The Labour Market Integration of Immigrants and their Children

Ana Sofia Damas de Matos

Dissertation submitted for the degree of Doctor of Philosophy in Economics at The London School of Economics and Political Science

September 2012

Declaration

I certify that this thesis is solely my own work other than where I have clearly indicated that it is the work of others (in which case the extent of any work carried out jointly by me and any other person is clearly identified in it).

I warrant that this authorization does not, to the best of my belief, infringe the rights of any third party.

Ana Sofia Damas de Matos

Abstract

This thesis examine three distinct aspects of the labour market integration of immigrants and their children in the host country.

The first chapter looks at the early careers of immigrants to shed light on the mechanisms driving the immigrant wage growth in the first years in the host country. I use a unique linked employer employee panel covering all wage earners in the private sector in Portugal to follow the careers of immigrant men. I show that in the first ten years in the country immigrants close one third of the initial immigrant-native wage gap and that one third of this wage catch-up is accounted for by immigrants gaining access to better paying firms. I then suggest an economic assimilation mechanism which highlights imperfect information about immigrant productivity and show that its predictions are in line with the data.

The second chapter offers a longer term perspective of the economic assimilation of immigrants by turning to the labour market performance of the second generation. The chapter uses a unique survey of children of immigrants from Turkey, Morocco and ex-Yugoslavia, and children of natives in 15 European cities to closely compare their educational and labor market outcomes. Although the second generation performs on average worse than the children of natives in most outcomes considered, all differences are explained by differences in socio-economic background.

While the first chapter focused on the dynamics of the wage gap over time, the third chapter studies the differences in the *level* of the wage gap across immigrant populations. The chapter provides a comparison of the wage gaps by country of origin in two major host countries, the UK and the US, in order to disentangle country of origin effects from immigrant selection. I show that the wage gaps by country of origin are strongly correlated in the two host countries and that virtually all the correlation is accounted for by differences in country of origin specific returns to education.

Acknowledgements

This dissertation was written under remarkable supervision at LSE and at U.C. Berkeley. I am grateful to Alan Manning for his advice, support and trust; to David Card for the time and energy he puts into advising his students; and to Barbara Petrongolo for her advice and encouragement.

I would also like to thank Miguel Portela without whose help and generosity I could not have written the main chapter (Chapter 1). I am indebted to him for welcoming me at the Universidade do Minho to work with the data and for taking the time to run thousands of programs in the data lab in Portugal while I was in London.

The first chapter benefited from many discussions with Giuseppe Berlingieri, Timothée Carayol, Laurent Eymard, Guy Michaels, Steve Pischke, Philip Kircher, Pat Kline, Sebastian Stumpner and Robert Zymek. Comments from Ana Rute Cardoso, Henry Farber, Fabian Lange, Enrico Moretti and Krishna Pendakur were very much appreciated. Thomas Liebig and Gilles Spielvogel provided useful comments on the second chapter.

Parts of the dissertation were presented at different stages at the following seminars and conferences from which I got valuable feedback: the Labour Markets Workshop at the LSE; the Labour Lunch at UC Berkeley; the 4th Meetings of the Portuguese Economic Journal (Faro, July 2010); the "Migration: Economic Change, Social Challenge" (London, April 2011) and the "Economic Analysis using linked employer-employee data: bringing together theory and empirics" (Porto, June 2011) conferences; and the Seminar on Integration on the Labour Market of the Children of Immigrants (Brussels, October 2009).

I also thank all those who contributed to making the Centre for Economic Performance and the Institute for Research in Labour and Employment great places to work.

Part of the work for the dissertation was undertaken with financial support from the Fundação para a Ciência e a Tecnologia (Portuguese Ministry of Science) under the grant SFRH/BD/23958/2005, for which I am thankful.

I am grateful to the Portuguese Ministry of Employment, Statistics Department, for access to the data used in the first chapter and to the TIES network for access to the data for the second chapter. Among the TIES researchers, I thank in particular Rosita Fibbi and Philippe Wanner for their support with the project and Liesbeth Heering and Laurence Lessard-Phillips for their help with the data.

Contents

List of Figures

List of Tables

Preface

13% of the OECD population in 2010 was foreign-born. The literature on the economic assimilation of immigrants studies how people born abroad, and often educated abroad, adapt to the host country labour market. How do immigrants perform in the labour market compared to natives? How does the performance change as time goes by? A motivation for migrating often invoked by labour migrants is the possibility of a better life for their offspring: how do immigrants' children fare compared to natives but also to the first generation? These are some of the main questions addressed by the literature.

Each chapter of this dissertation examines a distinct aspect of the labour market integration of immigrants and their children in the host country. The first chapter looks at the early careers of immigrants in Portugal to measure the wage catch up in the first years in the country and understand the mechanisms behind the wage growth. The second chapter analysis the long-term integration of immigrants by comparing the outcomes of the children of immigrants to the natives in European cities. The third chapter compares immigrants by country of origin in two major host countries, the UK and the US, to better understand the widely documented fact that the country of origin is a main predictor of the wages of immigrants in the host country.

In the first chapter, I focus on the labour market integration of immigrant men in the first years in the host country. I use a unique linked employer employee panel covering all wage earners in the private sector in Portugal to follow the careers of immigrant men. I show that in the first ten years in the country immigrants close one third of the initial immigrant-native wage gap and that one third of this wage catch-up is accounted for by firm heterogeneity. Immigrants remain in the same occupations but gain access to jobs in better paying firms. Over time immigrants move to firms that are larger, more productive and have a higher share of native workers. These patterns are similar for all the recent immigrants irrespective of their origin and in particular of whether their mother tongue is the host country's language. Motivated by these new stylized facts, I suggest an economic assimilation mechanism which highlights imperfect information about immigrant productivity. I build an employer learning model with firm heterogeneity and complementarities between worker and firm type. The initial uncertainty over immigrants' productivity prevents them from gaining access to the best jobs. Over time, productivity is revealed and immigrants obtain better firm matches. I derive predictions on the immigrant wage distributions over time, on their mobility patterns and on the productivity distribution of the firms they are matched with. The predictions of the model are in line with the data and are not trivially derived from competing explanations.

In the second chapter, I focus on a much longer time horizon. The decision to migrate takes into account not only the welfare of the migrant himself but also that of his family. The assimilation of the second generation is sometimes considered the best indicator of immigrant assimilation. There are several reasons why immigrants may not be expected to perform as well as comparable natives in the labour market. The question of the integration of the children of immigrants in the host country labour market is a very different issue from the integration of the first generation. The children of immigrants are raised in the host country, go through the same educational system than the natives and speak the host country language.

The chapter uses a unique survey of children of immigrants from Turkey, Morocco and ex-Yugoslavia, and children of natives in 15 European cities to closely compare their educational and labor market outcomes. Although the second generation performs on average worse than the children of natives in most outcomes considered, virtually all differences are accounted for by differences in family background. Parents' characteristics are also shown to be important predictors of other outcomes besides educational and labour market outcomes, such as the intermarriage of the second generation. Children from lower socio-economic background marry more often within the ethnic group. Moreover, the estimation of a simple intergenerational model of human capital shows that in European cities there is less correlation between the outcomes of parents and their children in immigrant than in native families.

While the first chapter focuses on the dynamics of the wage gap over time and the possible mechanisms behind the patterns observed, chapter three focuses on the differences in the levels of the wage gap across immigrant populations.

The chapter provides a comparison of the wage gaps by country of origin in two major host countries, the UK and the US. It is a well established fact in the literature that the country of origin is an important predictor of the wages of immigrants in the host country. It is nevertheless not clear why this is the case. On the one hand, there are characteristics of the country of origin that may make the workers more productive in any labour market, such as the quality of the educational system. On the other hand the selection of immigrants into a given host country differs by country of origin. The aim in this chapter is to compare immigrants from the same country of origin in two host countries to disentangle the different effects. I first show that there is a strong correlation between the wages of immigrants relative to that of natives by country of origin in the two host countries. Differences in returns to years of education by country of origin account for virtually all of the observed correlation. There is nevertheless a large part of the wage differences across countries of origin that remains unexplained by the country of origin effects.

Chapter 1

The Careers of Immigrants

1.1 Introduction

Over the past thirty years, the literature on the economic assimilation of immigrants has focused on measuring the immigrant-native wage gap and the speed at which the gap closes with time spent in the host country. According to Chiswick (1978) immigrants earnings' would equal and then exceed the natives' after 10 to 15 years of residence. Although this estimate has been shown to be overly optimistic, there is widespread evidence that immigrant wages catch up with the natives over time. Duleep and Dowhan (2002) and Lubotsky (2007) in particular present evidence using longitudinal data for the US.

A number of potential explanations for the wage catch-up have been proposed. Eckstein and Weiss (2004) summarize the channels through which immigrants assimilate as follows: "With the passage of time in the host country, immigrants invest in local human capital and search for better matches with local employers, and employers become less uncertain of the immigrant's potential and realized quality." Similar explanations are mentioned in Chiswick (1978), Borjas (2000) and LaLonde and Topel (1997). This quote refers to three models of the distribution of earnings which may be used to explain immigrant economic assimilation: a human capital, a search and matching and an employer learning model.

Surprisingly no research has focused on studying the relative importance of these channels. In fact, most empirical studies of immigrant wages start from a generic statement of the human capital model[1] and focus mainly on measuring the immigrant catch-up rate. Within the human capital framework, several contributions highlight the importance of different factors, such as speaking the host country language (Chiswick and Miller (1995)), the age at arrival in the host country (Friedberg (1992)) or the country of origin (Chiswick (1978),Borjas (2000)) in explaining the immigrant wage catch-up. However, no systematic attempt has been made to differentiate between immigrant economic assimilation channels.

[1]Borjas (2000) shows how different assumptions made on the human capital production function may lead to very different predictions in terms of immigrant wage patterns. Few papers take the human capital model seriously to investigate the mechanisms further. An exception is Eckstein and Weiss (2004) who assume an exogenous increase in the returns to immigrants' skills and model the investment in human capital with time spent in the host country.

This chapter is a first step to address this gap in the literature. I use a unique linked employer-employee panel to study the early careers of immigrants in Portugal. The contribution of this chapter is two-fold. First, exploiting the richness of the data, I document new immigrant assimilation patterns in the first years in the host country. In particular, I show that job mobility and firm heterogeneity play an important role in the assimilation process. Second, motivated by the stylized facts, I build an economic assimilation model based on employer learning with firm heterogeneity and complementarities between worker and firm type. I derive additional predictions from the model and show that they are in line with the patterns in the data and that they can not be trivially explained by a search and matching or human capital model.

I start the empirical analysis by measuring the immigrant wage catch-up rate. I document that upon arrival immigrants earn 34% less than natives of the same age and 16% less than natives of the same age working in the same region, industry and occupation. I show that the gap closes at a rate of 1 percentage point per year spent in the country. As I use a panel which covers virtually all workers in the private sector, selection concerns are reduced. Estimates with and without individual fixed effects are very similar showing that selection is not a major concern in this context. This estimate of the wage catch up is in line with the literature for the US. For instance, Lubotsky (2007), using longitudinal social security data, shows that immigrants' earnings catch up with the natives at a rate of 10 to 15 percentage points in 20 years.

Accounting for immigrant sorting across regions, industries and occupations does not change the estimated catch up rate significantly. Immigrants do not assimilate by changing occupations and moving to different industries. However, this chapter shows that they do assimilate by switching firms. In fact, the first years in the country are characterized by a very high job mobility rate and one third of the immigrant wage catch up is linked to moving to better paying firms. This finding relates to a small but growing literature which measures how the sorting of immigrants across firms relates to the wage gap between immigrants and natives. Evidence for Canada[2] indicates that wage differences between firms are more important than differences within firms in explaining the immigrant-native wage gap. I build on this literature and show that moving to better paying firms is an important channel through which immigrants move up the wage distribution.[3]

I then use the rich information in the data to focus more directly on the role of firms in the assimilation process. Over time, immigrants move to bigger and more productive firms and get access to longer term contracts. Immigrants tend to start their careers in firms with a high proportion of immigrant workers and over time they move to firms with a higher share of native workers.

Moreover, I show that the wage catch up and firm mobility patterns are very similar for

[2] See Aydemir and Skuterud (2008) and Pendakur and Woodcock (2009)

[3] Pendakur and Woodcock (2009) find evidence that immigrants who have spent more years in the host country work in less segregated and better paying firms than recent immigrants. However they are unable to rule out that this result may be driven by differences in characteristics of different cohorts of immigrants or by self-selection in out-migration. I estimate the wage regressions with firm *and* worker fixed effects, which allows to separate the effects.

all the recent immigrants irrespective of their origin and in particular of whether their mother tongue is Portuguese. This result is at odds with a human capital accumulation explanation of the wage catch up. One would expect immigrants who speak the language to suffer a lower wage penalty to begin with but also to catch up more slowly.

Motivated by this new set of empirical facts on immigrant careers, I suggest an economic assimilation mechanism which highlights imperfect information about the productivity of immigrants. The model presented is an employer learning model with firm heterogeneity and complementarities between worker and firm type. It builds on the employer learning model by Farber and Gibbons (1996) and Lange (2007). These models assume that firms are homogeneous and that workers are paid their expected marginal productivity, which is independent of the firm they work for. I introduce firm heterogeneity and an assignment mechanism to allocate workers to firms. The mechanism considered is similar to the one in the differential rents model presented in Sattinger (1993). Each firm hires one worker and workers are assigned to firms according to their expected productivity given the information available at the time. As there are complementarities between worker and firm productivity, workers with higher expected productivity are assigned to more productive firms.[4]

The focus of the model is on the uncertainty: I assume that the only difference between immigrants and natives entering the labour market is that there is more uncertainty about immigrants' productivity than about natives'. I consider this to be a reasonable assumption: Typically it is easier for employers to judge the skills of a native than those of an immigrant. For instance, the evaluation of prior experience and education is less straightforward in the case of immigrants.

In the model, firms produce subject to decreasing returns to skill and thus value certainty over worker productivity. This prevents immigrants from getting access to the more productive firms in the first years in the host country. With time spent in the labour market, the uncertainty over worker productivity decreases and workers get matched on average to more productive firms.

The predictions of the model on the mean wages and the job mobility patterns are in line with the stylized facts. The learning model also has strong predictions on the evolution of the distribution of immigrant wages over time, and in particular on the variance of wages. I take these predictions to the data and study the variance of wages of immigrants and natives entering the market in the same year over time. The variance of the log wages is higher for natives than for immigrants and increasing for both groups over time. I show that firm heterogeneity accounts for a significant part of the increase in the variance of log wages. These results are in line with the predictions of the model.

Finally, I show that the results are not trivially derived from a competing search and matching or human capital explanation.

[4]Two papers who combine complementarities in the production function and employer learning are Gibbons et al. (2005) and Groes et al. (2010). The complementarity I am assuming is between worker and firm type, whereas in these papers they refer to industry and worker type and occupation and worker type.

Section 2 of the chapter describes the data and presents some descriptive statistics on the immigrant population. Section 3 documents immigrant assimilation patterns. In section 4, I present an employer learning model with firm and worker heterogeneity and derive predictions on the distribution of immigrant wages. Section 5 compares the distribution of wages for immigrants over time against the predictions from the model and section 6 discusses other possible assimilation mechanisms and how they compare to the patterns in the data. Concluding remarks are presented in section 7.

1.2 Data and Descriptive Statistics

1.2.1 Data, Context and Sample Selection

Every year in November, firms registered in Portugal must hand in a detailed questionnaire ('Quadros de Pessoal') to the Portuguese Ministry of Labour. This process is mandatory for all firms in the private sector employing at least one wage earner. With the exception of the public service and domestic workers, virtually all wage earners in the Portuguese economy are covered by the survey. The questionnaire contains detailed information about the firm (the location, the volume of sales, the industry, etc.), the establishment (the location, the number of workers, the collective bargaining agreement, the industry, etc.) and the worker (age, gender, education, nationality, etc.). All workers, firms and establishments have a unique identifier which allows to track them over the years.

When a worker is not in the panel in a given year, it is impossible to distinguish whether he is unemployed, working in the public sector or in the informal sector. In the case of immigrants, in particular, when a worker drops out of the panel, it is impossible to know whether he has migrated to the home country (or to a third country).

Portugal, like Italy, Spain or Greece, has been an emigration country for most of the last century and this trend has only been reversed in the last 10 years. These traditional emigration countries are now experiencing large inflows of immigration. Net migration numbers between 2000 and 2007 are striking: there are an additional 4.6m legal immigrants in Spain, 2.6m in Italy and close to half a million in Portugal and Greece.[5] In order to deal with the large inflow of undocumented immigrants, the Portuguese government organized an "extraordinary regularization" in 2001. The foreign legal population in Portugal increased by 69% in that year. Approximately 183,000 individuals got a permit to live in the country for a year. The permits were renewable up to four times. After five years, immigrants could apply for a long-term residence permit. Having a work contract in Portugal was the main condition to obtain and renew a short-term residence permit. In 2003, bilateral agreements were signed with Brazil which allowed Brazilian immigrants residing in Portugal before July 2003 to obtain a long-term residence permit. Although there has been no major regularization programme since 2003, immigrants may apply for a residence permit if they are in the country, have a work contract

[5]These numbers represent respectively 10.5%, 4.2%, 3.7% and 2.7% of the countries' total populations in 2007, according to Eurostat.

and are registered with the social security.

I restrict the analysis to immigrants from the new immigration wave, that is immigrants who enter the labour market after 2001. In 2000, only 0.5% of workers in the data are immigrants, in 2002 immigrants represent 4% of workers. The data set covers only workers in the formal sector. As there is no direct information on the years immigrants have spent in the country, I build a proxy which indicates the first year the immigrant appears in the data, that is the first year the immigrant has a job in the formal sector. In all the analysis, the variable "years since migration", YSM, refers to years in formal employment, and the "cohort" the immigrant belongs to is the first year he is tracked in the data.

Figure 1.1 shows the mean hourly wages for the different cohorts of immigrants over time. The trend in mean wages is similar for all cohorts. The 2002 cohort captures a high proportion of the immigrants who took advantage of the 2001 regularization. These immigrants may have been working informally in the country in the previous years.[6] One may thus be concerned that this cohort is unusual. The trend in mean wages of the 2002 cohort is nevertheless similar to the other cohorts which eases this concern.

I use the information in the data on the workers' nationality to define immigrants as foreigners. In the short run naturalization is not an issue, since immigrants need at least six years of legal residence to be able to apply for Portuguese citizenship.[7]

I restrict the analysis to immigrant men. Women represent less than 30% of immigrant observations in the data and would need a separate analysis. Immigrant women in Portugal often get jobs as domestic workers and are hence not covered in the data. I restrict the sample used to native and immigrant men. In the 2002-2009 period, I follow the early careers of close to 120,000 immigrant men.

1.2.2 Descriptive Statistics

The immigrants considered in the data are divided into three main origin groups, representing more than 90% of the total number of immigrant observations: Immigrants from Eastern and South Eastern Europe (Eastern Europeans, in the text), Brazil, and the former Portuguese African colonies (Africa)[8].

Graphics 1.2 and 1.3 illustrate the number of immigrants in the data each year; and the number of immigrants who belong to each cohort, from 2002 until 2009. After the large increase of foreign legal residents in Portugal in 2001, the number of immigrants continued to increase. With worsening labour market conditions, the inflow of immigrants slowed down after 2005 and the stock of foreigners in the data actually decreased in 2006 and 2009. The representation of the main origin groups has also changed over the years. Immigrants from Eastern Europe are the group which took greatest advantage of the 2001 regularization (101,000 permits), in particular

[6] Detailed information on the construction of the panel and the variables is in the appendix.

[7] Also, if an individual is foreign for five years and then becomes Portuguese, he is considered to be an immigrant for the analysis. More details on the construction of the panel are presented in the appendix.

[8] The exact definitions of the groups are in the appendix. Immigrants from the EU15 represent 4.5% of immigrant observations and are excluded from the analysis.

citizens from the Ukraine (65,000 permits) and Moldova. The number of immigrants from Eastern Europe entering the country declined sharply over the years and, as figure 1.2 shows, even the stock of Eastern European immigrants is in decline. Brazilians started migrating later to Portugal, and by 2009 are the biggest of the three groups in terms of new migrants. Since 2007 Brazil is the most common citizenship of immigrants residing legally in Portugal. Immigrants from Africa are the oldest immigrant community in Portugal. Although this group also benefited from the 2001 regularization, there has been immigration from Africa, mainly from Cape Verde, since the 1980s. Until 2007 Cape Verdeans were the largest foreign community in Portugal. The assimilation patterns of this group turn out to be slightly different from those of the immigrants from the recent immigration wave.

Selected descriptive statistics of the data used are presented in table 1.1. Immigrants are younger than the native population, and they have worked in Portugal on average just a little more than 3 years. Immigrant men are very concentrated in a small number of industries: construction by itself accounts for more than 42% of the immigrant observations. Immigrants from different origin groups select into different industries: 46% percent of the observations for men from Eastern Europe and 56% from Africa are jobs in construction, whereas for Brazilians the proportion is only 34%. Brazilian immigrants are more likely to work in hotels and restaurants. Furthermore, immigrants from Eastern Europe are more evenly spread in the different regions of the country, whereas immigrants from Africa are very concentrated in the Lisbon metropolitan area where the traditional community has settled since the 1980s.

1.3 The Economic Assimilation of Immigrants

1.3.1 Measuring the Wage Catch-up

The main question in the immigrant assimilation literature is whether the gap in wages immigrants experience upon arrival decreases with time spent in the host country. Following the literature, I estimate equation (1.1) below by ordinary least squares. The log hourly wage of worker i in job j in year t is given by:

$$ln(HW)_{ijt} = \alpha FG_i + \gamma YSM_{it} + X_{ijt}\beta + \eta_i + \epsilon_{ijt} \tag{1.1}$$

The dependent variable is the worker's log hourly wage, FG is a dummy that indicates whether the individual is an immigrant and YSM are the years since migration. YSM is set to 0 for natives. The coefficient α measures the immigrant-native wage gap and γ the rate at which the gap decreases with years since migration[9]. I measure the wage gap and the wage catch-up controlling first only for a quartic in age, and then progressively controlling for region, industry and occupation. This specification is restrictive since it assumes that the returns to characteristics are the same for immigrants and natives but nevertheless represents a useful benchmark.

[9] Introducing higher order polynomials in YSM does not change the results. The effect of years in the country is close to linear in the first ten years in the country.

The results for the different specifications are presented in table 1.2. The mean hourly wage gap is 34.4% in the first year and decreases by 0.9 percentage points with each year spent in the country.[10] Adjusting by differences in sorting across regions and industries reduces the initial gap to 24.5% and accounting for occupational differences reduces the gap still further to 14.6%. More than half of the wage gap between natives and immigrants is due to differences in immigrant sorting into different regions, industries and occupations. The wage catch-up rate γ however is very stable across specifications. This result shows that the immigrant wage catch-up occurs within narrowly defined regions, industries and occupations. In the first years in the country, immigrants have higher wage growth than natives of the same age. The catch-up is not correlated to immigrants moving to different industries or occupations over time.

Cross-sectional calculations of the catch-up rate tend to over-estimate immigrant assimilation if more successful immigrants have a higher probability of remaining in the host country and less successful ones return to their home countries. I estimate all the specifications with individual fixed effects in order to address the selection concern. The results are presented in the last three columns of table 1.2. Controlling for individual fixed effects also does not change the γ significantly which indicates that the bias due to self-selection in out-migration is not a major concern in this context. Changing regions, industries and occupations is part of the assimilation process. I therefore choose the specification controlling only for a quartic in age and individual fixed effects as my preferred specification. The immigrant wage catch-up is set at 1 percentage point per year. This estimate is similar to the estimates for the US using panel data. Lubotsky (2007) evaluates the closing of the wage gap in the US at 10 to 15 percentage points in 20 years.

