.077	0.140	0.046	0.106
RSS	387827	264829	5757	6325	5555	4671
Jointly significance for earning equation						
Experience	241.1***	125.6***	.	.	.	.
Education	1416.4***	1586.2***	9.0***	18.6***	17.9***	41.4***
Job related	202.4***	415.4***	10.7***	25.3***	12.6***	13.9***
YSM	.	.	7.3***	1.3	16.6***	20.7***

Table 5: Estimated coefficients of linear model for current monthly earnings for employed by immigrant status and gender (***, ** and * respectively denote an observed significance level below 1%, between 1 and 5% and between 5 and 10%). Fixed effect panel data model.

	Natives		EU-15 imm		Non EU-15 imm.	
	Men	Women	Men	Women	Men	Women
Exper.	0.066 ***	0.064 ***				
Exper.^{2}	-0.001 ***	-0.001 ***				
Not spouse	-0.038 ***	-0.012	-0.016	-0.035	-0.011	-0.053
High ed.	0.136 ***	0.110 ***	-0.014	0.187	0.327 **	0.281 **
Interm. ed	0.041 ***	0.024 **	-0.023	0.079	0.050	0.107
Children	-0.022 ***	-0.064 ***	-0.044 *	-0.108 ***	-0.001	-0.066 *
Full-year	0.051 ***	0.049 ***	0.069 *	0.046 **	0.046 ***	0.023
Full-time	0.102 ***	0.105 ***	0.110 *	0.111 **	0.053	0.128 ***
Self em.	0.111 ***	0.089 ***	0.446 **	0.817	0.357 ***	0.154 ***
YSM			0.011	0.009	0.009	0.023 **
YSM^2			-0.000	-0.000	-0.001 ***	-0.001 **
Obs.	120078	84865	1958	1779	2660	2026
R-squared	0.136	0.138	0.030	0.041	0.061	0.029
Jointly significance for earning equation						
Experience	3938.3 ***	2741.0 ***	.	.	.	.
Education	53.4 ***	27.2 ***	0.1	1.0	3.3 **	3.1 **
Job related	171.0 ***	185.8 ***	4.4 ***	4.1 ***	20.6 ***	7.6 ***
YSM	.	.	0.7	0.5	8.2 ***	3.2 **

Table 6: Estimated coefficients of bias corrected current monthly earnings by immigrant status and gender (***, ** and * respectively denote an observed significance level below 1%, between 1 and 5% and between 5 and 10%).

	Natives		EU-15 imm		Non EU-15 imm.	
	Men	Women	Men	Women	Men	Women
Employment Status						
Age	0.170***	0.153***	0.170	0.153	0.170	0.153
Age^2	-0.019***	-0.013***	-0.019	-0.013	-0.019	-0.013
Not spouse	-0.289***	0.397***	-0.194	1.223***	0.299	0.665
High ed.	0.652***	1.255***	0.773	0.275	0.844	1.838***
Interm. ed	-0.111*	0.293***	-0.689	-0.279	-0.182	0.122
Children	-0.156***	-0.688***	0.175	-0.948***	0.274	-0.603***
Club	0.054	0.020	0.211	0.078	0.430	0.341
Talk	0.323***	0.320***	0.270	0.519**	0.468**	0.272
Meet	0.056	0.065*	0.514	-0.446*	-0.106	-0.070
HH size	-0.010	-0.157***	0.258	0.048	-0.050	-0.208
Numb.Employed	-0.002	0.041***	-0.112	0.038	0.010	0.024
Unempl.benefit	-0.180***	-0.174***	-0.190***	-0.267***	-0.323***	-0.286***
YSM			0.242***	0.127**	0.024	0.143***
YSM^2			-0.004*	-0.004***	-0.002	-0.002
Log-like.	-12365	-16796	-177	-334	-290	-370
Earning equation						
Exper.	0.052***	0.056***				
$Exper.^2$	0.000	-0.000				
Not spouse	0.008	-0.079**	-0.046	-0.146	-0.077	-0.311
High ed.	-0.083	-0.060	-0.413	0.244	0.378	0.216
Interm. ed	0.023	-0.024	0.151	0.075	0.104	0.110
Children	0.003	-0.030	-0.268**	-0.092	-0.040	-0.056
Full-year	0.038***	0.041***	0.015	0.053	0.028	0.025
Full-time	0.057**	0.083***	0.090	0.020	0.130	0.069
Not self em.	0.058***	0.093***	0.132	1.469***	0.349***	0.082
YSM			-0.054	0.042	0.014	0.019
YSM^2			0.001	-0.001	-0.001	-0.001
Obs.	167785	188456	2792	3968	3945	4647
Censored obs.	120078	84865	1958	1779	2660	2026
R^2	0.013	0.003	0.001	0.004	0.053	0.013
Jointly significance for selection equation						
Age	4611.7***	2641.8***	.	.	.	.
Education	78.1***	248.7***	7.0**	1.8	3.9	10.0***
Social rel.	76.6***	86.7***	3.8	6.5*	6.5*	3.0
Household	0.4	75.9***	1.8	1.0	0.1	1.8
YSM	.	.	7.1**	7.0**	0.2	8.8**
Jointly significance for earning equation						
Experience	344.910***	382.715***	.	.	.	.
Education	4.873*	1.060	1.784	0.554	0.579	0.214
Job related	49.773***	108.023***	0.389	8.486**	8.228**	1.125
YSM	.	.	0.909	0.960	0.254	0.377

Table 7: Severity of the endogeneity bias in the panel sample selection.

Ind.	Endogeneity					
	0		0.3		0.8	
	OLS	IV	OLS	IV	OLS	IV
1000	-0.008	-0.024	-0.196	-0.025	-0.379	-0.027
1000	-0.013	-0.025	-0.200	-0.028	-0.382	-0.029
5000	-0.004	-0.013	-0.193	-0.013	-0.376	-0.014
5000	-0.003	-0.013	-0.195	-0.014	-0.377	-0.020
10000	-0.001	-0.000	-0.190	-0.000	-0.376	-0.002
10000	-0.001	0.000	-0.191	-0.001	-0.377	-0.005

Note: The enters are $1 - \hat{\beta}$; the probability is a conditional logit with cov(Y,X)=0.8. Time length is 8.

Table 8: Estimated coefficients of bias corrected current monthly earnings by immigrant status and gender (***, ** and * respectively denote an observed significance level below 1%, between 1 and 5% and between 5 and 10%).

	Natives		EU-15 imm		Non EU-15 imm.	
	Men	Women	Men	Women	Men	Women
Employment Status						
Exper.	0.053 ***	0.057 ***				
Exper.^{2}	0.000	-0.000				
Not spouse	-0.012	-0.054	0.045	-0.258	0.135	-0.312
High ed.	-0.041	-0.036	-0.157	0.195	1.015	0.395
Interm. ed	0.022	-0.013	-0.057	0.191	0.737	0.257
Children	0.001	-0.053 **	0.015	-0.023	0.161	0.007
Full-year	0.039 ***	0.041 ***	0.050	0.096	-0.002	0.032
Full-time	0.046 *	0.065 ***	0.136	0.042	0.395	-0.014
Not self em.	0.110 ***	0.086 ***	0.398	5.061 ***	0.362	0.201
YSM			0.152	0.050	-0.179	-0.113
YSM^2			-0.003	0.001	0.011	0.004
Obs.	135254	154241	2156	3157	3101	3680
Censored obs.	96716	69629	1540	1412	2107	1633
R^2	0.015	0.002	0.026	0.003	0.013	0.022