Next, I run the regressions separately for different origin groups. Table 1.3 presents the preferred specification, which controls only for a quartic in age, with and without individual fixed effects, for the 3 main origin groups. The wage gap is similar for all origin groups. It is 6 percentage points lower for Brazilians than for immigrants from Eastern Europe. The gap for immigrants from Africa lies in between. After accounting for individual fixed effects, the wage catch-up rate is slightly higher than 1 percentage point for Brazilians and Eastern Europeans but immigrants from Africa lag substantially behind. These results show that speaking the host country language may not be as important as one might have imagined for immigrant assimilation. Eastern Europeans are the only immigrants whose mother tongue is not Portuguese, yet their wage growth is comparable to the one experienced by Brazilians. The descriptive statistics show that immigrants from Brazil self-select into different sectors and occupations than Eastern Europeans, but after this initial sorting the assimilation patterns are very similar.

1.3.2 A Distributional Approach

The previous results establish that there is immigrant wage catch-up as measured by the mean hourly wages. Comparing the whole distribution of log hourly wages of immigrants and natives

[10] The variable YSM is set to 1 in the first year an immigrant is in the country so the initial gap is $-0.353 + 0.009 = -0.344$

shows that the distribution of wages of immigrants is becoming more similar to that of the natives with time spent in Portugal. Figure 1.4 illustrates this point. The graphic shows the representation of immigrant wages in the distribution of native wages by years since migration, and more specifically in the entry year, after 5 years and 9 years in the country. For example, in the first year in the country on average 33% of immigrants earn less than the lowest decile of the native distribution. After 5 years in the host country, less than 5% of immigrants do so. With years spent in the country, the distribution of wages of immigrants widens and comes closer to the native wage distribution.

The calculations in this section use all cohorts and all years pooled together. One might worry that the results are confounded by cohort effects and selection. To address this concern, I do the same calculations for each cohort separately, for the whole cohort and for "stayers" only. I consider "stayers" immigrants who can be tracked in the data each year. The graphics in figure 1.5 show the results for the 2003 cohort. Immigrants move up the wage distribution also when considering only "stayers" of the same cohort. The results for all other cohorts and origin groups are similar and presented in the web appendix.

These results show that over time immigrants move up the wage distribution. In the next sections, I focus on a specific mechanism through which the catch up occurs: job mobility. I first estimate a linear probability model of job mobility; I then show that the wage catch up is linked to immigrants moving to better firms; and finally I present descriptives on the firms that immigrants work for over time.

1.3.3 Immigrant Job Mobility

A very strong empirical regularity in the data is that the immigrant job mobility is very high. Table 1.4 presents results on a linear probability model of changing employers. The dependent variable is a dummy that equals 1 if the worker-firm match will end in the next period, 0 if the worker is still working for the same firm in the next period. Only workers who are in the data in two consecutive years are considered in the analysis. On average 7% of native workers change employers in a given year. The rate is much higher for immigrants : after the first year in the host country, 26% of immigrants change employers[11]. The probability of changing firms for immigrants decreases by approximately 2.1 percentage points per year. In specifications (3) to (5) of table 1.4, I introduce other variables in the model. In line with the literature on job mobility, eg. Farber (1999), I control for a cubic in tenure and the current hourly wage in column (3), and account for differences in sorting across regions and industries (column (4)) and occupations (column (5)). Immigrants have on average lower tenure, lower wages and work in different industries and occupations than natives. These differences account partly for the differences in job mobility rates: there is nevertheless a remaining unexplained gap between immigrants and natives.

[11] The variable YSM is set to 1 in the first year an immigrant is in the country so the initial gap is $0.211 - 0.021 = 0.191$

1.3.4 The Role of Firms in Immigrant Assimilation

Introducing Firm Heterogeneity in the Wage Catch-up Estimations

Recent evidence from Canada[12] indicates that the immigrant native wage gap is associated to immigrant sorting across firms. Immigrants are not paid less than natives working in the same firm, but are systematically concentrated in firms that pay less, holding worker and job characteristics fixed. In this section, I look at whether with time spent in the host country immigrants move to firms that pay better, and if so, how much of the wage catch-up does this upward mobility account for.

I introduce firm heterogeneity in the wage equation estimated in the previous section in order to investigate whether the immigrant wage catch-up is related to immigrants moving to better paying firms over time. This estimation is a wage decomposition with individual and firm fixed effects following Abowd et al. (1999). This chapter is the first to present the AKM decomposition in the context of immigrant assimilation. I thus augment equation (1.1) as follows[13]:

$$ln(HW)_{ijt} = \alpha FG_i + \gamma YSM_{it} + X_{ijt}\beta + \eta_i + \mu_j + \epsilon_{ijt} \tag{1.2}$$

The estimation results are presented in table 1.5. Columns (1) to (3) reproduce the results from table 1.2 controlling for individual fixed effects. Columns (4) and (5) add firm fixed effects. Column (4) controls only for a quartic in age, whereas column (5) controls also for occupation. Comparing the estimates for the main coefficient of interest, the wage catch-up rate γ, with and without firm fixed effects, gives us an idea of the role of firm heterogeneity in immigrant assimilation. Controlling for firm fixed effects, in addition to region and industry, decreases the estimated catch-up rate from 1 to 0.6 percentage points. In the estimations controlling also for occupations, the rate decreases similarly from 0.9 to 0.6 percentage points. When analyzing the importance of sorting across firms in the immigrant wage gap, Pendakur and Woodcock (2009) show evidence that immigrants who have been in Canada for 10 years or more work in higher fixed effect firms than more recent immigrants. However, they can not exclude that this result may be due entirely to selection. The estimations with firm and individual fixed effects indicate that moving to higher paying firms is indeed an important channel through which immigrant wages catch up.

Table 1.6 shows the estimations for immigrants from the main origin groups. Comparing the estimates with and without firm fixed effects, the wage catch-up decreases from 1.3 to 0.9 percentage points for Eastern Europeans, 1.1 to 0.8 percentage points for Brazilians and from 0.3 to -0.1 percentage points for immigrants from Africa. Changing firms accounts for approximately one third of the wage catch-up for Eastern Europeans and Brazilians. For immigrants from Africa, all of the observed catch-up occurs by changing firms.

[12] The main papers are Aydemir and Skuterud (2008) and Pendakur and Woodcock (2009).

[13] I estimate this wage regression with two high dimensional fixed effects using the algorithm presented in Guimaraes and Portugal (2009) implemented in Stata through the command reg2hdfe.

Immigrants Climb up the ‘Firm Quality Ladder’ with Time Spent in the Host Country

Not much is known about firms that hire immigrants and how immigrants progress in the firm "quality ladder" with time spent in the host country. The previous section shows that immigrants sort into low-wage firms and part of the assimilation process goes through switching to better paying firms. In this section, I take a closer look at firms where immigrants work and at immigrant careers in the first years in the country from a firm perspective.

Figure 1.6 shows firm descriptives for firms where immigrants work over time. With years spent in the host country, a higher proportion of immigrants gains access to long-term contracts. Immigrants also become more integrated in the labour market: They start off their careers in firms with a very large share of immigrant workers[14], but are exposed to more native co-workers as time goes by. They also move to larger firms.

Firm fixed effects measure the firm wage premium, i.e., how firms in narrowly defined regions and sectors reward individuals working in the same occupation differently. The firm fixed effects are often thought of as a measure of firm productivity. Another more direct measure of firm productivity is the firm's volume of sales per worker. The firm fixed effects estimated in the previous section are net of the individual fixed effect. As a robustness check, I also estimate firm fixed effects using the same specification than in equation 1.2 but without individual fixed effects. All measures of productivity (volume of sales per worker and firm fixed effects estimated with or without individual fixed effects) show similar patterns: over time, immigrants move to firms which are on average more productive.

The results for all immigrant groups are similar and are presented in the web appendix.

One worry about these descriptive statistics is that they pool together all cohorts and do not deal with selective out-migration. For instance, if only immigrants who start off their careers in more productive firms remain in the country, the results would be driven exclusively by selection and would not tell us much about the assimilation process. To address this concern, I do the same calculations for all cohorts separately distinguishing between all the immigrants from a cohort and "stayers". The means are first calculated each year for all immigrants belonging to a certain cohort and then only for immigrants who can be tracked in the data each year. The graphics for the 2003 cohort are presented in figure 1.7. The graphics for all other cohorts are similar and are presented in the web appendix. There is no initial difference in the proportion of immigrants who hold long-term contracts comparing immigrants who remain in the panel all the years and all the immigrants in the cohort. However, as immigrants get long-term contracts, they become more likely to remain in formal employment in Portugal, which explains the divergent trends between the two groups. All the other graphics suggest a common analysis. Immigrants who stay in formal employment each year are the ones who start off in larger, more productive and more integrated firms. In terms of assimilation, the important aspect is that although the means are higher in levels for "stayers", the trends are in most cases parallel.

[14]For papers that analyze immigrant segregation in the workplace using linked employer-employee data see Andersson et al. (2010) and Dustmann et al. (2011).

The detailed calculations allowing for cohort effects and selection confirm the overall interpretation of the plots in figure 1.6. One of the channels of immigrant assimilation goes through moving to larger, more integrated and more productive firms.

The descriptives presented above show that immigrants move up the wage distribution with years spent in the host country labour market. A third of this upward mobility is linked to moving to firms that are more productive and that pay higher wages. In the next section, I build a model of immigrant economic assimilation based on firm heterogeneity and employer learning. When immigrants enter the labour market, little is known about their true productivity. There are complementarities between worker and firm type and firms value certainty over a worker's productivity. With high uncertainty about their types, immigrants begin their careers at the bottom of the firm productivity distribution. Over time, worker productivity is revealed and, on average, immigrants get better matches. I simulate the model and show in the subsequent section that it can account for many qualitative features of the data.

1.4 A Learning Model with Firm and Worker Heterogeneity

1.4.1 The Workers and the Firms

Each worker has a productivity η_i. This productivity is composed of three additive terms:

$$\eta_i = q_i + a_i + s_i$$

The term q is observed for all workers as, for example, skills easily observed at a job interview. The component a is unobserved for all workers and captures "true" ability or IQ. Finally, the term s is observed for natives but not for immigrants as, for example the quality of a worker's education. All three terms are independently drawn from normal distributions with means μ_a, μ_q and μ_s and standard deviations σ_a, σ_q and σ_s. The independence of a with respect to q and s is a strong assumption but common in the employer learning literature. The productivity η hence follows a normal distribution with mean $\mu_\eta = \mu_a + \mu_q + \mu_s$ and standard deviation $\sigma_\eta = (\sigma_a^2 + \sigma_q^2 + \sigma_s^2)^{\frac{1}{2}}$. In line with the employer learning literature[15], I assume the different components of worker productivity to remain unchanged over time.

The productivity of firms in the economy is assumed to follow a log normal distribution with mean μ_c and standard deviation σ_c.[16] The distribution of firms is taken as given in the model and is fixed over time. The productivity of each firm is known by all agents in the market and is constant over time. Each firm hires only one worker and takes the wage schedule as given. The worker i - firm j match at time t produces output:

$$y_{ijt} = c_j \left[K - (\exp\left(-(\eta_i + \epsilon_{it})\right))\right]$$

[15] Farber and Gibbons (1996) or Lange (2007)

[16] For evidence on the skewness of the firm productivity distribution in the US, see Bartelsman and Doms (2000)

where K is a large positive constant and $\epsilon_{it} \sim N(0, \sigma_\epsilon)$ is a random error to production.[17] For a given firm j, output is concave in the worker's ability η_i. The shape of the production function captures the idea that the quality of the machine (the firm productivity) limits the productivity of the worker. This production function ensures that the firm's expected output depends negatively on the uncertainty on the worker's productivity which will be a key element in the allocation of workers to firms in the model.

1.4.2 The Learning Process

Each period, all employers observe a noisy measure of the worker's productivity, $\eta_i + \epsilon_{it}$, and update their beliefs. There is symmetric learning: the current employer does not have more information about the worker's productivity than other potential employers. What is learnt about worker i at time t is also independent of the worker-firm match. Agents observe y_{ijt} and make their update on

$$\xi_{it} \equiv -\log\left(K - \frac{y_{ijt}}{c_j}\right) = \eta_i + \epsilon_{it}$$

The noise is assumed to be independent of all other variables in the model and is the same for immigrants and natives.

The normality assumptions make the learning process easily tractable. After a worker has spent x years in the labour market, the posterior distribution of worker i's type is a normal distribution with mean $\mu_{x,k,i}$ and standard deviation $\sigma_{x,k}$, where k is an index for immigrant fg or native nat. The expected productivity of an immigrant worker is:

$$\mu_{x,fg,i} = \frac{\sigma_\epsilon^2}{x(\sigma_a^2 + \sigma_s^2) + \sigma_\epsilon^2}(q_i + \mu_a + \mu_s) + \frac{\sigma_a^2 + \sigma_s^2}{x(\sigma_a^2 + \sigma_s^2) + \sigma_\epsilon^2}\sum_{l=0}^{x-1}\xi_{il}$$

and its variance is:

$$\sigma_{x,fg}^2 = \frac{\sigma_\epsilon^2(\sigma_a^2 + \sigma_s^2)}{x(\sigma_a^2 + \sigma_s^2) + \sigma_\epsilon^2}$$

For a native worker:

$$\mu_{x,nat,i} = \frac{\sigma_\epsilon^2}{x\sigma_a^2 + \sigma_\epsilon^2}(q_i + \mu_a + s_i) + \frac{\sigma_a^2}{x\sigma_a^2 + \sigma_\epsilon^2}\sum_{l=0}^{x-1}\xi_{il}$$

and

$$\sigma_{x,nat}^2 = \frac{\sigma_\epsilon^2\sigma_a^2}{x\sigma_a^2 + \sigma_\epsilon^2}$$

[17] Since η follows a normal distribution, there are workers who produce negative output. I choose m_η and K large enough such that this fraction of workers is negligible.

The expected worker productivity is a weighted average of the initial prior and the observed performance on the labour market. Initially, the weight on the prior is higher for natives as the prior is more precise. Over time the worker's expected productivity converges to the true productivity. The variance of the posterior is higher for immigrant workers as there is more uncertainty about them. Over time, the difference between the two groups decreases and in the limit the variance of the posterior tends to zero for every worker.

After x years in the labour market, the cross-sectional distribution of expected productivity for all immigrant workers of the same cohort is a Normal distribution with expected value

$$E\left(\mu_{x,fg}|I_x\right) = \mu_\eta$$

and variance[18]

$$V\left(\mu_{x,fg}|I_x\right) = \sigma_q^2 + \frac{x^2(\sigma_a^2+\sigma_s^2)^3}{(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2)^2} + \frac{x\sigma_\epsilon^2(\sigma_a^2+\sigma_s^2)^2}{(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2)^2}$$

Similarly for natives, expected productivity for all native workers of the same cohort, $\mu_{x,nat}$, follows a Normal distribution with expected value

$$E(\mu_{x,nat}|I_x) = \mu_\eta$$

and variance

$$V(\mu_{x,nat}|I_x) = \sigma_q^2 + \sigma_s^2 + \frac{x^2(\sigma_a^2)^3}{(x\sigma_a^2+\sigma_\epsilon^2)^2} + \frac{x\sigma_\epsilon^2(\sigma_a^2)^2}{(x\sigma_a^2+\sigma_\epsilon^2)^2}$$

Over time, the distribution of expected productivity becomes wider for both groups, while the mean always stays the same. Due to the initial information asymmetry between natives and immigrants, the distribution of expected productivity is always wider for natives. Over time, the two distributions converge.

1.4.3 The Assignment Mechanism

The expected production of a firm j that hires worker i conditioned on all information available about the worker after x periods in the labour market is:[19]

$$E\left(y_{ijt}\right) = c_j\left[K - \exp\left(-\mu_{x,k,i} + \frac{1}{2}(\sigma_{x,k}^2+\sigma_\epsilon^2)\right)\right]$$

Firms prefer to hire workers with a higher risk-adjusted expected productivity $\mu_{x,k,i} - \frac{1}{2}\sigma_{x,k}^2$. Within a group and cohort, firms prefer workers with a higher expected productivity $\mu_{x,k,i}$. The term $\sigma_{x,k}$ introduces a distortion across groups and cohorts: For a given expected productivity, firms prefer workers for whom expected productivity is more certain. This introduces an

[18] The calculation is in the appendix.

[19] This expression comes from the fact that $\exp(-(\eta_i+\epsilon_{i,t}))$ follows a log normal distribution with mean $\exp(-\mu_{x,k,i} + \frac{1}{2}(\sigma_{x,k}^2+\sigma_\epsilon^2))$

advantage for older cohorts and natives in the labour market.

For each cohort of natives or immigrants at each level of experience in the labour market, $\mu_{x,k} - \frac{1}{2}\sigma^2_{x,k}$ follows a normal distribution with expected value

$$M_{x,k} = E\left(\mu_{x,k} - \frac{1}{2}\sigma^2_{x,k}|I_x\right) = \mu_\eta - \frac{1}{2}\sigma^2_{x,k}$$

and variance

$$V_{x,k} = V\left(\mu_{x,k} - \frac{1}{2}\sigma^2_{x,k}|I_x\right) = V(\mu_{x,k})$$

The distribution of $\mu_{x,k} - \frac{1}{2}\sigma^2_{x,k}$ for *all* workers, immigrants and natives, of a given cohort after x years in the market is hence a mixture of two normal distributions. The C.D.F. of this distribution is:

$$F(t) = p\Phi\left(\frac{t - M_{x,fg}}{V^{\frac{1}{2}}_{x,fg}}\right) + (1-p)\Phi\left(\frac{t - M_{x,nat}}{V^{\frac{1}{2}}_{x,nat}}\right)$$

where Φ is the C.D.F of the standard normal distribution and p is the proportion of immigrants in the cohort.

Assuming that each worker remains in the labour market for T periods, that all cohorts are similar and that the proportion of immigrants is constant across years, the C.D.F. of the distribution of $\mu_{x,k} - \frac{1}{2}\sigma^2_{x,k}$ for *all* workers in the market in a given year is:

$$F(t) = \sum_{x=1}^{T} \frac{p}{T}\Phi\left(\frac{t - M_{x,fg}}{V^{\frac{1}{2}}_{x,fg}}\right) + \frac{1-p}{T}\Phi\left(\frac{t - M_{x,nat}}{V^{\frac{1}{2}}_{x,nat}}\right)$$

An efficient equilibrium at time t consists of an assignment of workers to firms and a wage schedule that maximize expected aggregate output. In such an assignment, each period workers are matched to firms according to the worker's risk-adjusted expected productivity and the firm's productivity. Worker i is assigned to firm j with productivity $c^*_j\left(\mu_{x,k,i} - \frac{1}{2}\sigma^2_{x,k}\right)$, such that

$$G\left(c^*_j\left(\mu_{x,k,i} - \frac{1}{2}\sigma^2_{x,k}\right)\right) = F\left(\mu_{x,k,i} - \frac{1}{2}\sigma^2_{x,k}\right)$$

where G is the C.D.F. of firm productivity. This assignment means that workers and firms are matched by their relative position in the probability distributions. In a discrete setup, this would mean that the nth worker, in order of decreasing expected worker productivity, will be employed by the nth firm, in order of decreasing firm productivity. This has to hold in an efficient equilibrium and follows from the firm-worker complementarity.

In this setup there is no need to solve a dynamic problem as every period the distributions of firms and workers' expected productivity are the same and there are no moving costs. Each period there is a new equilibrium based on all available information. Facing a wage schedule

$w(z)$, where z is risk adjusted worker productivity, firm j maximizes expected profits:

$$\max_{z} \left\{ c_j \left[K - \exp\left(-z + \frac{1}{2}\sigma_\epsilon^2 \right) \right] - w(z) \right\}$$

The first order condition implies that the expected marginal product must equal the marginal increase of the wage.[20] In equilibrium, this is only true for the proposed assignment, so I can write:

$$w'(z) = b\, c^*(z) \exp(-z)$$

where $b = \exp\left(\frac{1}{2}\sigma_\epsilon^2\right)$ is a constant. The wage schedule in the economy can be found by integrating this expression:

$$w(x) = b \int_A^x c^*(z) \exp(-z) dz$$

where A is the minimum worker productivity. Since there exists no closed-form solution for the optimal firm match $c^*(z)$, no explicit solution for the wage can be found. In the following subsection, the model's predictions on the moments of the wage distribution will thus be derived by simulation.

The shape of the wage schedule is governed by decreasing returns to skill, captured by $\exp(-z)$, and the match function $c^*(z)$. Decreasing returns alone would make the wage schedule concave. This is counteracted by the equilibrium assignment, according to which better workers work at better firms. Depending on the rate at which the optimal match function increases, the wage schedule can be locally convex or concave, but is in all cases increasing in worker productivity. The graphics in figure 1.8 plot the optimal firm match $c^*(z)$ and the wage $w(z)$ as a function of worker risk-adjusted productivity $z = \mu_{x,k,i} - \frac{1}{2}\sigma_{x,k}^2$. For the parameters chosen, the firm match function is strictly convex. In general, its exact shape depends on the parameters of the underlying skill and productivity distributions of workers and firms. In particular, the convexity of $c^*(x)$ is related to the skewness of the firm productivity distribution. If the firm productivity distribution is heavily right-skewed, then a marginal improvement in worker skill is associated with an increasingly better firm match, thus making the optimal match function convex.

1.5 Comparing the Predictions of the Model to the Data

In this section, I derive predictions from the model presented in the previous section on the distribution of wages and on job mobility patterns. I first show that the predictions on the immigrant mean wage, mean firm productivity and job mobility over time are in line with the stylized facts of section 3. I then take the additional prediction of the model on the variance of wages to the data.

The model does not have a closed form solution for the optimal worker firm match as c^* is

[20] The second order condition holds, since the cross-derivative of expected production is positive. See Sattinger (1993).

the inverse of the C.D.F. of a log normal distribution. I therefore simulate the model. There are 600,000 workers who each spend 30 periods in the labour market and immigrants represent 10% of workers in each cohort.

1.5.1 The Predictions of the Model and the Stylized Facts

In the empirical analysis in section 3, I highlighted three main stylized facts about the immigrant wage catch-up:

1. Immigrant wages catch up to the wages of natives of the same age group
2. In the first years in the country, immigrants exhibit high job mobility rates which decrease over time
3. Part of the immigrant wage catch-up is explained by immigrants moving to better paying and more productive firms

In this first section, I show how the model accounts for these stylized facts.

The Mean Firm Productivity and the Mean Wage over Time

In the model, the distribution of the risk-adjusted expected productivity for a cohort of immigrants moves to the right and becomes wider over time: the right-shift in the distribution is due to less uncertainty about immigrant true productivity: σ^2_{xk} decreases. The widening of the distribution comes from employer learning about each worker's true productivity.

Mean Firm Match: As firms reward certainty over the worker's productivity, new entrants on the market are matched to less productive firms on average. Among new entrants, immigrants have a higher uncertainty than natives and hence occupy on average the bottom of the firm productivity distribution. Over time, uncertainty decreases and workers gain access to better firms. This effect shifts the distribution of their firm matches to the right and hence increases the mean firm productivity over time. This effect is stronger for immigrants than for natives of the same cohort as there is more to learn about immigrants.[21]

Mean Wage: The reduced uncertainty about productivity also improves immigrants' wages through two main effects. First, as described above, they gain access to better firms, thus increasing their marginal product. Second, their expected marginal product increases due to reduced uncertainty: $\exp(-\mu_{xki} + \frac{1}{2}\sigma^2_{xk})$ declines. Job mobility thus accounts for only a part of the total wage gains in the model.[22]

[21] If the match function $c^*(x)$ is convex, as in the present simulation, there is another effect on the mean firm match: as true worker productivity is revealed, and the variance of the expected productivity distribution of a cohort rises, the mean match increases. However, as explained earlier, the local convexity of $c^*(x)$ depends on the exact parameter values chosen. This effect is second-order relative to the shift of the worker productivity distribution.

[22] Again, the local curvature of the wage function together with the increasing variance of the expected productivity distribution exerts a second order effect on mean wages.

The model also predicts an increase in the mean of the log firm match and the mean of the log wage for an entering cohort of workers. The same mechanisms that increase the mean wage and the mean firm match also lead to increases in the log of these variables.[23]

The graphics of figure 1.9 show the mean log firm productivity and the mean log wage for an entry cohort of immigrants over time. The left hand side graphics compare an entry cohort of immigrants to natives of the same cohort, and the right hand side graphics compare an entry cohort of immigrants to the whole native labour force. The mean log wage of immigrants is increasing and part of the increase is due to firm heterogeneity. Comparing immigrants and natives of the same cohort, the mean log wage is initially higher for natives as they start their careers in better firms. Over time, the mean log wage for both groups increases, more so for immigrants as there is initially more uncertainty about their productivity.

The model thus generates predictions that are consistent with the first and third stylized fact of the data: On average, immigrants catch up to natives of the same age group, and part of this catch up is accounted for by moving to better firms. To sensibly derive predictions on job mobility, a variant of the model is discussed in the next subsection.

Job Mobility

The assignment model presented above has very strong continuity assumptions and a restrictive one to one match. This way of modeling allows to solve explicitly for the optimal worker-firm match and hence to simulate the patterns of the firm productivity distribution over time. In this continuous version of the model, all workers move jobs every period as information is revealed. In order to make the predictions on job mobility more realistic, I make a small change to the model above and assume that there is a finite number of firms, and that each has a fixed number of jobs. Firms are ordered by their productivity level: $0 < c_1 < c_2 < .. < c_m$. All the other ingredients of the model remain the same.

As before, an equilibrium is defined by an assignment of workers to firms and a wage schedule. I can define $m-1$ worker risk-adjusted expected productivity thresholds, l_j, so that workers with risk-adjusted expected productivity $\mu_{x,k} - \frac{1}{2}\sigma^2_{x,k} \in [l_j, l_{j+1}]$ are assigned to the firm of productivity c_j. The wages are derived in the same way as in the model above. I assume that there are no moving costs. Workers switch firms when their risk-adjusted expected productivity is revealed to be much higher or much lower than expected - that is, when their expected productivity crosses a threshold l_j.[24]

Comparing immigrant and native workers, the model yields a main prediction: Immigrant workers switch firms more often than natives do, but the difference in job mobility between the two groups decreases over time. There is initially more uncertainty about immigrant productivity and more updating for immigrants each period. The difference between the two groups decreases over time as extra information each period represents a smaller and smaller part of

[23] Since the log wage function is concave the second order effect of an increasing variance of expected productivity now depresses the mean log wage.

[24] The distribution of the changes in risk-adjusted expected productivity for a cohort over time is derived in the appendix.

all information available about the worker. This prediction is in line with the stylized fact on immigrant job mobility from section 3. Immigrants move jobs more often than natives but at a decreasing rate.[25]

1.5.2 Taking an Additional Prediction of the Model to the Data

The Variance of Wages over Time

The model considered is an employer learning model and as such generates clear predictions on the second moment of the wage distribution. In this section, I show that the model predicts an increase in the dispersion of wages for immigrants over time and that this increase arises through switching firms.

Variance of Wages: As worker productivity is revealed, the distribution of expected productivity for a cohort of workers widens over time. This effect increases the variance of wages since workers are paid according to their expected marginal product. In the present model, this effect is magnified by worker assignment to heterogeneous firms. As the distribution of expected productivity widens over time, so does the distribution of firm productivity workers are matched to. If the c^* schedule is convex, then the dispersion of firm productivity will further increase due to a second effect: As new entrants move up the firm-quality ladder, they gain access to increasingly better firms. This is related to the underlying skewness of the firm productivity distribution. The distribution of assigned firm productivity for these workers widens and further contributes to the increase in the variance of wages. According to the model, we should thus see an increasing profile of the variance of log wages[26] for a cohort over time and this increase arises in the model through switching firms. If we consider a model with a finite number of firms, not all of the increase in the variance of wages is related to switching firms: the dispersion of immigrant wages increases even within the same firm as employers learn about worker productivity.

In the next section, I conduct an empirical analysis of the variance of log wages for immigrant and native workers entering the labour market between 2002 and 2009 in order to take this prediction to the data.

The Variance of Wages in the Data

I start by estimating equation (1.3) below by ordinary least squares:

$$\ln(HW)_{it} = origin_i * cohort_i * year_t + \epsilon_{it} \tag{1.3}$$

The variable *origin* is a dummy for each origin group: native, Brazilians, Eastern Europeans,

[25]The model considered is silent on the effect of tenure. A possible way to introduce the impact of tenure is to add accumulation of employer specific human capital. This generates moving costs which depend on the firm productivity. Solving for the extended model implies solving for a dynamic equilibrium instead of the stable equilibrium in the previous section.

[26]The increase in the variance of wages will also raise the variance of log wages. In the present simulation, this effect dominates the effect coming from the higher mean of wages, which depresses the variance of log wages.

Africans and other immigrants; *cohort* is a dummy variable for each entry cohort from 2002 until 2009; and *year* denotes a dummy for calendar year, from 2002 until 2009. This specification is a more general form of the specification used in section 3. The aim is to estimate the dispersion of the wages net of all mean effects. The residual estimated from this regression represents the part of the log wages which is not explained by the evolution of the mean log wages of workers from a given group and cohort over time.

The graphics in figure 1.11 plot the variance over time of the residuals estimated for natives and immigrants of the 2003 cohort under different specifications. I focus on the 2003 cohort as an example, the same analysis is conducted for all other cohorts in the web appendix. The first plot uses the specification of equation (1.3), the following plots add controls first for age groups, region and industry; then occupations; and finally firm fixed effects.

The variance of log wages is higher for natives than for immigrants and it is increasing over time for both groups. This stylized fact holds true independent of the exact specification considered. Controlling for region, industry and occupation explains part of the difference in the level of the variance of log wages between natives and immigrants. Immigrants have more undifferentiated log wages because they sort into more similar industries and occupations than natives. However, the increase in the variance profiles over time remains the same. Controlling for firm heterogeneity has a different effect. The increasing variance profile of immigrants and natives is flattened when firm heterogeneity is taken into account. I interpret this effect as evidence that new entrants on the market sort through changing firms. This effect is particularly strong in the first years in the labour market.

Figure 1.12 presents the same results but restricting the sample to workers from the 2003 cohort who are in employment every year. The patterns are very similar to those in figure 1.11 which shows that selection out of the labour market does not have an effect in these estimations. The results are similar for all cohorts and origin groups. This is shown in the web appendix.

The stylized facts are in line with the predictions of the model on the variance of log wages. The variance of log wages is higher for natives than for immigrants as initially more is known about native productivity. Natives have a higher variance of expected productivity and gain access to a wider range of firms. The variance of log wages is increasing over time for all new entrants in the market. As productivity is revealed, workers are sorted and work at more diverse firms. This mechanism is consistent with the stylized fact that firm heterogeneity explains part of the increase in the variance of log wages.

1.6 Competing Theories of the Distribution of Wages

1.6.1 The Learning Model with Firm and Worker Heterogeneity

The model of the distribution of wages presented in section 4 is a model in which the type of workers is unknown and as productivity is revealed workers are assigned to more productive firms. The predictions of the model are consistent with the empirical analysis presented in section 5. The mean wages and the variance of wages are increasing over time. Both of these effects are partly explained by switching firms and the probability of switching firm decreases over time.

To model the difference between natives and immigrants, I assumed that there is initially more uncertainty about immigrant productivity than native productivity. Two stylized facts are in line with this assumption: immigrants switch firms more often than natives; and the variance of wages is higher for natives than for immigrants.

An additional prediction from the learning model is that the variance of the changes in expected productivity of workers of the same cohort declines over time. With time spent in the labour market there is progressively less to be learnt about the worker's productivity. This is the mechanism which leads to the decrease in job mobility over time. Initially, as there is more uncertainty about immigrants, the variance of the changes in expected productivity is higher for immigrants than for natives. The variance decreases for both groups over time but faster for immigrants than for natives.[27] In order to investigate this prediction, I first estimate the following equation with ordinary least squares:

$$\Delta \ln(HW)_{it} = origin_i * cohort_i * year_t + \epsilon_{it} \tag{1.4}$$

This equation is similar to the one used to estimate the variance of log wages in the previous section. I calculate the variance of the residual for immigrants and natives for each cohort, each year. I consider in this calculation only "stayers", that is workers who remain in employment every year. I then estimate the following regression by weighted least squares:

$$Var(\hat{\epsilon}_{lt}) = \alpha FG_l + \beta EXP_l + \gamma FG_l * EXP_l + year_t + \epsilon_{lt} \tag{1.5}$$

l refers to a origin-cohort (for ex. natives belonging to the 2003 cohort), t refers to calendar time. Table 1.7 presents the estimations. The variance of the wage growth is on average higher for immigrants than for natives and decreases for both groups over time. I interpret the fact that the variance of the wage growth is higher for immigrants than for natives as evidence of the higher uncertainty over immigrant productivity.

In the next sections I consider two competing models of the wage distribution and investigate whether they match the stylized facts on immigrant economic assimilation. Table 1.6.3 compares the predictions of the three competing explanations to the stylized facts.

1.6.2 Search Model

A competing model of the distribution of wages which may be useful in the context of understanding the immigrant wage catch up is a search model. This class of models departs from the perfect competition framework and introduces search frictions. Workers need time to receive wage offers, and as they do, they climb up the wage distribution. I assume the difference between immigrants and natives to be that immigrants have less "search capital" upon arrival in

[27]The distribution of the changes in risk-adjusted expected productivity for a cohort over time is derived in the appendix.

the country and over time they receive wage offers at an increasing frequency.

In order to be more specific, let us consider a simple search model: The distribution of wages is exogenous, workers get wage offers from a wage distribution with C.D.F F. Offers arrive at a rate $\lambda(x)$. If the new wage offer is higher than the current wage the worker switches jobs, if not he remains with the same employer. This model is a simple on the job search model as for instance Burdett (1978). I abstract in this simple model from unemployment. Workers remain in employment all periods. When taking the predictions of the model to the data I consider only "stayers", that is, workers of an entry cohort who are always in employment. The difference between immigrant and native workers in the model is then modeled by a different arrival rate of wage offers. Immigrants are assumed to have initially lower search capital, $\lambda_{fg}(0) < \lambda_{nat}(0)$, but the rate of arrival increases faster for immigrants than for natives $\lambda'_{fg}(x) > \lambda'_{nat}(x)$.

According to this simple model, the mean wages of workers of a cohort increases and the increase is due to switching firms. All workers move up the wage distribution as they receive more wage offers over time. Workers also switch firms at a decreasing rate with time spent in the labour market. As workers move to better firms, the probability of receiving a better offer decreases over time. These two predictions are in line with the patterns in the data for new entrants. However, another prediction of the search model is that the distribution of wages of a cohort over time becomes truncated to the left. Workers who started off in the worse jobs move up over time, faster than the workers who started with a higher relative wage. This mechanism implies that the variance of wages of a cohort decreases over time.[28]

Figure 1.13 shows a simulation of the mean and variance of the log wages and wage growth of the model above for an entry cohort in the labour market. In this simulation, the probability of receiving a wage offer each period is constant and set equal to 0.1. The mean wage is increasing and the variance of wages is decreasing with time spent on the market. The exact shape of the curves depends on the assumption on the arrival rate $\lambda(x)$ but these two results hold for all cases. A decreasing variance of wages is in contradiction with the patterns in the data for new entrants in the market, immigrants and natives. Independently of the precise assumption on the difference between immigrants and natives entering the labour market, a simple search model is not compatible with the increase in the variance of wages for "stayers" over time, as documented in section 5.

1.6.3 Human Capital Model

Another competing model is based on human capital accumulation.

Let us consider the following setup: There are complementarities between worker skill and firm productivity, as in the model above. Over time, workers accumulate human capital and become more productive. A possible assumption to model the difference between natives and immigrants is that immigrants have an initial lower level of human capital but that they accu-

[28]This prediction on the monotonicity of the variance of wages only holds when considering only workers who remain in employment every year. The model is the same than the one in Manning (2000), however he finds that the patterns of the variance of wages are non-monotonic: this is due to the effect of workers who accept a job offer after an unemployment spell.

mulate human capital in the first years in the host country faster than natives. Let us assume also that the human capital function is concave: there are decreasing returns to investment in human capital.

New entrants in the market start off at the bottom of the firm distribution since they have the lowest levels of human capital. Over time, as their human capital stock increases, they gain access to better firms and the mean wages increase. As the productivity of workers increases at a decreasing rate, the job mobility rate decreases over time. The prediction on the immigrant wage catch up and on job mobility are the same than those in the model of section 4.

To derive predictions on the variance of wages, an extra assumption is needed which is that workers accumulate human capital heterogeneously. This implies that as workers accumulate human capital, the wages of a cohort become more dispersed. The variance of the wage growth also decreases over time as there are decreasing returns to human capital accumulation.

As this specific example illustrates, a human capital model can explain any set of stylized facts, if the appropriate assumptions are made. It is therefore not really testable. Distinguishing between heterogeneous accumulation of human capital and learning is an unsolved problem in the literature, and goes beyond the scope of this chapter.

Competing Theories of the Distribution of Wages

	Learning Model	Search Model	Human Capital Model
Basic Set up	Employer learning with complementarities between worker and firm type	On-the-job-search	Human capital accumulation with complementarities between worker and firm type
Immigrants and Natives	Higher initial uncertainty about immigrant productivity	Lower initial search capital for immigrants	Lower initial human capital for immigrants
Immigrants and Natives over Time		Immigrants accumulate search capital faster than natives	Immigrants accumulate human capital faster than natives
Other Features	Firms value certainty over the worker's productivity		Decreasing returns to human capital accumulation
Stylized Facts			
Immigrant wage catch up	✓	✓	✓
High but decreasing job mobility for immigrants	✓	✓	✓
Switching firms accounts for part of the catch up	✓	✓	✓
Variance of wages increases over time	✓	X	if heterogeneous accumulation of human capital
Variance of the wage growth decreases over time	✓	X	if heterogeneous accumulation of human capital

1.7 Conclusion

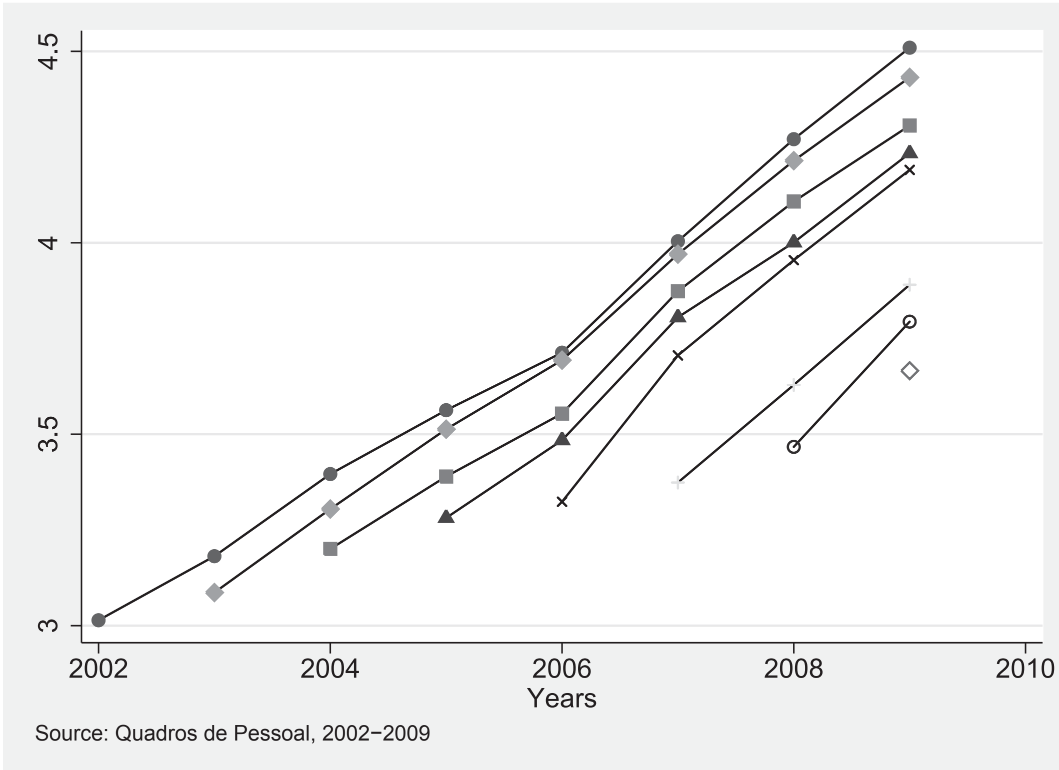
Although there is widespread evidence that immigrant wages catch-up to the wages of comparable natives with years spent in the host country, the mechanisms through which wages catch-up are not well understood. I use a unique linked employer employee panel for Portugal to study the careers of immigrants in the first years in the host country. The data allows following all workers in the private sector in the country and provides detailed information on the firms.

I show that immigrant wages catch up to the natives of the same age at a rate of 10 percentage points in 10 years. Immigrants exhibit very high job mobility rates and one third of the wage catch-up is associated to moving to better paying firms. Sorting across occupations explains a large part of the immigrant-native wage gap but changing occupations does not contribute to the catch-up. Over time, immigrants move to bigger, better paying and more productive firms. They tend to start their careers in segregated firms but the share of native co-workers increases as time goes by. The proportion of immigrants with a long term contract also increases with years spent in the labour market.

Motivated by these new stylized facts, I suggest a model of immigrant economic assimilation which highlights the role of uncertainty about immigrant productivity. Workers and firms are heterogeneous and firms value certainty over worker productivity. The model predicts that immigrants start their careers in the host country working in low productivity firms. Over time, they get access to more productive firms and move up the wage distribution. I derive additional predictions from the model on the variance of wages. In line with the model, immigrant wages become more dispersed with time spent in the host country and the increase in dispersion is associated with firm heterogeneity.

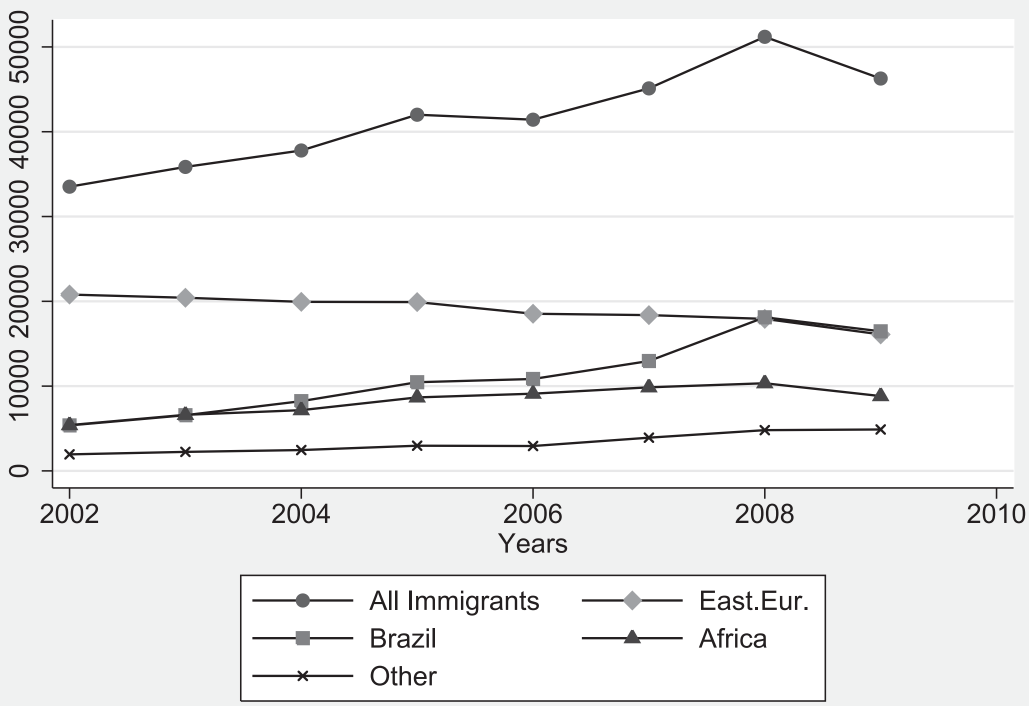
Finally, I consider two competing explanations of the immigrant wage catch-up: search and human capital accumulation. The predictions on the evolution of the variance of wages of immigrants from a simple search model are not in line with the patterns in the data. A human capital accumulation model with heterogeneous agents may be consistent with the data. Distinguishing between the predictions from a learning model and from a human capital model with heterogeneous agents is an unsolved problem in the literature, which is beyond the scope of this chapter.

Figure 1.1: Mean Hourly Wages for Immigrants by Cohort



.

Figure 1.2: Number of Immigrant Workers in the Data



Source: Quadros de Pessoal, 2002−2009

Figure 1.3: Region of Origin of Immigrants by Cohort

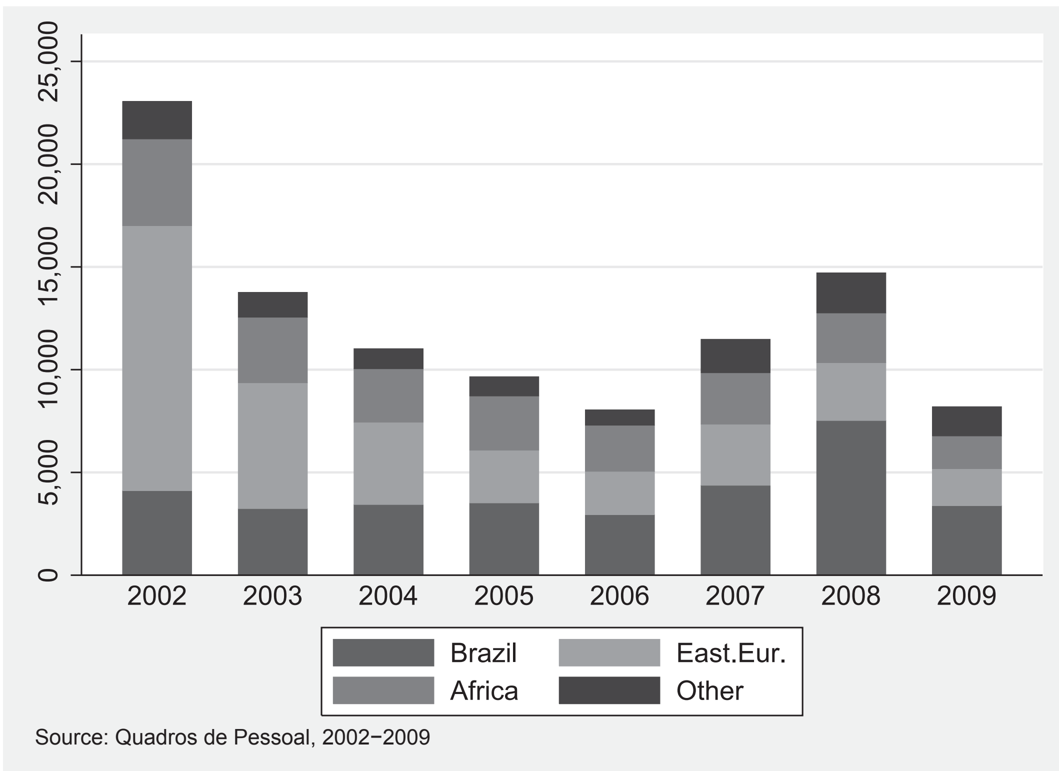
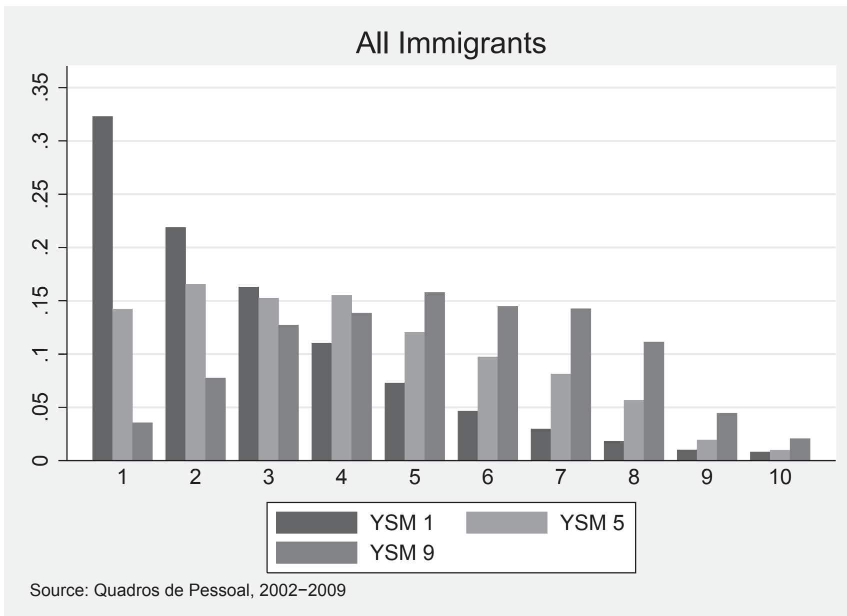
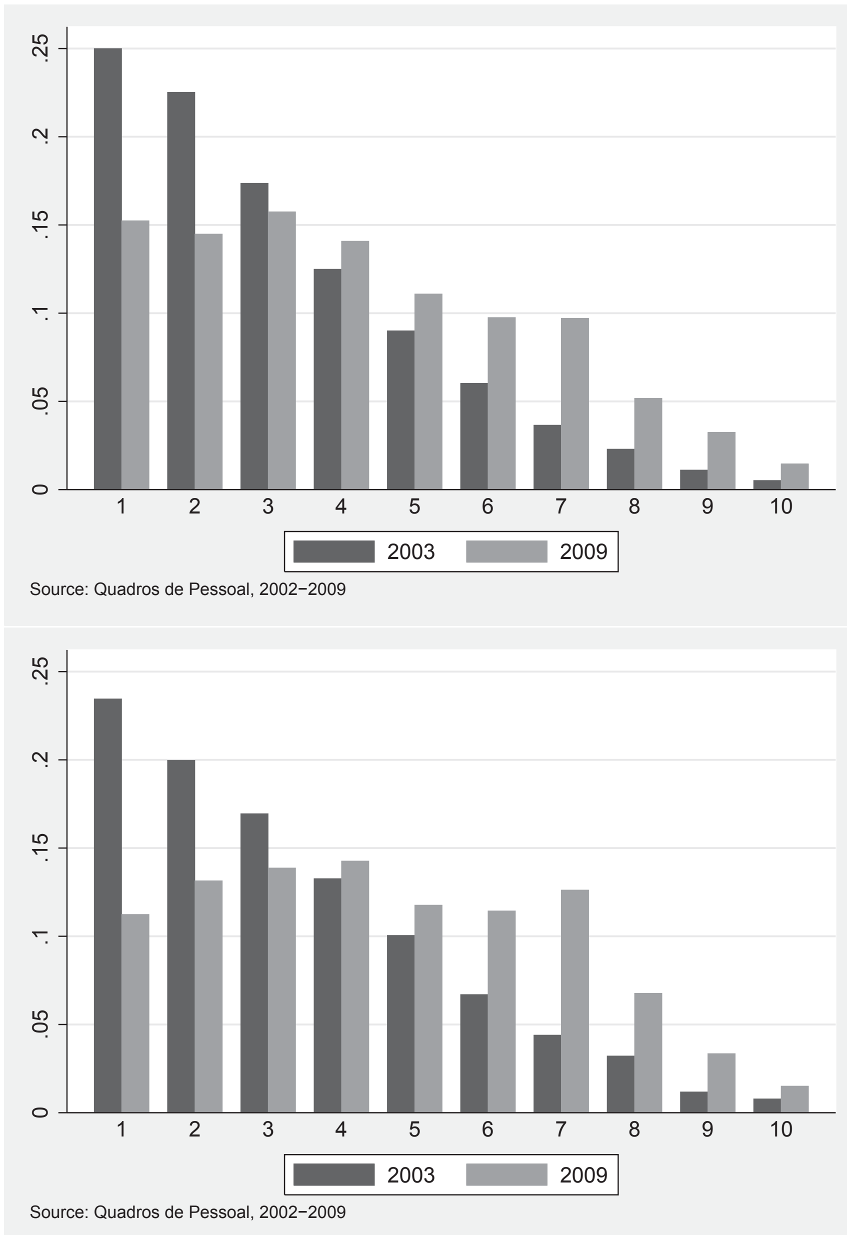


Figure 1.4: Representation of Immigrant Wages in the Distribution of Native Wages by Year since Migration



Note: The graphic illustrates the representation of immigrant wages after 1, 5 and 9 years in the country in the native wage distribution. With years spent in the country, the distribution of wages of immigrants comes closer to the one of the natives.

Figure 1.5: Representation of Immigrant Wages in the Distribution of Native Wages by Years since Migration, 2003 Cohort



Note: The top graphic is for all immigrants of the 2003 cohort and the bottom one is for immigrants of the 2003 cohort who remain in the data every year the ”stayers”. The comparison group is natives who are in the data in 2003 and natives who are in the data in 2009.

Figure 1.6: Climbing up the 'Firm Quality Ladder'

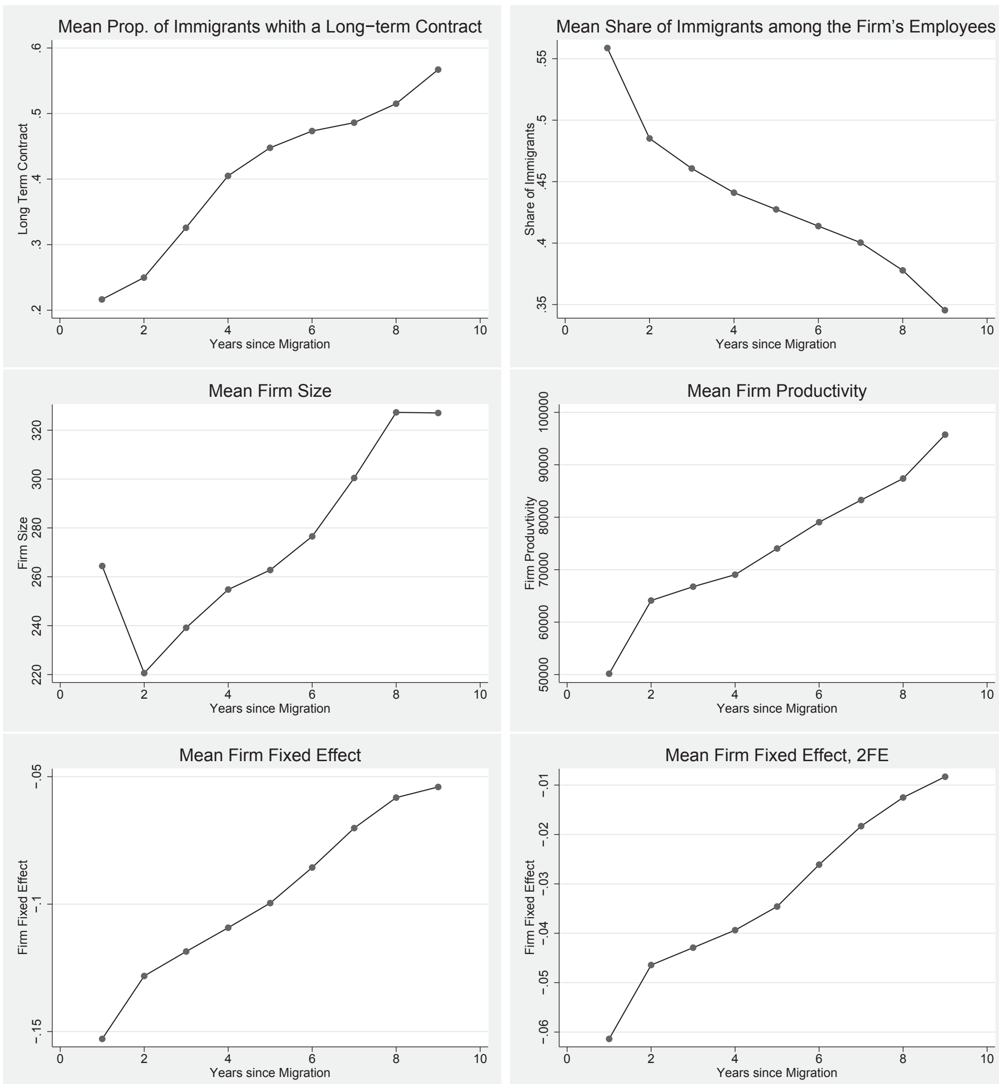


Figure 1.7: Climbing up the 'Firm Quality Ladder', 2003 Cohort

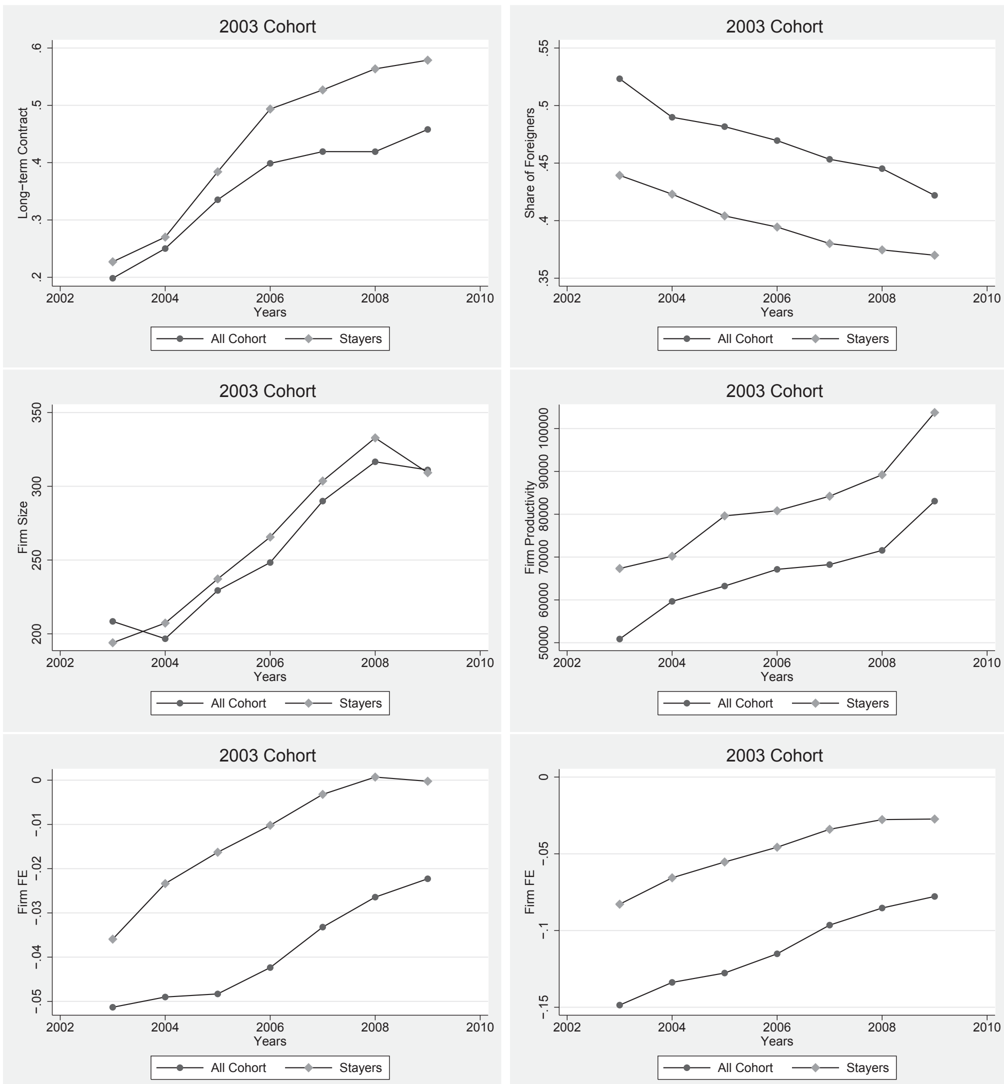
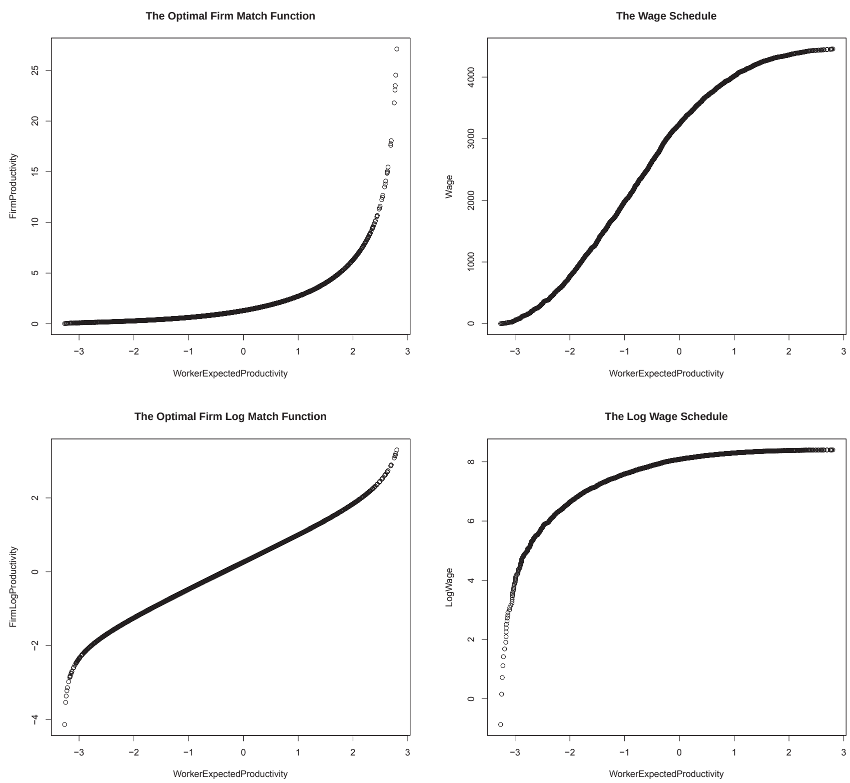
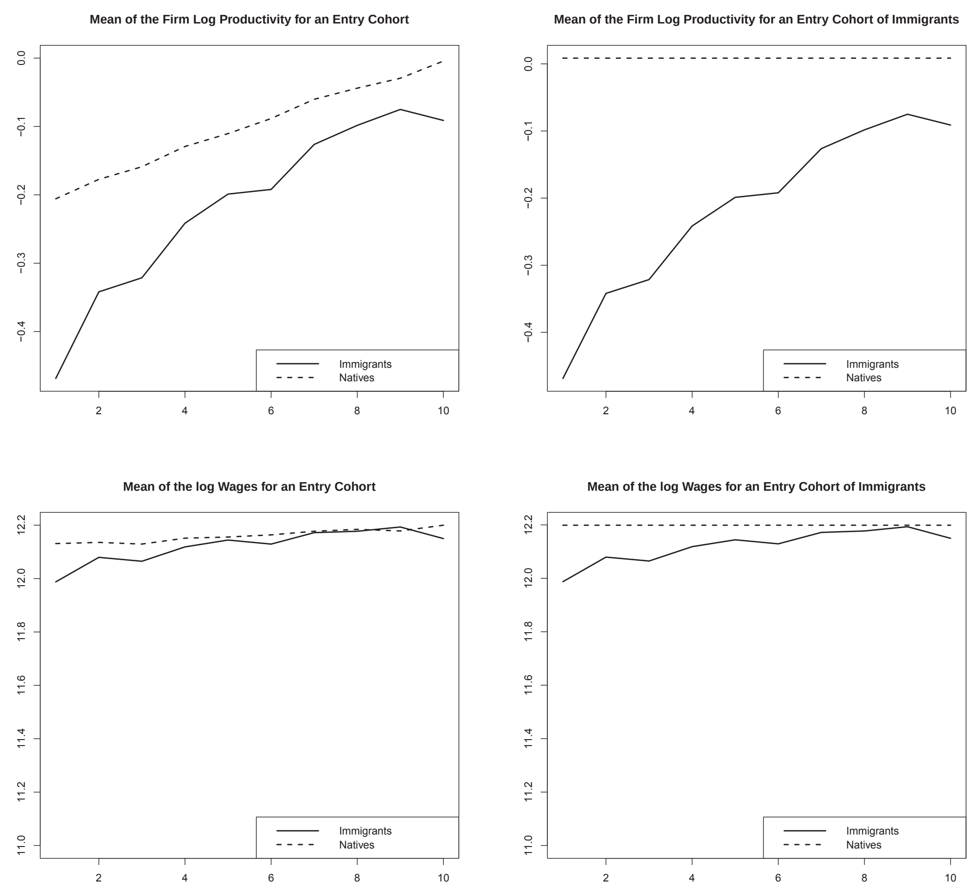


Figure 1.8: An Employer Learning Model with Firm and Worker Heterogeneity



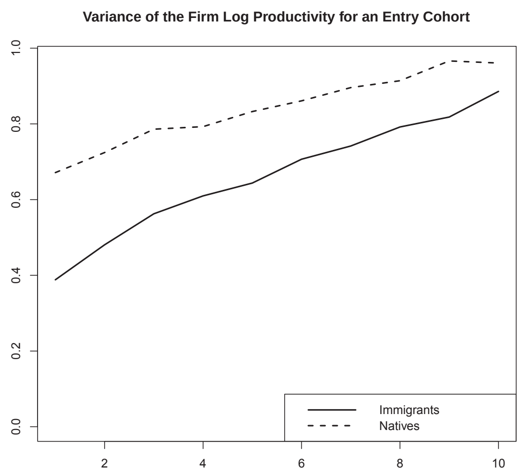
Note: The plots of the expressions derived in the model are drawn setting all means equal to 0, $\mu_q = \mu_a = \mu_s = \mu_\epsilon = \mu_c = 0$ and $\sigma_q^2 = 0.5$, $\sigma_a^2 = 1.5$, $\sigma_s^2 = 1$, $\sigma_\epsilon^2 = 30$, $\sigma_c^2 = 1$.

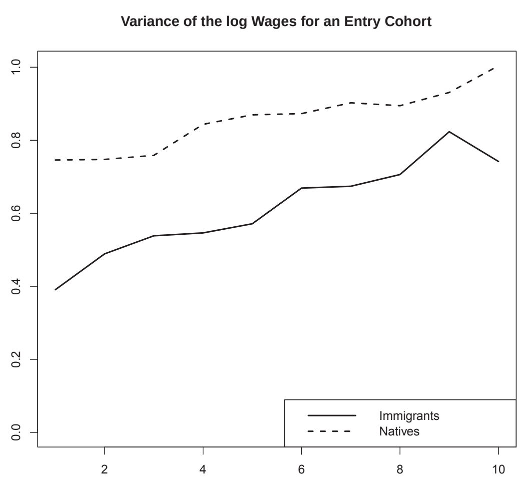
Figure 1.9: Predictions on the Mean Log Firm Productivity and the Mean Log Wages



Note: The top left and the bottom left graph compare immigrants of an entering cohort with the natives of the same cohort. The figures on the right compare immigrants of an entering cohort to all natives in the economy. The plots of the expressions derived in the model are drawn setting all means equal to 0, $\mu_q = \mu_a = \mu_s = \mu_\epsilon = \mu_c = 0$ and $\sigma_q^2 = 0.5$, $\sigma_a^2 = 1.5$, $\sigma_s^2 = 1$, $\sigma_\epsilon^2 = 30$, $\sigma_c^2 = 1$.

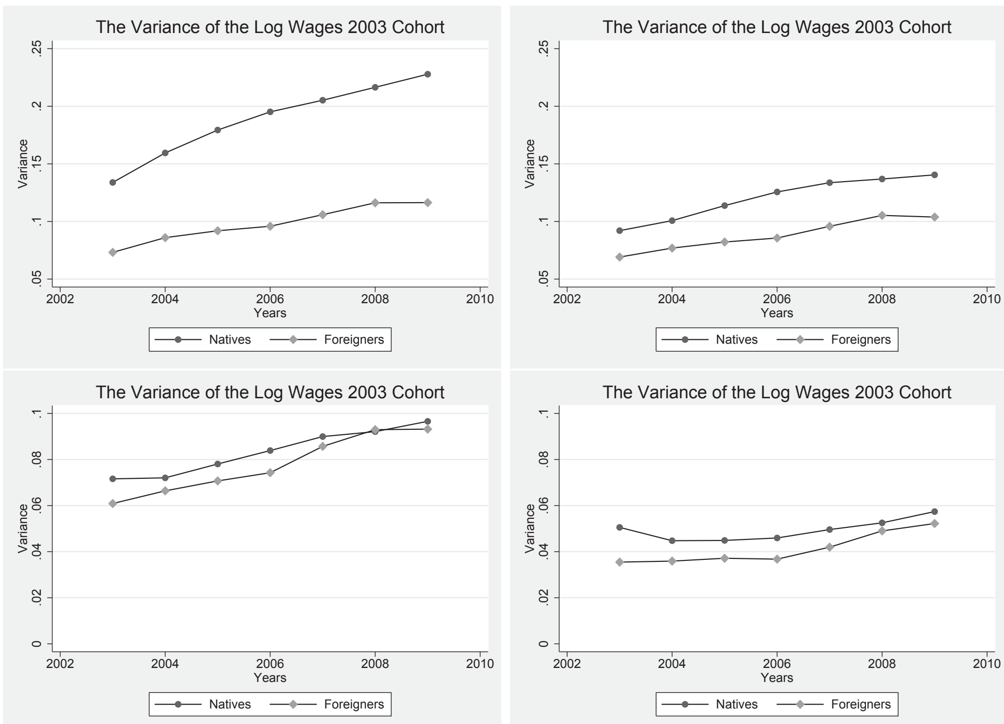
Figure 1.10: Predictions on the Variance of Log Firm Productivity and the Variance of Log Wages





Note: The plots of the expressions derived in the model are drawn setting all means equal to 0, $\mu_q = \mu_a = \mu_s = \mu_\epsilon = \mu_c = 0$ and $\sigma_q^2 = 1$, $\sigma_a^2 = 1.5$, $\sigma_s^2 = 0.5$, $\sigma_\epsilon^2 = 30$, $\sigma_c^2 = 1$.

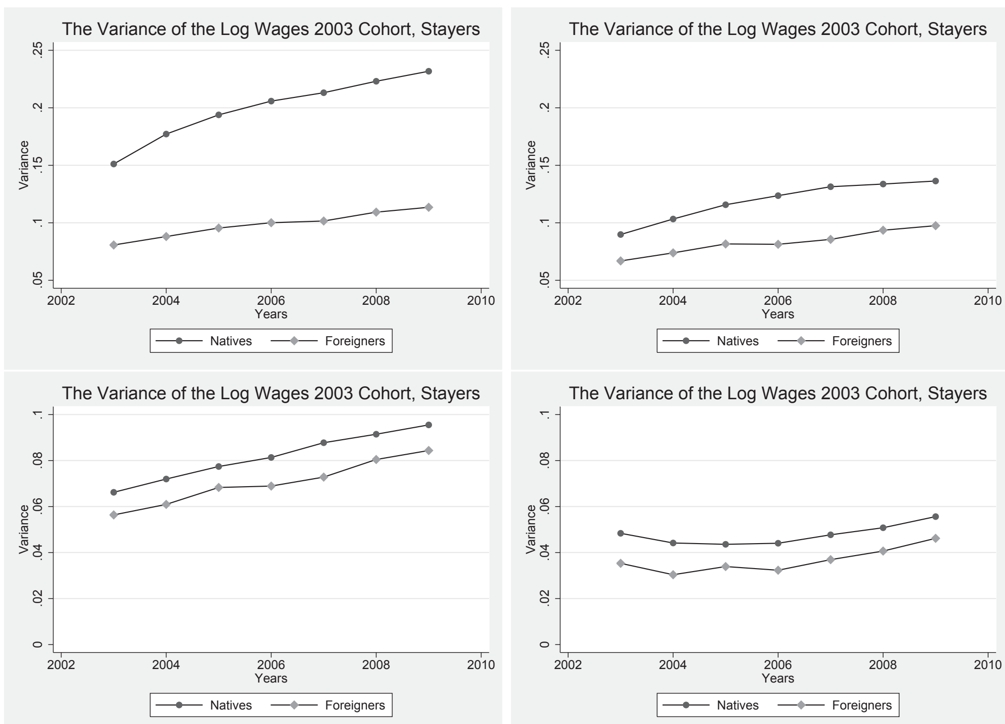
Figure 1.11: The Variance of Log Wages



Notes: The plots represent the variance of the residual by year and cohort for all natives and immigrants of the 2003 cohort estimated by least squares from the following specification: $\ln(HW)_{ijt} = origin_i * cohort_i * year_t + \epsilon_{it}$, and controlling progressively by age group and industry (top right), occupation (bottom left) and firm heterogeneity (bottom right).

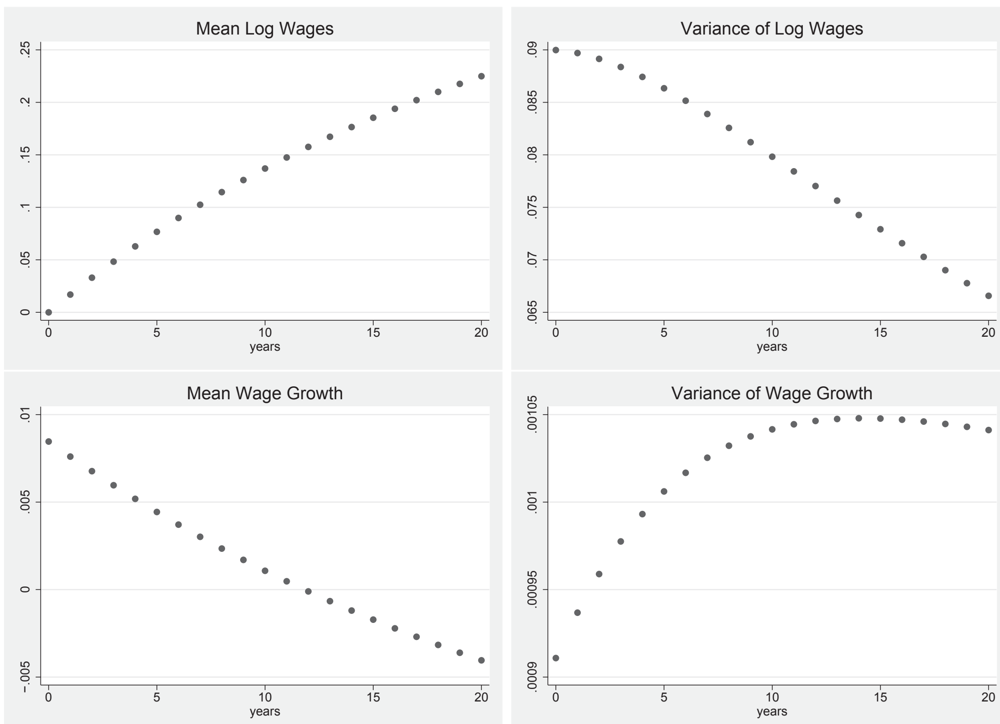
The variance of log wages is higher for immigrants than for natives and increasing for both groups over time. Firm heterogeneity explains the increase in the variance in particular in the first years in the labour market.

Figure 1.12: The Variance of Log Wages, Stayers



Notes: The plots are the same than those in figure 1.11 but consider only workers from the 2003 cohort who remain employment each year. The patterns are very similar, which show that selection out of the labour market does not affect the results.

Figure 1.13: Predictions on the Mean and Variance of Wages of a Search Model



Notes: The plots represent the patterns of the mean and variance of log wages and wage growth for an entry cohort. Workers are assumed to stay in employment every period. The probability to receive a wage offer in any given period is set to 0.1.
The decrease in the variance of wages of a cohort over time is not compatible with the stylized facts in section 5.

Table 1.1: Population Selected Means

	Natives	All Immigrants	East. Europ.	Brazil	Africa
Age	38.7	35.1	36.6	32.6	35.3
YSM	0	3.3	3.6	3.0	3.2
By Origin					
East.Eur.	0	0.49	1	0	0
Brazil	0	0.22	0	1	0
Africa	0	0.20	0	0	1
By Region					
Alentejo	0.05	0.06	0.07	0.06	0.02
Algarve	0.04	0.14	0.19	0.11	0.08
Centro	0.22	0.18	0.26	0.15	0.07
Lisboa	0.29	0.49	0.32	0.55	0.77
Norte	0.40	0.13	0.16	0.13	0.06
By Industry					
Manufacturing	0.31	0.16	0.24	0.12	0.06
Construction	0.18	0.42	0.46	0.34	0.56
Wholesale and retail trade	0.19	0.10	0.08	0.13	0.07
Hotels and restaurants	0.04	0.09	0.04	0.18	0.06
Transport, storage and communication	0.08	0.05	0.06	0.06	0.02
Real estate, renting and business activities	0.09	0.13	0.09	0.13	0.19
Number of Workers		117,964	47,279	34,913	23,810
Number of Observations	8,506,801	339,986	152,008	89,001	65,977

Notes: This table shows the mean age for natives and immigrants of the three main origin groups and the "years since migration" (YSM) for immigrants; the distribution of immigrants by origin; and the distribution of immigrants and natives by region and industry. Only recent immigrants who have entered the labour market after 2001 are considered in the analysis. All the differences in means between groups are very significantly different from 0.
Source: Quadros de Pessoal, 2002-2009.

Table 1.2: Immigrant Wage Catch up

	(1)	(2)	(3)	(4)	(5)	(6)
FG	-0.353	-0.254	-0.152			
	(0.0012)	(0.0012)	(0.0011)			
YSM	0.009	0.009	0.008	0.010	0.010	0.009
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age (quartic)	Yes	Yes	Yes	Yes	Yes	Yes
Region		Yes	Yes		Yes	Yes
Industry		Yes	Yes		Yes	Yes
Occupation			Yes			Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE				Yes	Yes	Yes
N	7,543,209	7,543,209	7,543,209	7,543,209	7,543,209	7,543,209
R^2	0.105	0.456	0.608	0.313	0.324	0.332

Notes: The dependent variable is log hourly wages. Standard errors are in parentheses.
'FG' is a dummy for foreigners. 'YSM' is the interaction between 'FG' and years since migration. 'Region' is a set of 27 dummy variables (nutse3) accounting for the region of the country the establishment is located in; 'Industry' is a set of 211 dummy variables accounting for the industry of the establishment at the 3 digit level (cae rev2.1); 'Occupation' is a set of 110 dummy variables accounting for the occupation of the individual at the 3 digit level (cnp94).
FG measures the wage gap and YSM the wage catch up. Sorting into regions, sectors and occupations explains half of the wage gap between natives and immigrants. Immigrants wages grow at a rate of approximately 1 percentage point faster than natives. The catch up is not correlated to immigrants moving industries or occupations. Estimations with and without individual heterogeneity are similar and show that the result is not driven by selection.
Source: Quadros de Pessoal, 2002-2009.

Table 1.3: Immigrant Wage Catch up by Origin Group

	East.Eur. (1)	Brazil (1)	Africa (1)	East.Eur. (4)	Brazil (4)	Africa (4)
FG	-0.377	-0.314	-0.346			
	(0.0016)	(0.0024)	(0.0024)			
YSM	0.010	0.020	0.006	0.013	0.011	0.003
	(0.0004)	(0.0007)	(0.0006)	(0.0005)	(0.0007)	(0.0007)
Age (quartic)	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE				Yes	Yes	Yes
N	7,395,761	7,334,379	7,319,209	7,395,761	7,334,379	7,319,209
R^2	0.100	0.148	0.096	0.315	0.315	0.315

Notes: The dependent variable is log hourly wages. Standard errors are in parentheses.
See table 1.2 for the definitions of the variables used.
The wage gap upon entry is highest for immigrants from Eastern Europe and lowest for Brazilians. The wage catch up rate accounting for individual fixed effects is above 1 percentage point for Brazilians and Eastern Europeans but immigrants from Africa lag substantially behind.
Source: Quadros de Pessoal, 2002-2009.

Table 1.4: Job Mobility

	(1)	(2)	(3)	(4)	(5)
FG	0.211	0.191	0.117	0.083	0.080
	(0.0013)	(0.0013)	(0.0013)	(0.0013)	(0.0013)
YSM	-0.021	-0.017	-0.010	-0.008	-0.007
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.0003)
Age		-0.036	-0.022	-0.020	-0.019
		(0.0011)	(0.0011)	(0.0011)	(0.0011)
Quartic in Age		Yes	Yes	Yes	Yes
Tenure			-0.023	-0.019	-0.019
			(0.0001)	(0.0001)	(0.0001)
Cubic in Tenure			Yes	Yes	Yes
Hourly wage			-0.001	-0.001	-0.001
			(0.00003)	(0.00003)	(0.00005)
Region				Yes	Yes
Industry				Yes	Yes
Occupation					Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Cst	0.071	0.582	0.400	0.368	0.389
N	5,731,442	5,731,442	5,731,442	5,731,442	5,731,442
R^2	0.010	0.011	0.049	0.081	0.082

Notes: The dependent variable is 1 if the worker will be working in a different firm next period, 0 if he stays with the same employer. Standard errors are in parentheses.
See table 1.2 for the definitions of the variables used.
The probability of changing employers is higher for immigrants than for natives. This probability declines with years spent in the labour market.
Source: Quadros de Pessoal, 2002-2009.

Table 1.5: Immigrant Wage Catch-up and Firm Fixed Effects

	(1)	(2)	(3)	(4)	(5)
YSM	0.010	0.010	0.009	0.006	0.006
	(0.0003)	(0.0003)	(0.0003)	(0.0002)	(0.0002)
Age (quartic)	Yes	Yes	Yes	Yes	Yes
Region		Yes	Yes		
Industry		Yes	Yes		
Occupation			Yes		Yes
Year FE	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes
Firm FE				Yes	Yes
N	7,543,209	7,543,209	7,543,209	7,543,209	7,543,209
R^2	0.313	0.324	0.332	0.945	0.945

Notes: The dependent variable is log hourly wages. Standard errors are in parentheses.
See table 1.2 for the definitions of the variables used.
These regressions control for firm fixed effects in the wage catch-up estimations. Comparing the estimates for γ in this table and table 1.2 shows that the coefficient decreases from 1ppt to 0.6ppt, or from 0.9 to 0.6ppt when controlling also for occupations. Changing firms accounts for a third of the immigrant wage catch-up.
Source: Quadros de Pessoal, 2002-2009.

Table 1.6: Immigrant Wage Catch-up and Firm Fixed Effects by Origin Group

	East.Eur. (1)	East.Eur. (2)	Brazil (3)	Brazil (4)	Africa (5)	Africa (6)
YSM	0.009	0.008	0.006	0.006	-0.001	-0.001
	(0.0002)	(0.0002)	(0.0004)	(0.0004)	(0.0004)	(0.0004)
Age (quartic)	Yes	Yes	Yes	Yes	Yes	Yes
Region						
Industry						
Occupation		Yes		Yes		Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Individual FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE Yes	Yes	Yes	Yes	Yes	Yes	
N	7,395,761	7,395,761	7,334,379	7,334,379	7,319,209	7,319,209
R^2	0.945	0.946	0.946	0.946	0.946	0.946

Notes: The dependent variable is log hourly wages. Standard errors are in parentheses.
See table 1.2 for the definitions of the variables used.
Comparing the estimates in this table to those in table 1.3, the estimated γ decreases from 1.3ppt to 0.9ppt for Eastern Europeans and from 1.1ppt to 0.8ppt for Brazilians. A third of the wage catch-up occurs when changing firms for these two groups. All of the wage catch-up for immigrants from Africa occurs when changing firms as the estimated γ is close to 0 in this estimation.
Source: Quadros de Pessoal, 2002-2009.

Table 1.7: The Variance of the Wage Growth

FG	0.0072	0.0056
	(0.0009)	(0.0022)
EXP	-0.0025	-0.0026
	(0.0009)	(0.0002)
YSM		0.0004
		(0.0005)
Year FE	Yes	Yes
N	56	56
R^2	0.808	0.807

Notes: The dependent variable is the variance of the residual estimated from equation (1.4) for a origin-cohort at each calendar year. 'FG' is a dummy for foreigners.'EXP' are the years of experience in the labour market and 'YSM' is the interaction between 'FG' and 'FG'. Standard errors are in parentheses.
In line with the model, the variance of the wage growth is higher for immigrants than for natives and both decrease over time.

1.A Data Appendix

The data used in the paper is a linked employer-employee panel. The information is collected yearly by the Ministry of Labour in Portugal and the questionnaire is compulsory for all firms that employ at least one wage earner. All firms, establishments and workers have a unique identifer. An observation is a worker-firm match in a given year.

1.A.1 Building the Panel

Pooling all observations from 2000 until 2009, the initial data has 27m observations. In this section, I present details on the checks which were made to construct the panel adequately.

Workers are identified by their social security number. I start by identifying workers with an invalid social security number. In most cases, an invalid social security number is coded with a 0. It may be that immigrants as they first appear in the data have not been attributed a social security number. Deleting these observations would lead to ignoring information related to the first year in the panel and underestimating the number of years spent in formal employment. In order to recover the potential first year in the panel of immigrant observations, I match observations with a 0 social security number with observations in the following year by gender, date of birth, nationality (Portuguese or not), and firm identifier. The profile of an individual with an invalid social security number may hence only be recovered if he works in the same firm the following year. This correction allows to recover 240,595 observations. 478,347 observations still have an invalid social security number after this correction and are hence deleted.

I exclude workers who have several jobs at some point in their careers. The paper focuses on the career of immigrants and in particular on the importance of job mobility. The cases where workers have multiple jobs would need special attention. I discard these profiles: 2.8m observations in total, 18% of immigrant and 10% of native observations.

I then check for basic inconsistencies in the workers' profiles. Individuals for whom there are changes in gender or in immigrant status over time are allocated the gender and immigrant status reported more than half the times. 50,208 and 90,695 observations are dropped when after this correction no conclusion is reached. Individuals with a decreasing age profile are also identified and dropped from the analysis: 305,661 observations, 1.3% of native and 1.4% of immigrant observations. A last profile consistency check concerns wages. Individuals with an inconsistent wage growth profile (log hourly wage growth smaller than -0.5 or bigger than 2) are deleted[29]. In total, 1,073,426 observations were deleted: 4.6% of native and 5.9% of immigrant observations.

1.A.2 Sample Selection

For the analysis, I use only a sub-sample of individuals from the full panel. I restrict the analysis to men, as the careers of women would need a separate analysis. 44% of native observations are

[29] This correction follows Cardoso (2005)

from female workers but only 35% of immigrants.[30] This leads to discarding 8.4m observations. Only individuals working in the mainland of Portugal are considered. Workers who work in the islands (Madeira, Aores) at some point of their career are excluded: 3% of immigrant and 4.7% of native observations. The data has a low coverage of agriculture, the whole industry is hence excluded. 5% of immigrant observations and 2.5% of native observations are deleted. Family workers and self-employed workers were dropped from the sample, only wage earners were considered to make wage progression comparisons meaningful. This accounts to excluding 13% of native but only 4% of immigrant observations. Part-time workers are also excluded, which accounts to 3.8% of native and 3.3% of immigrant observations.

1.A.3 Immigrant Cohorts and Origin Groups

Origin Groups

I exclude immigrants from the EU15 from the analysis. These immigrants benefit from the same conditions in the labour market than native workers and have very different characteristics than the other immigrant groups. The three main immigrant origins are Eastern and South-Eastern Europe, the former Portuguese African colonies (African Countries of Portuguese Official Language), and Brazil. The residual group represents less than 10% of the total number of immigrant observations in the data. The countries considered in the Eastern and South-Eastern Europe group are Slovakia, Poland, the Czech Republic, Hungary, Slovenia, Latvia, Estonia, Lithuania, Romania, Russia, Moldova, Ukraine and the former Yugoslavia. The countries belonging to the PALOP are Cape Verde, Mozambique, Angola, Guinea Bissau and So Tom and Prncipe. Similarly to the consistency checks above, workers who exhibit changes in origin are identified and attributed the origin declared over half the times. Workers for which no conclusion may be drawn have the origin variable set to missing.

Cohorts

The paper focuses on immigrants from the new immigration wave to Portugal. I consider only immigrants first tracked in the data after 2001. All immigrants already in the data in 2000 are dropped from the sample.[31] This amounts to dropping approximately 7% of all immigrant workers in the data between 2000 and 2009.

The information on the date of arrival in the country is not available. The first time an immigrant is observed in the panel is used as a proxy. This captures the first time the worker is in formal employment, since the data set covers all wage earners in the private sector in Portugal. The cohort is defined as the first year the immigrant appears in the data and the years since migration are calculated as the difference between the calendar year and the cohort year. Moreover, a correction using the tenure variable is made to this calculation. Immigrants

[30] 71% of the observations for Eastern European immigrants are from male workers.

[31] I use the whole panel from 1987 to check whether immigrants who are classified as new immigrants are in the data at an earlier point in time. I find that less than 10% of the immigrant observations considered in the analysis can be tracked before 2000. The correction using only the year 2000 is thus an acceptable correction.

in their first year in the panel are assumed to have arrived in the country at their arrival in the firm. Consider for example an immigrant who is first observed in the data in 2003, but whose tenure indicates that he has already worked in the same firm for two years. He is considered to have been in the country since 2001.

Tenure for the purposes of the analysis refers to the time spent working in the same firm. If an individual's tenure is reported decreasing in the same firm (this may, for instance, be due to a change in contract) then the number of years considered as tenure is the time since the beginning of the first contract with the firm. After this correction, if there are still different dates of entry in the firm across the years, the correct value is considered to be the one taken over half of the times. When no conclusion may be reached after these corrections, the individual's tenure is set to missing.

1.A.4 Specific Data Issues and Robustness Checks

Dealing with the Missing Year of 2001

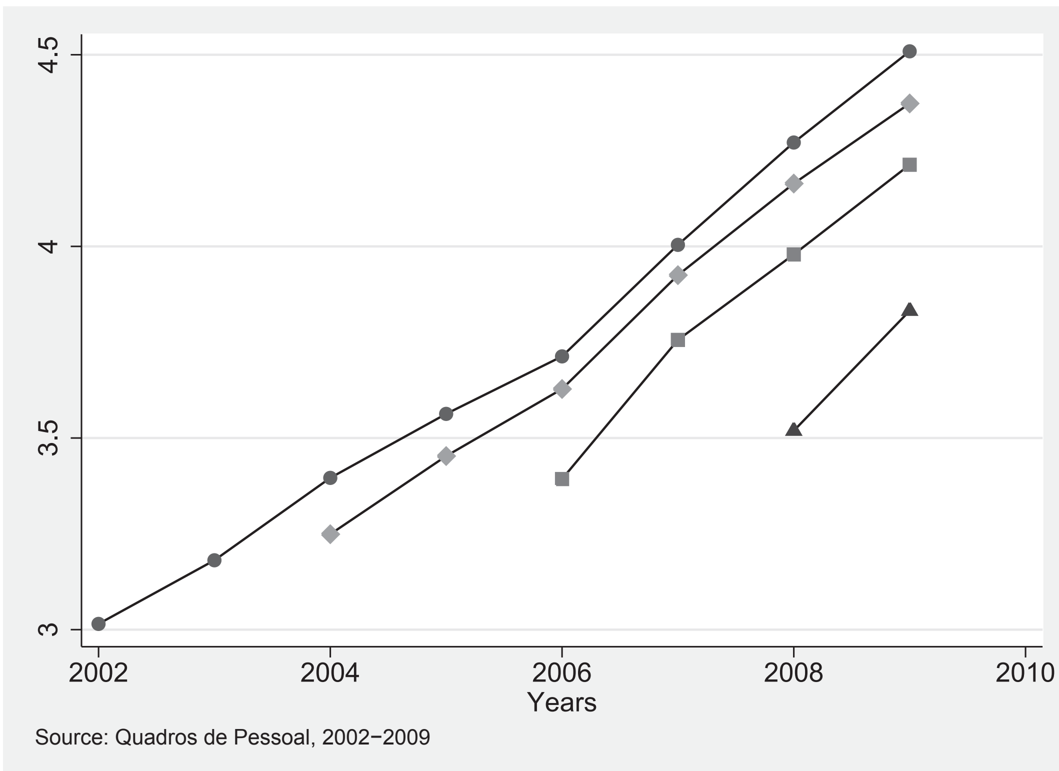
The data for the year of 2001 is not available. The analysis in the paper starts the analysis in 2002. I use the year 2000 to identify immigrants who are already in the country in 2000.

The missing data for 2001 poses several challenges and particularly in defining the immigrant cohorts. Workers who first start working in 2001 but change employers between 2001 and 2002 are allocated to the 2002 cohort; whereas workers who remain with the same employer are allocated to the 2001 cohort. A fraction of the 2002 cohort is made up of workers who have an extra year of experience in the Portuguese market. If movers are positively selected, these extra workers are also the "best" workers of the 2001 cohort. Figure 1.1 in section 2 shows that the pattern of the mean wages for the immigrants of the 2002 cohort does not appear to be very different from the one of the other cohorts.

I perform another robustness check: I redefine cohorts as two-year instead of one-year cohorts. I classify immigrants who first appear in the data in 2001[32] or 2002 as belonging to the first cohort; immigrants who are first tracked in 2003 or 2004 belong to the second cohort, etc. The years since migration variable is re-calculated accordingly. Figure 1.14 plots the mean wages for these newly defined cohorts over time. I re-estimate all the main empirical specifications in section 3 using this two-year cohorts and find little difference in the main results.

[32]The correction using the tenure variable allows to allocate immigrants to the 2001 cohort even if the data for 2001 is missing.

Figure 1.14: Mean Hourly Wages for Immigrants by Two-year Cohort



Dealing with the Changes in the Industry Classification in 2007

In 2007, the classification used for the industries in Portugal changed.[33] I use the 'old' classification in all the analysis. I do not use the official table created by the Portuguese Institute of Statistics to convert the new classification into the old one. This would lead to over-estimate the frequency with which workers change industries as the classifications are very different. In fact, when I consider firms which existed in 2006 and 2007 and which are all classified in a given 'new' industry in 2007, there is a very large dispersion in the industries the firms belonged to in 2006. For all firms which can be tracked in the data before 2007, I assign the industry observed before 2007 in the 'old' classification to the observations from 2007 until 2009. I then drop all observations belonging to firms who enter the market in 2007 and later.

One may worry that excluding these observations may lead to a selection bias. I perform two robustness checks to address this concern. First, I re-do all the main estimations in the main empirical section which do not use the industry dummies including the firms entering the market after 2007; second, I re-do the analysis excluding the years after 2007. I find little difference in the main results.

1.B Model Appendix

1.B.1 The Variance of the Expected Productivity for a Cohort

I calculate the variance of the distribution of the expected worker productivity ($\mu_{x,k}$) for a given cohort after x years in the labour market.

For immigrants:

$$\mu_{x,fg} = \frac{\sigma_\epsilon^2}{x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2}(q+\mu_a+\mu_s) + \frac{\sigma_a^2+\sigma_s^2}{x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2}\sum_{l=0}^{x-1} y_l$$

Replacing $\sum_{l=0}^{x-1} y_l = x(a+s+q) + \sum_{l=0}^{x-1}\epsilon_l$,

$$\mu_{x,fg} = q + \frac{\sigma_\epsilon^2(m_s+m_a) + (\sigma_a^2+\sigma_s^2)(x(a+s)+\sum_{l=0}^{x-1}\epsilon_l)}{x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2}$$

q, a, s and ϵ are independent random variables. The variance of the above expression is hence:

$$V\left(\mu_{x,fg}|I_x\right) = \sigma_q^2 + \frac{x^2(\sigma_a^2+\sigma_s^2)^3}{(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2)^2} + \frac{x\sigma_\epsilon^2(\sigma_a^2+\sigma_s^2)^2}{(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2)^2}$$

Similarly for natives:

$$\mu_{x,nat} = \frac{\sigma_\epsilon^2}{x\sigma_a^2+\sigma_\epsilon^2}(q+\mu_a+s) + \frac{\sigma_a^2}{x\sigma_a^2+\sigma_\epsilon^2}\sum_{l=0}^{x-1} y_l$$

[33] The classification until 2006 is cae rev 2.1 and from 2007 it is cae rev 3.

$$V\left(\mu_{x,nat}|I_x\right) = \sigma_q^2 + \sigma_s^2 + \frac{x^2(\sigma_a^2)^3}{(x\sigma_a^2+\sigma_\epsilon^2)^2} + \frac{x\sigma_\epsilon^2(\sigma_a^2)^2}{(x\sigma_a^2+\sigma_\epsilon^2)^2}$$

1.B.2 The Distribution of the Changes in Risk-adjusted Expected Productivity for a Cohort

The change in worker risk-adjusted expected productivity from year $x-1$ to year x

$$\Delta_{x,k} = \left(\mu_{x,k} - \frac{1}{2}\sigma^2_{x,k}\right) - \left(\mu_{x-1,k} - \frac{1}{2}\sigma^2_{x-1,k}\right)$$

follows a normal distribution. I calculate the mean and variance of its distribution for immigrants and natives of a given cohort.

The mean of $\Delta_{x,k}$ is the mean of $-\frac{1}{2}\sigma^2_{x,k} + \frac{1}{2}\sigma^2_{x-1,k}$ as the mean of expected productivity over a cohort of workers is constant over time.

For immigrants:

$$M(\Delta_{x,fg}) = \frac{1}{2}\frac{\sigma_\epsilon^2(\sigma_s^2+\sigma_a^2)^2}{(\sigma_\epsilon^2 + x(\sigma_a^2+\sigma_s^2))(\sigma_\epsilon^2 + (x-1)(\sigma_a^2+\sigma_s^2))}$$

For natives:

$$M(\Delta_{x,nat}) = \frac{1}{2}\frac{\sigma_\epsilon^2\sigma_a^4}{(\sigma_\epsilon^2 + x\sigma_a^2)(\sigma_\epsilon^2 + (x-1)\sigma_a^2)}$$

The variance of $\Delta_{x,k}$ is the variance of $\mu_{x,k} - \mu_{x-1,k}$ as $\sigma^2_{x,k}$ is constant across workers of the same cohort in all periods in the market.

For immigrants:

$$\mu_{x,fg} - \mu_{x-1,fg} = q + \frac{\sigma_\epsilon^2(m_s+m_a) + (\sigma_a^2+\sigma_s^2)(x(a+s) + \sum_{l=0}^{x-1}\epsilon_l)}{x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2}$$

$$-q - \frac{\sigma_\epsilon^2(m_s+m_a) + (\sigma_a^2+\sigma_s^2)((x-1)(a+s) + \sum_{l=0}^{x-2}\epsilon_l)}{(x-1)(\sigma_a^2+\sigma_s^2)-\sigma_\epsilon^2}$$

Re-writing:

$$\frac{\left((x-1)(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2\right)\left(\frac{\sigma_\epsilon^2(m_s+m_a)+(\sigma_a^2+\sigma_s^2)(x(a+s)+\sum_{l=0}^{x-1}\epsilon_l)}{x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2}\right)}{\left(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2\right)\left((x-1)(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2\right)}$$

$$-\frac{\left(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2\right)\left(\frac{\sigma_\epsilon^2(m_s+m_a)+(\sigma_a^2+\sigma_s^2)((x-1)(a+s)+\sum_{l=0}^{x-2}\epsilon_l)}{(x-1)(\sigma_a^2+\sigma_s^2)-\sigma_\epsilon^2}\right)}{\left(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2\right)\left((x-1)(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2\right)}$$

Noting that a, s and ϵ are independent random variables, collecting terms in a, s and ϵ, and calculating the variance:

$$V(\Delta_{x,fg}) = \sigma_\epsilon^2 \frac{((\sigma_a^2+\sigma_s^2)^2(x-1)+(\sigma_a^2+\sigma_s^2)\sigma_\epsilon^2)^2+(\sigma_a^2+\sigma_s^2)^4(x-1)+\sigma_\epsilon^2(\sigma_a^2+\sigma_s^2)^3}{[(x(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2)((x-1)(\sigma_a^2+\sigma_s^2)+\sigma_\epsilon^2)]^2}$$

Similarly for natives:

$$V(\Delta_{x,nat}) = \sigma_\epsilon^2 \frac{(\sigma_a^4(x-1)+\sigma_a^2\sigma_\epsilon^2)^2+\sigma_a^8(x-1)+\sigma_\epsilon^2\sigma_a^6}{[(x\sigma_a^2+\sigma_\epsilon^2)((x-1)\sigma_a^2+\sigma_\epsilon^2)]^2}$$

Chapter 2

The Labour Market Integration of Immigrants and their Children

2.1 Introduction

As the children of the post-war guest workers arrive on the labour markets across Western Europe, their integration becomes a major policy concern for the OECD countries. The integration of the children of immigrants, or second generation[1], has little in common with the immigrants' integration process. These are children born and raised in the host country. Most of the issues raised when dealing with the integration of immigrants such as the learning of the host country language, the relative quality of schooling in the home country or the lack of an extended social network are at odds in the context of the second generation. However, there is increasing evidence that in Europe the children of immigrants lag behind the children of natives in educational achievement and in labour market outcomes[2]. The focus of this chapter is to understand the source of some of these differences at the European level.

One of the first difficulties that arise when comparing the outcomes of children of immigrants and children of natives[3] is that immigrant parents often have a lower educational background than the average native parents. It is a well known fact that the outcomes of children are strongly influenced by parental background, as measured by parents' education, occupation or income. It is therefore important to compare children of natives and immigrants with similar family characteristics to understand to which extent the gap in outcomes may be linked to background. If immigrants have on average a lower background than natives, an immigrant child may not be expected to do as well as the average child of native-born parents. A few papers have addressed the impact of parental background on the second generation performance in individual European countries, such as Riphahn (2003) for Germany, or Van Ours and Veenman (2003) for the Netherlands. Rigorous cross-country comparisons, based on these papers, are difficult

[1]The second generation is defined in this context as the children born in the host country who have at least one foreign-born parent.

[2]See Heath et al. (2008) for a recent literature review.

[3]The term "natives" refers to individuals born in the host country. The TIES network uses the term "comparison group" instead.

to make since the methodology used and the outcomes measured are hardly comparable.

The first contribution of the chapter is to address the impact of background differences between immigrant and native families at the European level. I take advantage of a recent survey of children of immigrants from Turkey, Morocco and ex-Yugoslavia, in 15 European cities, The Integration of the European Second Generation (TIES) project, which has very detailed information on individual and family history. In particular, I focus on the impact of parental background on the educational, labour market and marriage market outcomes of the three second generation groups. All the analysis is done relatively to a comparison group of children of natives living in the same cities. Educational achievement and earnings are commonly used measures of economic assimilation. Second generation inter-ethnic marriage has been shown to be highly correlated to labour market outcomes in the US, as in Card et al. (2000), and is often looked upon as the ultimate measure of social assimilation. Marrying a native has also been shown by Furtado and Theodoropoulos (2009) to have a positive impact on labour market outcomes of the US second generation. The aim here is to take advantage of the detailed family information to determine which parental characteristics are correlated with a higher probability of the child marrying a native. I show that parental background accounts for a big part of the gap in educational and labour market outcomes between children of natives and children of immigrants. Parental characteristics are also shown to be important predictors of the extent to which second generation children marry within the ethnic group.

The long term consequences of immigration for the host country depend greatly on the intergenerational mobility of immigrants. Immigrants outcomes get closer to the natives' average during their life cycle but have been shown not to converge, see for example in the US case Lubotsky (2007). A more important question is whether the outcomes of their children and grandchildren can be expected to converge to those of the offspring of natives. It is a well established fact that the outcomes of parents and children are correlated but that there is regression towards the mean between generations. For instance, children of parents in the lower tail of the income distribution will have on average a lower income than the mean but will be closer to the mean than the parents were. In terms of intergenerational assimilation, it is important to know to which extent there is regression towards the mean in immigrant families. It is a priori not clear whether the transmission should be higher or lower for immigrants than for natives. On the one hand, immigrants suffer disadvantages that their children do not, as in most cases they completed their education in the home country and are not fluent in the host country language. This could lead to a higher mobility for children of immigrants than children of natives. On the other hand, cultural factors have been shown to have an impact on labour market outcomes and could slow the convergence of immigrant descendants' outcomes to the natives'. Fernandez and Fogli (2009), for instance, show for the US that second generation fertility and female labour market participation are closely correlated to those of the country of origin of the immigrant parents, making the convergence towards the native mean slower.

The second contribution of this chapter is to quantify the average intergenerational transmission of human capital for immigrant and native families at the European level using different measures of educational achievement and occupational status. Some recent papers have looked at this issue for different European countries, such as Bauer and Riphan (2007) for Switzerland, Gang and Zimmermann (2000) for Germany, or Hammarstedt and Palme (2006) for Sweden. The approach here is different so that the analysis is made at the European level for three second generation groups, the children of Turkish, Moroccan and Yugoslav immigrants. I show that the intergenerational transmission of educational outcomes and occupational status is on average lower for immigrant than native families.

The remainder of the chapter is organized as follows. Section 2 offers a short introduction to the Turkish, Moroccan and Yugoslav migration to Europe and describes the data used for the analysis; Section 3 studies the outcomes of the second generation in the educational, labour and marriage markets; Section 4 quantifies the intergenerational transmission of human capital in immigrant families; Section 5 concludes and points the direction to further research.

2.2 Context and Data

The Integration of the European Second Generation (TIES)[4] is a comparative research project on the outcomes of the children of immigrants from Turkey, Morocco and ex-Yugoslavia in major European cities. The project is based on an international survey that has been administered to approximately 10,000 individuals aged 18 to 35 years old in 15 cities in 8 countries (Vienna and Linz in Austria, Brussels and Antwerp in Belgium, Paris and Strasbourg in France, Berlin and Frankfurt in Germany, Amsterdam and Rotterdam in The Netherlands, Madrid and Barcelona in Spain, Zrich and Basel in Switzerland and Stockholm in Sweden). The main focus of the project is on the offspring of Turkish immigrants: children of Turkish immigrants have participated in the survey in all countries except Spain; whereas children of immigrants from Morocco have been selected in Belgium, The Netherlands and Spain; and children of immigrants from ex-Yugoslavia have been surveyed in Germany, Austria and Switzerland.

The main Turkish, Moroccan and Yugoslav migration flows to Western Europe started in the post-war period and were driven by the area's unprecedented growth. At the beginning of the 1960s, Western Germany, soon followed by other European countries, signed the first guest-workers agreements. Immigrants, mainly from Turkey, would come to Western Europe for a short period to help Western economies deal with labour shortages and would return to the home countries with new skills to enhance the countries' industrialization processes. Turkey signed agreements with Austria, Belgium, The Netherlands, France, and Sweden. Morocco

[4]The TIES survey was carried out by survey bureaus under supervision of the nine national TIES partner institutes: Netherlands Interdisciplinary Demographic Institute (NIDI), Institute for Migration and Ethnic Studies (IMES), University of Amsterdam in the Netherlands; the Institute for Social and Political Opinion Research (ISPO), University of Leuven in Belgium; the National Institute for Demographic Studies (INED) in France; the Swiss Forum for Migration and Population Studies (SFM), University of Neuchtel in Switzerland; the Centre for Research in International Migration and Ethnic Relations (CEIFO), University of Stockholm in Sweden; the Institute for Migration Research and Intercultural Studies (IMIS), University of Osnabrck in Germany, the Institute for the Study of Migration (IEM), Pontifical Comillas University of Madrid in Spain, and the Institute for European Integration Research (EIF), Austrian Academy of Sciences in Austria. For more information on the TIES project, see http://www.tiesproject.eu/index.php/lang=en.

signed similar agreements in the 1960s with West Germany, France, Belgium and The Netherlands. Yugoslavia also participated in the late 1960s in guest worker programs with Austria and Germany. Although these waves of migration were supposed to be temporary and host countries tried to put a stop to immigration during the mid 1970s downturn, most guest workers settled in Western Europe. During the 1980s and 1990s, European host countries saw their stocks of Turkish, Moroccan and Yugoslav immigrants increase sharply through family reunifications. In the early 2000s, Moroccan descendants amount to approximately 300,000 in The Netherlands and 200,000 in Belgium. Turkish descendants in Western Europe are estimated at more than 3 million. The offspring of immigrants from the successor countries of the former Yugoslavia account for approximately 1 million people in Germany and close to 400,000 in Switzerland. The timing of the TIES survey coincides with the arrival of the children of immigrants in the European labour markets.

Several aspects make the TIES project a unique and extremely valuable instrument for the study of the second generation in Europe. Firstly, the data is extremely detailed and contains information on virtually all aspects of economic, social and political integration (education, labour markets, income, parental background, housing and neighbourhood, social relations and political participation, etc.). The fact that the questionnaire was the same in the different countries makes European comparisons possible. Secondly, and perhaps, more importantly, a comparison group, consisting of children of parents born in the host countries, was also taken into account, making the comparison between children of natives and children of immigrants possible at the city level. The project should allow not only to compare, for instance, children of Turkish immigrants in Switzerland and in Germany but also relative differences between the outcomes of children of Turkish nationals and children of natives in the different cities.

For the analysis, I use data on all countries, except Spain and Belgium[5], and all origin groups. Only individuals out of full-time education are considered in the analysis in order to make the analysis on educational achievement and labour market performance meaningful. As the population of the survey is relatively young, this means that only two thirds of the total sample is considered in the estimations. Table 2.1 presents the number of individuals considered in the sample by city and origin group. The definition of second generation used throughout the paper is children with a least one foreign-born parent. Most second generation individuals in the sample considered have both parents foreign-born, only approximately 6% have one foreign-born and one native parent. This distinction does not make much difference in the outcomes analyzed and is hence not presented in most of the results. The only exception is the marriage market outcomes.

A major drawback of the data set is the fact that the weighting scheme of the different countries is hardly comparable. For the moment, all the estimations are made unweighted. One should hence be careful not to extend the results, and in particular the descriptive characteristics of the sample, to the population of the cities considered. Measures of association, on the

[5]Data from Spain does not have a sample of children of Turkish immigrants and data from Belgium is not available.

other hand, like regression coefficients, should not be too different in weighted and unweighted estimations, in particular if the sampling rate does not depend too much on the outcome variable. Comparing the TIES sample descriptive statistics with other main surveys for the different countries should alleviate this concern and is an important point of the research agenda. For a discussion on weighted against unweighted estimations, see Korn and Graubard (1995).

2.3 Parental Background and the Integration of the Second Generation

The majority of the European second generation in the TIES are the children of the post-war guest workers. For the most part, the guest workers had low levels of education and came from a rural background. Given the lower educational background of their parents, the children of immigrants may not be expected to do on average as well as the children of native-born parents. It is crucial to compare the second generation to native children with similar background in order to understand to which extent the difficulties they face are specific to migrant families. From a policy perspective, progress on this issue should indicate whether suitable policies are policies targeted at low educated, low income families or policies aimed specifically at the children of immigrants.

2.3.1 The Empirical Model

In this section, I use a very simple model to analyze differences in educational and labour market outcomes of children of immigrants and children of natives across 13 major European cities holding family characteristics constant. The originality of the analysis is to compare children of different origins at the city level. The strength of the TIES survey is the very detailed personal and family history that makes these comparisons possible at the local level. Table 2.2 contains the means values for all variables used by ethnic group for the individual and the parents.

I estimate models of the type:

$$Y_i = X_i\beta + Z_i\gamma + \alpha_k SG_{i,k} + \mu_j City_{i,j} + \epsilon_i$$

Y_i represents an outcome measure in education, the labour market or the marriage market of individual . In particular, the measures used for educational achievement are: a dummy for having attended some form of higher education; the labour market outcomes are monthly earnings, a dummy for being employed and labour force participation for women; as a marriage market outcome, I use a dummy to indicate whether the individual married within the ethnic group.

$SG_{i,k}$ are dummy variables corresponding to each second generation group, Turkish, Moroccan and ex-Yugoslav. The reference group is the children of natives.

X_i is a vector of individual characteristics and Z_i of parental characteristics.

All estimations have city dummies to capture regional variation in the outcomes considered. In this model, there are no interactions between the ethnic group dummies and the city fixed effects. The assumption made is that the slopes for each group are the same in the different cities. Although the assumption is strong, the results are robust to introducing the interactions. The results of the unrestricted model are presented in tables 2.10 to 2.14 in the appendix. The background coefficients are also the same for all groups. I do not allow for different returns to characteristics between ethnic groups. This means that for instance, a European father's high-school diploma is assumed to have the same impact on the child's probability of attending higher education than the equivalent foreign diploma for a Turkish father. The differences in returns will hence be captured by the second generation group dummies. Although the quality of education varies between countries, the aim is to keep the specifications as simple as possible, even if this implies some additional restrictions.

The individual characteristics are limited to age, gender and in some cases education. Although the TIES survey has a large choice of individual variables, I want to capture the total effect of parental background and over controlling may hide part of the effect.

The choice of parental background variables is also larger than in most surveys. I use variables that account for parental educational and cultural background, family composition and the parents' labour market status at a prime age. Parents' education is coded in 4 levels corresponding to primary, lower-secondary, higher-secondary education, and higher education (Father edu1 to Father edu4 and Mother edu1 to Mother edu4). A proxy for parents' background is the quantity of books (Books) owned by the household when the individual attended high-school. This variable is coded in 5 categories: 0-10, 11-25, 26-50, 51-100, more than 100 books and has proved to be a good proxy in the education literature. Introducing this variable helps capturing family background characteristics that may not be captured by the parents' education. This may be particularly relevant in the case of immigrant families, as foreign education levels are hardly comparable to the host countries. Family composition is accounted for by the individual's number of siblings (Siblings), which has been shown in the literature to be relevant in particular for educational achievement. Black et al. (2005) show for Norway that not only the size of the family matters but also the birth order. However accounting for the number of older siblings, instead of the total number of siblings, makes no significant difference in the specifications below. The size of the family matters for most outcomes whereas the birth order does not. Parents' labour market situation is captured using: a distinction between high-skilled and low-skilled occupation for the father when the individual was aged 15 (F.high-skill) ; and the mother's labour force participation status also at age 15 (M. labour force). It is important to have these last two variables when the individual was aged 15, since the parents' labour market status at a prime working age matters a priori more than at the time of the survey when most parents are already retired. Also for immigrants, there is often a downgrading in labour market status upon arrival in the host country. After 15 years or more of residence, the labour market situation of the immigrant parents should have significantly improved and stabilized.

2.3.2 Results

Educational Achievement

It is among individuals who attended higher education that the children of immigrants in the TIES sample used seem to lag behind the most significantly. Since higher education degrees are a ticket to the best paying jobs, this disadvantage has long term consequences for the life cycle earnings of the second generation. I use as a measure of educational attainment, an indicator for whether the individual attended university or some equivalent form of higher education. Table 2.3 presents the results of a linear probability model of higher education attendance. All specifications control for age, gender and city fixed effects. Column (1) presents average differences between ethnic groups in higher education attendance for the whole sample considered and column (2) presents these differences only for the individuals who have no missing values for the background variables used in the model. Background characteristics are added in columns (3) to (5). All other tables in this section are presented in the same way.

The three second generation groups in the TIES sample attended higher education less often than natives living in the same cities. The difference is particularly large for the Moroccan and Turkish second generation at 25 percentage points. Controlling for parental education increases the fit of the model, the adjusted r-squared goes from 0.20 to 0.26. More importantly, parents' education reduces the higher education attendance gap by more than half for all groups considered. The effect is particularly large for the Moroccan second generation whose parents in the 4 cities considered have particularly low levels of education, whereas their children do much better as shown on the summary statistics in table 2.2. Both mother and father's education are significant in the model. Running a separate model for men and women shows that the mothers' education is more relevant for the daughter's than the son's higher education attendance[6].

The Books variable, as an extra proxy for parental background, reduces the gap further by 4 percentage points for all groups. Family background differences not accounted for by parents' education thus also play a part in the educational achievement gap of the TIES second generation. The number of siblings has a negative impact on the higher education attendance rate. This result is in line with the education literature. As immigrant families in the sample are on average larger than native families, family composition plays a part in explaining the educational gap.

Finally, among the second generation, differences in the family's integration, as captured by the usage of the host country language, also have a significant impact on educational achievement. Children of natives and children of immigrants with similar family background and who mainly speak the host country language within the household have on average virtually the same rate of higher education attendance. The average difference in higher education attendance between children of immigrant families that always speak the host country language among them and those who never do is 12 percentage points.

[6]The results are not shown but are available on the demand.

The apparently very large differences between ethnic groups in higher education attendance in the sample thus seem to be to a large extent explained by background characteristics. Table 2.10 in the appendix presents the results for the three second generation groups in each city with the same specifications than in table 2.3[7].

Labour Market Outcomes

Several labour market outcomes are analyzed in this section: wages and employment rates but also women's labour force participation, and occupational status, measured by a distinction between high-skilled and low-skilled jobs.

Table 2.4 presents the results of a linear probability model of employment[8]. Employment rates are lower for the second generation than for natives in the sample. Controlling for the individual's education level decreases the gap slightly. The family background variables are not strong predictors of the employment rates once education has been taken into account. Separate estimations for men and women show that the differences in employment rates are higher for women than men, and differences in background are more important for women than men[9]. As before, the different coefficients for the second generation groups by city are shown in table 2.11 in the appendix. Differences in employment rates are not strongly related to background once education has been taken into account.

When looking at labour market outcomes for women, the labour force participation is an important indicator. Table 2.5 presents the results of a linear probability model of labour force participation for women in the TIES sample. Second generation women with Turkish and Moroccan origins have on average much lower participation rates than children of natives at the city level, 19 and 13 percentage points respectively.

Accounting for differences in education reduces the Turkish and Moroccan gap by approximately one third. Second generation women from these two groups have on average lower education and hence face a lower cost of staying out of the labour force. Parental background explains half of the remaining gap for the Turkish second generation and the entire gap for the Moroccan second generation[10]. The individual's education decision is likely to be correlated with the decision to participate in the labour force. For example, a woman who decides not to participate in the labour force may decide not to complete an advanced schooling degree as she will not get the return from the schooling investment. Also, as the individual's education is partly explained by the parents' own education, the education variable may be hiding the full

[7]The coefficients on the control variable are not shown as they are very similar to the ones in table 2.3.

[8]The dependent variable is a dummy that equals one if the individual is employed full-time at the time of the survey and equals 0 if the individual participates in the labour force but is unemployed. Part-time workers are excluded from the analysis.

[9]The results are not shown but are available on demand.

[10]The results for the Moroccan second generation are an average for 4 cities and those for the Yugoslav second generation an average for 6 cities, compared to 13 cities for natives and the Turkish second generation. Introducing interactions between cities and second generation groups allows estimating the coefficients at the city level. The results presented in table 2.12 in the appendix are robust.

effect of parental background. However, not accounting for the individual's education in the specifications does not have an effect on the other coefficients presented in table 2.5[11].

The coefficient on the number of siblings is significantly negative. Women from larger families participate less in the labour force even after accounting for ethnicity. Having a mother who participated in the labour force at her prime age is also a significant predictor of labour force participation of the daughter.

As for the educational outcomes results, the language index is large and significant. Speaking mostly the host country language within the family is associated with a higher probability for women to participate in the labour force.

All in all, taking into account background characteristics strongly decreases differences in labour force between ethnic groups. The results are robust to allowing for different city effects for the different ethnic groups. The results are presented in table 2.12 in the appendix.

Once employed, occupational status is another measure of a labour market outcome. I use the distinction between low-skilled and high-skilled occupations. This classification is based on the occupation classification, ISCO 88, at the one digit level[12]. Results of a linear probability model are presented in table 2.6.

Second generation individuals have less often high-skilled occupations than the children of natives: the difference is close to 15 percentage points for each second generation group. A large part of this gap is linked to the lower educational attainment of the second generation. Accounting for educational attainment strongly decreases the gap. Parental background reduces the gap further: the gap becomes insignificant for all groups. I introduce in these specifications a variable that indicates whether the individual's father had a high-skilled occupation when the respondent was aged 15 years old. The father's occupational status is a very strong predictor of the child's own occupational status. Once this variable and the educational attainment have been accounted for, parent's education has no significant impact on the individual's probability of having a high-skilled occupation.

The TIES survey does not contain detailed information on the individual's wage, instead respondents were asked to situate their monthly earnings in one out of ten 500 euro groups: less than 500, between 500 and 1000, and so on, until 5000 or more. Although in Switzerland and Sweden, the amounts are stated in Swiss franc and Swedish kronor, respectively, the earning bands are approximately equivalent. I use the midpoints of each band as a measure for monthly earnings. All estimations have city fixed effects that account for differences in prices. Very few respondents work part-time[13], less than 5% of the sample, thus only full-time workers are considered for the estimations. Table 2.7 presents estimation results of a linear regression model of log monthly earnings.

[11] The results are not shown but are available on demand.

[12] The occupational classification is for the moment not available for France. This country is hence not taken into account in these specifications.

[13] Individuals working less than 20 hours a week are considered to be part-time workers.

At the city level, in the TIES sample, the Turkish and the Moroccan second generations earn on average significantly less than natives; whereas the children of Yugoslav immigrants earn approximately the same than natives. Accounting for the individual's education decreases the group differences since, as discussed earlier, there is important between group education variation. After accounting also for family background, the earnings gap decreases substantially in the sample considered. The TIES survey has detailed job information. Even after controlling for the size of the firm, the sector, the occupation at the 1 digit level and the private/public status of the workplace, the results remain very similar to the ones in table 2.7[14]. Separate estimations for men and women show that the earnings gap is higher for women than men[15].

Marriage Market Outcomes

The marriage of the children of immigrants to natives may be seen as an alternative measure of social integration. In the economics assimilation literature, an alternative assimilation index used is the rate of inter-marriage of the children of immigrants. Inter-marriage of the second generation has been shown in the US case to be correlated to other economic assimilation measures, see Card et al. (2000), and also to promote the second generation spouse labour market outcomes through extended networks, see Furtado and Theodoropoulos (2009). The index considered in this section is the marriage rate within the immigrant group: the dependent variable is one if the individual marries an immigrant from the parents' country of origin or a second generation from the same origin group; it is zero if the individual marries someone whose both parents are native born[16]. The aim of this section is to understand to which extent marriage decisions are correlated to family background and investigate which parents are more likely to have children who marry outside the immigrant group.

The regression results presented in table 2.8 are obtained from a linear probability model of the inter-marriage index described above. The regressions are run only on second generation individuals, since the outcome measured is not relevant for the children of natives. The baseline is now the Turkish second generation. As in the previous sections, interactions between the second generation groups and the city effects allow to compare for each city the difference in outcomes of the different groups. Differences in marriage rates within the ethnic group between cities may be due to differences in the sizes of the immigrant local communities. The interactions between the origin groups and the city effects account for these potential differences[17].

In the TIES sample, marriage within the immigrant group is the norm for the children of Turkish and Moroccan immigrants but less so for the Yugoslav second generation. Women

[14]The results are not shown but are available on request.

[15]The results are not shown but are available on request.

[16]There are only 20 cases in the sample considered of a second generation individual marrying an immigrant or a second generation from another origin, and 51 cases where a second generation marries someone from a country that is neither the host country, neither one of the three origin countries. These observations are not taken into account in the estimations.

[17]The model with interactions between cities and second generation groups is not shown but results are available from the author. All coefficients remain virtually the same than those in table 2.8.

in the TIES sample are also more often married within the ethnic group. I use all the variables describing family background used in the previous models and add a dummy to indicate whether one of the parents is native-born. The marriage choice of the parents is particularly important in the marriage market[18]. In fact, having one native-born parent strongly decreases the probability for men and women of marrying within the ethnic group. The mother's characteristics come out to be more relevant than the father's. Education and labour market status are important predictors of marriage decisions. Children with more educated parents and with a working mother tend to marry more outside of the immigrant group. Family size is also positively correlated with marriage within the group: women from larger families intra-marry more even after controlling for parental education. Separate results for men and women indicate that background variables are more important predictors for women than for men. All the background variables are also correlated with the individual's educational achievement. However, even accounting for the individual's education level, the parental background variables still have an impact on the marriage choices. Note that in specifications that account for the individual's education, column (4), and specifications that ignore this variable, column (3), the coefficients on the parents' education remain stable. The language index, as in the previous models, has a strong effect also on the inter-marriage rate. Children of immigrant families more assimilated in terms of the host country language tend to inter-marry significantly more. Note that this effect remains strong after controlling for having one native-born parent.

2.4 Intergenerational Assimilation

The descriptive results above show that, for quite diverse outcomes, parental background explains most of the differences between second generation groups and children of natives. The question that arises in the intergenerational assimilation is how mobile are the children of immigrants compared to the natives'. Is the weight of parental background heavier or lighter in immigrant families? In the next section, I compare the intergenerational transmission of human capital for immigrant and native households for a set of outcomes.

A simple model of intergenerational transmission of human capital, based on Solon (1999), may be written as follows:

$$Ychild_i = \alpha + \beta parent_i + \epsilon_i$$

$Ychild_i$ is an outcome of interest (related in this context to the educational or labour market performance) of the individual surveyed in the TIES sample. $Yparent_i$ measures the same outcome for the mother and/or the father when the child was aged 15. Ideally, the child and the parents' outcomes should be measured exactly at the same stage in the life cycle, say at age 40. β is the coefficient of interest, it measures intergenerational mobility, or how correlated parent and child outcomes are. A higher β corresponds to a lower intergenerational mobility. The constant α captures differences in mean outcomes across generations. For the estimations,

[18] Having a native-born parent is not significantly correlated with the educational and labour market outcomes presented above in the TIES sample.

I use a slightly transformed model:

$$Ychild_i = \alpha + \beta Yparent_i + \gamma_k Yparent_i * SG_{i,k} + \delta_k SG_{i,k} + \mu_j City_{i,j} + \mu_{j,k} City_{i,j} * SG_{k,j} + \epsilon_i$$

The city fixed effects interacted with the different ethnic groups capture regional variations in outcomes. I do not interact city dummies with the parents' outcomes to allow for different rates of intergenerational transmission at the city level. Introducing dummies for the different origins allows the constants in the model to differ between children of immigrants of the different origins and children of natives. The coefficients of interest in this setting are the γ_k: the difference in intergenerational transmission between ethnic groups. β represents the intergenerational transmission for natives.

Table 2.9 presents the estimations of a simple intergenerational model for different outcomes: higher education attendance; two occupational measures: the International Socio-Economic Index of occupational status (ISEI)[19], and the dichotomous variable for high-skilled versus low-skilled occupation; and the labour force participation for women. In the intergenerational assimilation literature, earnings is a more commonly used measure, however the TIES survey does not inquire about parental earnings. The ISEI and F. high skill measures are built using the father's occupation when the respondent was aged 15 and are to a certain extent correlated to earnings.

Let us start by the three first columns in table 2.9. In all three cases, there is a high correlation between the child and the father's outcomes, between 25 and 32 percentage points. The correlation between immigrant fathers and children's outcomes is lower since the coefficient of interest is negative or insignificant. Some of these coefficients are not significant due to high standard error[20]. More research is still needed on the link between parents and children's outcomes but this simple model indicates that on average children and parents' characteristics are more strongly correlated for native families than for immigrant ones, indicating the possibility of a greater intergenerational mobility for immigrant children.

The labour force participation of women is a different case. Having a mother who participated in the labour force increases the daughter's probability of participating in the labour force by 4 percentage points. The correlation in outcomes between generations seems stronger for second generation women, although the large standard errors make the coefficients insignificant. This result would be in line with recent research on the second generation in the US. In fact, second generation female labour market outcomes, and labour force participation in particular, have been shown to be strongly influenced by the parents' country of origin, see Blau et al. (2008), and to be correlated with the labour force participation of women in the parents' country of origin, see Fernandez and Fogli (2009).

[19] This index captures the attributes of occupations that convert education into income. The index was derived by the optimal scaling of occupation groups to maximize the indirect effect of education on income through occupation and to minimize the direct effect of education on income, net of occupation (both effects being net of age). For more information, please refer to Ganzeboom et al. (1992).

[20] Estimations in columns (2) and (3) do not take France into account since the ISCO 88 coding is for the moment unavailable for this country. Column (3) also does not take Austrian data into account.

2.5 Conclusion

In all educational and labour market outcomes analyzed in the chapter, the children of Turkish, Moroccan and Yugoslav immigrants lag behind the children of natives in 13 major European cities.

The large educational attainment gap between children of immigrants and children of natives at the city level is on average largely explained by differences in family background. The parents' education levels are strong predictors of the child's own education and both mothers' and fathers' educations seem to matter. The number of books in the household when the individual attended high school has often been used in the literature as a proxy for parental background. This variable allows accounting for differences in the child's upbringing environment that may not be reflected solely by the parent's education levels. Introducing the number of books in the analysis also explains part of the educational achievement gap. Comparing children of immigrants and children of natives whose parents have similar educational levels and who grew up in a somewhat similar environment reduces the differences in educational achievement greatly. Children from larger families are also on average less likely to complete higher education. Among children of immigrants, children who more often speak the host country language within the family (with the parents and siblings) are more likely to attend higher education. The analysis has been focused on the higher education rate of attendance as a measure of educational achievement. Further research should investigate whether these findings also hold for alternative educational achievement measures.

The labour force participation of second generation women is significantly lower than the participation of children of natives in the sample. Part of the gap in labour force participation between the two groups can be explained by differences in educational levels. Children of immigrants are less educated and thus opt out of the labour force more often. After accounting for the individual's own education level, the parents' education is not a strong predictor of labour force participation. The mother's labour force participation, on the other hand, seems to be a strong determinant of the daughter's own decision to participate in the labour force. Among the children of immigrants, the extent to which the host country language is spoken within the family is again positively correlated with labour force participation.

Differences in employment rates for children of natives and children of immigrants may partly be explained by differences in education levels. Less educated individuals have on average lower employment rates. After accounting for differences in educational attainment of the children, parental background seems to have virtually no effect on the employment rate. Family composition and the fluency in the host country language also appear to be uncorrelated with the likelihood of being in employment. This suggests that the impact of parental background on children of migrants' labour market outcomes is mainly indirect (i.e. through its impact on educational attainment).

Once employed, children of immigrants earn lower wages and are less likely to have high-skilled occupations than children of natives living in the same cities. Lower education levels, once again, partly explain the differences observed. Parent's education and the host country language fluency are not strongly correlated with the individual's wage or with his/her occupational status. A notable exception is the strong correlation between the individual's probability of having a high-skilled job and whether the father had himself a high-skilled job at his prime age. Nevertheless, after controlling for educational differences, the gap between children of natives and second generation in these two labour market outcomes is (already) fairly small in the sample.

The children of immigrants' marriage decisions, and in particular the decision whether to marry within the ethnic group, are correlated with parental background. Children who have more educated parents and who were brought up in smaller families tend to marry more outside of the ethnic group. The results also indicate that the parents' marriage decisions are strong predictors of the child's own marriage patterns. Children who have one native-born parent seem to have a much higher probability of marrying a native. Speaking the host country language within the family is also correlated with marrying outside of the group. Other factors are relevant for marriage decisions, such as religion and other cultural differences, and would be worth investigating in further research.

One major concern is the extent to which the results presented above may be generalized to the whole population of the cities considered. The introduction of appropriate weights in the analysis and the comparison of the TIES sample with other available, albeit more general, data sources on the European second generation should ease this concern.

The aim of the chapter was to present a broad picture of the second generation outcomes at the European level. Results at the city level[21] indicate that although the outcome gaps are of different magnitudes in the different cities, the results go broadly in the same direction whatever the European city considered. Future research should aim at understanding the specificities of each European country and the reasons for cross country differences in the second generation integration process.

The results suggest that there is less correlation between parents' and children's outcomes in immigrant families compared with native ones. This result indicates the possibility of a greater intergenerational mobility for immigrant children. However, intergenerational mobility is linked to the institutional setting of a country. There are large differences between OECD countries in the extent to which parents' and children's outcomes are correlated[22]. Or putting it differently, children from the lowest social classes have different probabilities of climbing the social ladder depending on the country they live in. A better understanding of cross country differences in intergenerational mobility more generally and how these relate to the schooling system and labour market institutions would be a crucial next step in the research agenda on this important issue. The second step would then be to analyze cross-country differences between the intergenerational mobility of immigrants and native families and how these relate to countries' institutional settings.

[21]The results for the different outcomes are presented in the appendix.

[22]See Corak (2006) for a comparative overview of intergenerational mobility in Europe.

Table 2.1: Effectif Numbers by City

	Natives	Turkish SG	Moroccan SG	Ex-Yugoslav SG
Amsterdam	174	149	130	0
Basel	153	141	0	122
Berlin	213	221	0	165
Frankfurt	222	210	0	170
Linz	133	130	0	160
Paris	111	110	0	0
Rotterdam	173	177	132	0
Stockolm	186	185	0	0
Strasbourg	108	165	0	0
Vienna	155	173	0	217
Zurich	133	131	0	165

Source: The TIES survey was carried out by survey bureaus under supervision of the nine national TIES partner institutes: Netherlands Interdisciplinary Demographic Institute (NIDI), Institute for Migration and Ethnic Studies (IMES), University of Amsterdam in the Netherlands, the Institute for Social and Political Opinion Research (ISPO), University of Leuven in Belgium; the National Institute for Demographic Studies (INED) in France; the Swiss Forum for Migration and Population Studies (SFM), University of Neuchatel in Switzerland; the Centre for Research in International Migration and Ethnic Relations (CEIFO), University of Stockholm in Sweden; the Institute for Migration Research and Intercultural Studies (IMIS), University of Osnabruck in Germany, the Institute for the Study of Migration (IEM), Pontifical Comillas University of Madrid in Spain, and the Institute for European Integration Research (EIF), Austrian Academy of Sciences in Austria. For more information on the TIES project, see http://www.tiesproject.eu/index.php/lang=en.

.

Table 2.2: Summary Statistics

	Natives	Turkish SG	Moroccan SG	Ex-Yugoslav SG
Male (d)	0.49	0.5	0.51	0.5
Age	28.45	25.91	25.23	26.73
Ed1 (d)	0.01	0.06	0.14	0.03
Ed2 (d)	0.08	0.23	0.22	0.12
Ed3 (d)	0.5	0.58	0.42	0.73
Ed4 (d)	0.4	0.14	0.22	0.11
Siblings	1.46	2.66	4.97	1.61
Books 0-10 (d)	0.06	0.27	0.36	0.18
Books 11-25 (d)	0.12	0.29	0.3	0.28
Books 26-50 (d)	0.18	0.22	0.17	0.24
Books 51-100 (d)	0.21	0.12	0.12	0.15
Books 100+ (d)	0.42	0.11	0.05	0.15
Language Index	0	1.41	1.41	1.13
1 parent native-born (d)	0	0.05	0.06	0.1
High-skill (d)	0.6	0.43	0.4	0.32
ISEI	49.55	41.97	44.88	42.34
LFP (d)	0.87	0.7	0.72	0.82
Empl (d)	0.92	0.84	0.81	0.89
AT	0.16	0.17	0	0.37
CH	0.16	0.15	0	0.29
DE	0.25	0.24	0	0.34
FR	0.12	0.15	0	0
NL	0.2	0.18	1	0
SE	0.11	0.1	0	0
Mother ed1 (d)	0.03	0.55	0.74	0.18
Mother ed2 (d)	0.32	0.22	0.14	0.33
Mother ed3 (d)	0.47	0.16	0.11	0.39
Mother ed4 (d)	0.19	0.07	0.01	0.09
M. labour force (d)	0.63	0.43	0.26	0.69
Father ed1 (d)	0.03	0.48	0.67	0.13
Father ed2 (d)	0.19	0.22	0.12	0.26
Father ed3 (d)	0.47	0.16	0.15	0.46
Father ed4 (d)	0.31	0.13	0.05	0.15
F.high-skill (d)	0.65	0.47	0.2	0.43
FISEI	48.27	34.21	32.31	34.77
N	1761	1792	262	989

Notes: Ed1 to Ed4 are 4 dummies representing 4 levels of education: primary, lower secondary, higher secondary and college. Mother ed1 to Mother ed4 and of the individual. Books represents in 5 categories the number of books ownedFather ed1 to Father ed4 represent the same levels for the mother and the father by the household when the individual was in high-school. The Language Index represents the frequency the second generation speaks the host country language within the family at the time of the survey, the index goes from 0 to 3. 1 parent native-born is a dummy that is 1 if the individual is a second generation and has one parent who is native-born. High skill is a dummy that is 1 if the individual has a high-skilled occupation. ISEI is the International Socio-Economic Index of occupational status of the current occupation based on the ISCO 88 coding. LFP is a dummy for labor force participation de ned only for women. Empl is a dummy for employment. M. labor force is a dummy that equals 1 if the mother participated in the labor force when the individual was 15 years old. F. High-skill is a dummy that equals 1 if the occupation of the father when the individual was aged 15 is considered high-skilled according to the ISCO 88 classi cation. FISEI is the the International Socio-Economic Index of occupational status (ISEI) of the father when the individual was aged 15 based on the ISCO 88 coding.
Source: TIES

Table 2.3: Higher Education

	(1)	(2)	(3)	(4)	(5)
Turkish SG	-0.22***	-0.23***	-0.12***	-0.07***	-0.02
	(-0.01)	(-0.02)	(-0.02)	(-0.02)	(-0.02)
Moroccan SG	-0.25***	-0.21***	-0.05	0	0.04
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Yugoslav SG	-0.16***	-0.15***	-0.08***	-0.04*	0.01
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Men	-0.01	-0.01	-0.01	0.00	0.00
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Father edu2			-0.00	-0.02	-0.02
			(0.02)	(0.02)	(0.02)
Father edu3			0.06***	0.04*	0.03
			(0.02)	(0.02)	(0.02)
Father edu4			0.23***	0.18***	0.17***
			(0.03)	(0.03)	(0.03)
Mother edu2			0.01	-0.01	-0.02
			(0.02)	(0.02)	(0.02)
Mother edu3			0.04	0.01	-0.01
			(0.02)	(0.02)	(0.02)
Mother edu4			0.16***	0.12***	0.11***
			(0.03)	(0.03)	(0.03)
Siblings			-0.02***	-0.01***	-0.01***
			(0.00)	(0.00)	(0.00)
Language Index					-0.05***
					(0.01)
Books				Yes	Yes
Constant	-0.20***	-0.24***	-0.36***	-0.44***	-0.43***
	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)
Observations	4779	3445	3445	3445	3445
Adjusted R2	0.192	0.197	0.260	0.290	0.294

Notes: Linear probability model of higher education attendance. All specifications control for age, gender and city fixed effects. See the note on Table 2 for the definitions of the variables used. Standard errors in parentheses. *** p¡0.01, **¡0.05, *p¡0.1.
Source: TIES

Table 2.4: Employment

	(1)	(2)	(3)	(4)	(5)
Turkish SG	-0.02	-0.08*	-0.06	-0.08*	-0.05
	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Moroccan SG	-0.10**	-0.11**	-0.10**	-0.12**	-0.08*
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Yugoslav SG	0.06*	0.11***	0.11***	0.12***	0.13***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Men	0.01	0.01	0.02	0.02*	0.02*
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Edu2			0.04	0.05	0.05
			(0.04)	(0.04)	(0.04)
Edu3			0.17***	0.18***	0.17***
			(0.04)	(0.04)	(0.04)
Edu4			0.21***	0.21***	0.21***
			(0.04)	(0.04)	(0.04)
Father edu2				-0.07***	-0.08***
				(0.02)	(0.02)
Father edu3				-0.04**	-0.05**
				(0.02)	(0.02)
Father edu4				-0.05**	-0.05**
				(0.02)	(0.02)
Mother edu2				0.02	0.01
				(0.02)	(0.02)
Mother edu3				0.03*	0.03
				(0.02)	(0.02)
Mother edu4				0.03	0.02
				(0.02)	(0.03)
Siblings				0.00	0.00
				(0.00)	(0.00)
Language Index					-0.03***
Books				Yes	Yes
Constant	0.52***	0.54***	0.43***	0.44***	0.45***
	(0.04)	(0.04)	(0.06)	(0.06)	(0.06)
Observations	4156	3022	3022	3022	3022
Adjusted R2	0.056	0.065	0.090	0.093	0.095

Notes: Linear probability model of employment. All specifications include age, gender and city fixed effects. See the note on Table 2 for the definitions of the variables used. Standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1.
Source: TIES

Table 2.5: Labour Force Participation for Women

	(1)	(2)	(3)	(4)	(5)
Turkish SG	-0.18***	-0.18***	-0.12***	-0.07**	-0.04
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Moroccan SG	-0.15***	-0.17***	-0.09*	0.01	0.03
	(0.04)	(0.05)	(0.05)	(0.06)	(0.06)
Yugoslav SG	-0.03	-0.04	-0.02	-0.00	0.03
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Edu2			0.05	0.04	0.04
			(0.06)	(0.06)	(0.06)
Edu3			0.31***	0.29***	0.28***
			(0.06)	(0.06)	(0.06)
Edu4			0.40***	0.37***	0.36***
			(0.06)	(0.06)	(0.06)
Father edu2				-0.02	-0.02
				(0.03)	(0.03)
Father edu3				-0.02	-0.02
				(0.03)	(0.03)
Father edu4				-0.01	-0.02
				(0.04)	(0.04)
Mother edu2				0.05	0.04
				(0.03)	(0.03)
Mother edu3				0.05	0.04
				(0.03)	(0.03)
Mother edu4				0.02	0.01
				(0.04)	(0.04)
Siblings				-0.02***	-0.02***
				(0.01)	(0.01)
M. labour force				0.04*	0.04*
				(0.02)	(0.02)
Language Index					-0.03*
					(0.02)
Books				Yes	Yes
Constant	0.96***	0.99***	0.79***	0.77***	0.78***
	(0.06)	(0.07)	(0.09)	(0.10)	(0.10)
Observations	2417	1689	1689	1689	1689
Adjusted R2	0.065	0.050	0.119	0.125	0.126

Notes: Linear probability model of labor force participation. All specifications control for age and city fixed effects. See the note on Table 2 for the definitions of the variables used. Standard errors in parentheses. *** p¡0.01, ** p¡0.05, *p¡0.1.
Source: TIES

Table 2.6: High-skilled Jobs

	(1)	(2)	(3)	(4)	(5)
Turkish SG	-0.12***	-0.09***	-0.01	0.03	0.06*
	(0.02)	(0.02)	(0.02)	(0.03)	(0.03)
Moroccan SG	-0.17***	-0.18***	-0.15**	-0.10	-0.07
	(0.04)	(0.06)	(0.06)	(0.07)	(0.07)
Yugoslav SG	-0.13***	-0.12***	-0.07***	-0.02	0.01
	(0.02)	(0.03)	(0.02)	(0.03)	(0.03)
Men	-0.08***	-0.08***	-0.07***	-0.07***	-0.06***
	(0.01)	(0.02)	(0.02)	(0.02)	(0.02)
Edu2			-0.11	-0.10	-0.10
			(0.08)	(0.08)	(0.08)
Edu3			0.02	0.02	0.01
			(0.08)	(0.08)	(0.08)
Edu4			0.34***	0.30***	0.30***
			(0.08)	(0.08)	(0.08)
Father edu2				-0.04	-0.04
				(0.03)	(0.03)
Father edu3				0.01	0.01
				(0.03)	(0.03)
Father edu4				0.01	0.01
				(0.04)	(0.04)
Mother edu2				0.00	-0.01
				(0.03)	(0.03)
Mother edu3				-0.02	-0.03
				(0.03)	(0.03)
Mother edu4				0.01	-0.01
				(0.04)	(0.04)
F. high skill				0.11***	0.11***
				(0.03)	(0.03)
Siblings				0.00	0.00
				(0.01)	(0.01)
Language Index					-0.04**
					(0.02)
Books				Yes	Yes
Constant	-0.25***	-0.19***	-0.09	-0.12	-0.12
	(0.05)	(0.07)	(0.10)	(0.11)	(0.11)
Observations	3501	2178	2178	2178	2178
Adjusted R2	0.243	0.248	0.330	0.339	0.340

Notes: Linear probability model of high-skilled job status. All specifications control for age, gender and city fixed-effects. See the note on Table 2 for the definitions of the variables used. Standard errors in parentheses. *** p¡0.01, **p¡0.05, * p¡0.1.
Source: TIES

Table 2.7: Earnings

	(1)	(2)	(3)	(4)	(5)
Turkish SG	-0.11***	-0.11***	-0.05***	0.00	0.02
	(0.01)	(0.02)	(0.02)	(0.02)	(0.03)
Moroccan SG	-0.15***	-0.12***	-0.08**	0.01	0.02
	(0.03)	(0.04)	(0.04)	(0.04)	(0.05)
Yugoslav SG	-0.03*	-0.04*	-0.00	0.03	0.04*
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)
Men	0.14***	0.15***	0.16***	0.16***	0.16***
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
Edu2			0.05	0.05	0.05
			(0.06)	(0.06)	(0.06)
Edu3			0.15***	0.13**	0.13**
			(0.06)	(0.06)	(0.06)
Edu4			0.35***	0.31***	0.31***
			(0.06)	(0.06)	(0.06)
Father edu2				-0.00	-0.00
				(0.03)	(0.03)
Father edu3				0.01	0.00
				(0.03)	(0.03)
Father edu4				-0.01	-0.01
				(0.03)	(0.03)
Mother edu2				0.05*	0.04*
				(0.02)	(0.02)
Mother edu3				0.03	0.03
				(0.02)	(0.03)
Mother edu4				0.06*	0.05*
				(0.03)	(0.03)
Siblings				-0.01*	-0.01*
				(0.01)	(0.01)
Language Index					-0.02
					(0.01)
Books				Yes	Yes
Constant	4.75***	4.72***	5.00***	4.86***	4.86***
	(0.20)	(0.25)	(0.25)	(0.25)	(0.25)
Observations	3234	2415	2415	2415	2415
Adjusted R2	0.448	0.444	0.485	0.489	0.489

Notes: OLS model of the log monthly earnings, as described in the text. All specifications control for age, age2, gender and city fixed effects. See the note on Table 2 for the definitions of the variables used. Standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1
Source: TIES

Table 2.8: Endogamous Marriage

	(1)	(2)	(3)	(4)	(5)
Moroccan SG	-0.01	0.01	-0.01	0.01	0.07
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Yugoslav SG	-0.24***	-0.18***	-0.11***	-0.12***	-0.10***
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)
Men	-0.03*	-0.07***	-0.07***	-0.07***	-0.07***
	(0.02)	(0.03)	(0.02)	(0.02)	(0.02)
Edu2				0.03	0.06
				(0.07)	(0.07)
Edu3				0.01	0.06
				(0.06)	(0.06)
Edu4				-0.12*	-0.05
				(0.07)	(0.07)
1 parent native-born			-0.33***	-0.31***	-0.21***
			(0.06)	(0.06)	(0.06)
Father edu2			-0.04	-0.04	-0.01
			(0.04)	(0.03)	(0.03)
Father edu3			-0.08**	-0.08**	-0.05
			(0.04)	(0.04)	(0.04)
Father edu4			-0.09*	-0.08*	-0.04
			(0.05)	(0.05)	(0.05)
Mother edu2			-0.09**	-0.09**	-0.08**
			(0.04)	(0.04)	(0.04)
Mother edu3			-0.00	-0.00	0.00
			(0.04)	(0.04)	(0.04)
Mother edu4			-0.12*	-0.12*	-0.10*
			(0.06)	(0.06)	(0.06)
Siblings			0.01*	0.01	0.01
			(0.01)	(0.01)	(0.01)
M. labour force			0.02	0.03	0.02
			(0.03)	(0.03)	(0.03)
Language Index					0.13***
					(0.02)
Books				Yes	Yes
Constant	0.79***	0.83***	0.90***	0.84***	0.73***
	(0.08)	(0.10)	(0.11)	(0.12)	(0.12)
Observations	1361	887	887	887	887
Adjusted R2	0.14	0.131	0.222	0.232	0.273

Notes: Linear probability model of the marriage assimilation rate described in the text. The reference group is the children of Turkish immigrants. All specifications control for age, gender and city fixed effects. See the note on Table 2 for the definitions of the variables used. Standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1.
Source: TIES

Table 2.9: Inter-generational Mobility

	Higher Education	High-skill	ISEI	LFP
Yparent	0.30***	0.31***	0.33***	0.04
	(0.02)	(0.03)	(0.03)	(0.03)
Yparent*Turkish SG	-0.17***	-0.33***	-0.29***	0.07**
	(0.03)	(0.06)	(0.06)	(0.04)
Yparent*Moroccan SG	-0.46***	-0.22*	-0.12	0.03
	(0.12)	(0.13)	(0.14)	(0.08)
Yparent*Ex-Yugoslav SG	-0.10**	-0.10	-0.05	0.05
	(0.04)	(0.07)	(0.06)	(0.05)
Constant	0.53***	0.55***	40.09***	0.86***
	(0.03)	(0.06)	(2.05)	(0.04)
Observations	4351	2852	1837	2330
Adjusted R2	0.241	0.244	0.189	0.075

Notes: The models are estimated with an OLS. All specifications control for gender, second generation group, city fixed effects and city fixed effects interacted with each second generation group. Standard errors in parentheses. *** p¡0.01, ** p¡0.05, * p¡0.1.
Source: TIES

2.A Appendix

2.A.1 Tables Introducing City Fixed Effects

Table 2.10: Higher Education

	(1)	(2)	(3)	(4)	(5)
Turkish SG*Linz	-0.07	-0.00	0.04	0.05	0.13**
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Turkish SG*Amsterdam	-0.35***	-0.34***	-0.19***	-0.14***	-0.06
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Basel	-0.19***	-0.16***	-0.05	-0.01	0.04
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Berlin	-0.16***	-0.17***	-0.14***	-0.11**	-0.06
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Frankfurt	-0.10***	-0.08	-0.08	-0.06	-0.02
	(0.04)	(0.06)	(0.05)	(0.05)	(0.05)
Turkish SG*Paris	-0.20***	-0.22***	-0.08	-0.03	0.01
	(0.05)	(0.06)	(0.06)	(0.05)	(0.06)
Turkish SG*Rotterdam	-0.30***	-0.31***	-0.21***	-0.15***	-0.07
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Stockholm	-0.30***	-0.29***	-0.12**	-0.07	-0.02
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Strasbourg	-0.43***	-0.51***	-0.33***	-0.26***	-0.21***
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Vienna	-0.12***	-0.12**	-0.02	0.03	0.09*
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Zurich	-0.24***	-0.25***	-0.12**	-0.09*	-0.04
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Moroccan SG*Amsterdam	-0.37***	-0.35***	-0.18***	-0.13**	-0.07
	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Moroccan SG*Rotterdam	-0.23***	-0.18***	-0.03	0.02	0.08
	(0.04)	(0.05)	(0.06)	(0.05)	(0.06)
Yugoslav SG*Linz	-0.12***	-0.09*	-0.05	0.04	0.10*
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Basel	-0.16***	-0.12**	-0.07	-0.00	0.05
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Berlin	-0.15***	-0.14***	-0.10**	-0.08*	-0.05
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Frankfurt	-0.04	-0.05	-0.03	-0.02	0.02
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Vienna	-0.03	0.01	0.07	0.09*	0.14***
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Zurich	-0.29***	-0.28***	-0.17***	-0.13***	-0.07
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Observations	4779	3445	3445	3445	3445

Notes: Linear probability model of higher education attendance. All specifications control for age, gender, city fixed effects and city fixed effects interacted with each second generation group. The coefficients on the main variables are virtually the same than those in table 3. See the note on Table 2.2 for the definitions of the variables used. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.11: Employment

	(1)	(2)	(3)	(4)	(5)
Turkish SG*Linz	-0.08**	-0.05	-0.05	-0.04	0.01
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Amsterdam	-0.14***	-0.18***	-0.14***	-0.15***	-0.10**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Turkish SG*Basel	-0.01	-0.03	-0.00	0.00	0.03
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Turkish SG*Berlin	-0.02	-0.08*	-0.06	-0.08*	-0.05
	(0.03)	(0.04)	(0.04)	(0.05)	(0.05)
Turkish SG*Frankfurt	-0.02	0.02	0.03	0.01	0.03
	(0.03)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Paris	-0.06	-0.05	-0.02	-0.02	0.01
	(0.04)	(0.05)	(0.04)	(0.05)	(0.05)
Turkish SG*Rotterdam	-0.10***	-0.15***	-0.11***	-0.12***	-0.06
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Turkish SG*Stockholm	-0.03	-0.05	-0.03	-0.04	-0.01
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Turkish SG*Strasbourg	0.00	-0.02	0.02	0.02	0.05
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Turkish SG*Vienna	-0.08**	-0.06	-0.04	-0.04	-0.00
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Turkish SG*Zurich	-0.08**	-0.08*	-0.06	-0.05	-0.02
	(0.04)	(0.04)	(0.04)	(0.04)	(0.05)
Moroccan SG*Amsterdam	-0.14***	-0.18***	-0.13***	-0.15***	-0.12**
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Moroccan SG*Rotterdam	-0.10**	-0.11**	-0.10**	-0.12**	-0.08*
	(0.04)	(0.04)	(0.04)	(0.05)	(0.05)
Yugoslav SG*Linz	-0.12***	-0.11***	-0.12***	-0.12***	-0.08*
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Yugoslav SG*Basel	0.03	0.03	0.04	0.06	0.09**
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Yugoslav SG*Berlin	0.06*	0.11***	0.11***	0.12***	0.13***
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Yugoslav SG*Frankfurt	-0.04	-0.07*	-0.05	-0.05	-0.02
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)
Yugoslav SG*Vienna	-0.01	-0.03	-0.02	-0.03	0.01
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Yugoslav SG*Zurich	0.02	0.00	0.01	0.03	0.07
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)
Observations	4156	3022	3022	3022	3022

Notes: Linear probability model of employment. All specifications control for age, gender, city fixed effects and city fixed effects interacted with each second generation group. The coefficients on the main variables are virtually the same than those in table 4. See the note on Table 2.2 for the definitions of the variables used. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.12: Labor Force Participation for Women

	(1)	(2)	(3)	(4)	(5)
Turkish SG*Linz	-0.15**	-0.20**	-0.19**	-0.16**	-0.12
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
Turkish SG*Amsterdam	-0.29***	-0.28***	-0.22***	-0.15**	-0.11
	(0.06)	(0.07)	(0.07)	(0.07)	(0.08)
Turkish SG*Basel	-0.12*	-0.12	-0.08	-0.02	0.00
	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)
Turkish SG*Berlin	-0.26***	-0.18**	-0.13*	-0.08	-0.06
	(0.06)	(0.08)	(0.07)	(0.08)	(0.08)
Turkish SG*Frankfurt	-0.20***	-0.18**	-0.15**	-0.10	-0.08
	(0.05)	(0.07)	(0.07)	(0.07)	(0.07)
Turkish SG*Paris	-0.18**	-0.18**	-0.11	-0.05	-0.03
	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)
Turkish SG*Rotterdam	-0.28***	-0.24***	-0.17***	-0.10	-0.06
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
Turkish SG*Stockholm	-0.08	-0.10	-0.07	-0.02	0.01
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Turkish SG*Strasbourg	-0.16**	-0.17**	-0.06	0.03	0.06
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Turkish SG*Vienna	-0.22***	-0.27***	-0.23***	-0.15**	-0.12*
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Turkish SG*Zurich	0.05	0.02	0.06	0.12	0.14*
	(0.07)	(0.08)	(0.08)	(0.08)	(0.08)
Moroccan SG*Amsterdam	-0.17***	-0.24***	-0.14*	-0.02	0.01
	(0.07)	(0.08)	(0.07)	(0.08)	(0.08)
Moroccan SG*Rotterdam	-0.23***	-0.19***	-0.12*	-0.01	0.01
	(0.06)	(0.07)	(0.07)	(0.08)	(0.08)
Yugoslav SG*Linz	-0.12*	-0.19***	-0.19***	-0.18**	-0.15**
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Yugoslav SG*Basel	-0.04	-0.04	-0.03	-0.02	-0.00
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Yugoslav SG*Berlin	-0.16***	-0.18***	-0.15**	-0.12*	-0.11*
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Yugoslav SG*Frankfurt	0.05	0.04	0.05	0.05	0.07
	(0.06)	(0.06)	(0.06)	(0.06)	(0.06)
Yugoslav SG*Vienna	0.10*	0.14**	0.13*	0.17**	0.20***
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Yugoslav SG*Zurich	0.03	-0.01	0.01	0.04	0.07
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Observations	2417	1689	1689	1689	1689

Notes: Linear probability model of labor force participation. All specifications control for age, city fixed effects and city fixed effects interacted with each second generation group. The coefficients on the main variables are virtually the same than those in table 5. See the note on Table 2.2 for the definitions of the variables used. Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$.

Table 2.13: High-skilled jobs

	(1)	(2)	(3)	(4)	(5)
Turkish SG*Linz	-0.04	-0.04	-0.03	-0.02	0.04
	(0.06)	(0.08)	(0.07)	(0.07)	(0.08)
Turkish SG*Amsterdam	-0.17***	-0.24***	-0.10	-0.05	-0.00
	(0.06)	(0.09)	(0.08)	(0.09)	(0.09)
Turkish SG*Basel	-0.10*	-0.13**	-0.06	0.02	0.05
	(0.06)	(0.06)	(0.06)	(0.06)	(0.07)
Turkish SG*Berlin	-0.25***	-0.27***	-0.18**	-0.15*	-0.12
	(0.05)	(0.08)	(0.08)	(0.08)	(0.08)
Turkish SG*Frankfurt	-0.16***	-0.13	-0.09	-0.08	-0.06
	(0.05)	(0.08)	(0.08)	(0.08)	(0.08)
Turkish SG*Paris	0.09	0.09	0.15**	0.19***	0.22***
	(0.07)	(0.07)	(0.07)	(0.07)	(0.07)
Turkish SG*Rotterdam	-0.25***	-0.12	-0.04	-0.00	0.06
	(0.06)	(0.08)	(0.08)	(0.08)	(0.09)
Turkish SG*Stockholm	-0.26***	-0.26***	-0.17**	-0.11	-0.07
	(0.06)	(0.09)	(0.08)	(0.08)	(0.09)
Turkish SG*Strasbourg	0.08	0.06	0.24***	0.27***	0.30***
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
Turkish SG*Vienna	-0.09	-0.10	-0.05	-0.02	0.02
	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)
Turkish SG*Zurich	-0.08	-0.04	0.05	0.11*	0.14**
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
Moroccan SG*Amsterdam	-0.22***	-0.27***	-0.22**	-0.16	-0.13
	(0.06)	(0.10)	(0.10)	(0.10)	(0.10)
Moroccan SG*Rotterdam	-0.19***	-0.15*	-0.12	-0.07	-0.03
	(0.06)	(0.09)	(0.08)	(0.09)	(0.09)
Yugoslav SG*Linz	-0.19***	-0.23***	-0.21***	-0.18***	-0.14**
	(0.06)	(0.07)	(0.06)	(0.07)	(0.07)
Yugoslav SG*Basel	-0.13**	-0.10	-0.05	0.03	0.06
	(0.06)	(0.07)	(0.06)	(0.06)	(0.07)
Yugoslav SG*Berlin	-0.19***	-0.22***	-0.14**	-0.10	-0.08
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Yugoslav SG*Frankfurt	-0.05	-0.05	-0.03	0.02	0.04
	(0.05)	(0.07)	(0.06)	(0.06)	(0.06)
Yugoslav SG*Vienna	-0.06	0.02	0.02	0.02	0.06
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Yugoslav SG*Zurich	-0.18***	-0.18***	-0.07	-0.00	0.04
	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
Observations	3501	2178	2178	2178	2178

Notes: Linear probability model of high-skilled job status. All specifications control for age, gender, city fixed effects and city fixed effects interacted with each second generation group. The coefficients on the main variables are virtually the same than those in table 6. See the note on Table 2.2 for the definitions of the variables used. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Table 2.14: Earnings

	(1)	(2)	(3)	(4)	(5)
Turkish SG*Linz	-0.14***	-0.16**	-0.16***	-0.13**	-0.10
	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
Turkish SG*Amsterdam	-0.19***	-0.19***	-0.10*	-0.03	-0.01
	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
Turkish SG*Basel	-0.11**	-0.07	-0.02	0.05	0.06
	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
Turkish SG*Berlin	-0.12***	-0.13*	-0.07	-0.03	-0.01
	(0.04)	(0.07)	(0.07)	(0.07)	(0.07)
Turkish SG*Frankfurt	-0.10**	-0.11*	-0.09*	-0.06	-0.05
	(0.04)	(0.06)	(0.06)	(0.06)	(0.06)
Turkish SG*Paris	-0.01	-0.02	0.04	0.10*	0.11**
	(0.05)	(0.06)	(0.05)	(0.05)	(0.06)
Turkish SG*Rotterdam	-0.15***	-0.16***	-0.08	-0.01	0.01
	(0.04)	(0.05)	(0.05)	(0.05)	(0.06)
Turkish SG*Stockholm	-0.04	-0.04	0.02	0.09*	0.10*
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Turkish SG*Strasbourg	-0.12**	-0.16***	-0.04	0.04	0.06
	(0.05)	(0.05)	(0.05)	(0.06)	(0.06)
Turkish SG*Vienna	-0.20***	-0.20***	-0.15**	-0.08	-0.07
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Turkish SG*Zurich	-0.04	-0.04	0.02	0.07	0.08
	(0.05)	(0.06)	(0.05)	(0.06)	(0.06)
Moroccan SG*Amsterdam	-0.16***	-0.10	-0.02	0.06	0.08
	(0.05)	(0.06)	(0.06)	(0.06)	(0.07)
Moroccan SG*Rotterdam	-0.18***	-0.17***	-0.14***	-0.05	-0.04
	(0.05)	(0.06)	(0.06)	(0.06)	(0.06)
Yugoslav SG*Linz	-0.09*	-0.11**	-0.11**	-0.06	-0.05
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Basel	0.02	0.03	0.07	0.10*	0.12**
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Berlin	-0.12**	-0.12**	-0.06	-0.06	-0.05
	(0.05)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Frankfurt	-0.01	0.00	0.02	0.03	0.04
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Vienna	-0.02	-0.02	-0.02	0.03	0.04
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Yugoslav SG*Zurich	-0.03	-0.03	0.03	0.06	0.08
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)
Observations	3234	2415	2415	2415	2415

Notes: OLS model of the log monthly earnings, as described in the text. All specifications control for age, age^2, gender, city fixed effects and city fixed effects interacted with each second generation group. The coefficients on the main variables are virtually the same than those in table 7. See the note on Table 2.2 for the definitions of the variables used. Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1.

Chapter 3

Immigrant Wage Gaps in the UK and the US[1]

3.1 Introduction

The literature on immigrant assimilation has focused on measuring and explaining the immigrant wage penalty: immigrants with similar education and labour market experience earn less than natives. A very robust finding in the literature is that the immigrants' country of origin is a strong predictor of the size of the wage gap. Holding observable individual characteristics fixed, immigrants from Canada for instance perform better in the US labour market than immigrants from Mexico. This result was presented in Chiswick's early work, Chiswick (1978), and also in Jasso and Rosenzweig (1986) and Borjas (2000), among many others.

This paper focuses on better understanding the country of origin effect. The factors that explain that the immigrants from certain countries of origin perform better in a host country labour market than others may be classified into two wide categories. Some factors are specific to the home country, such as characteristics of the home country that make the worker more productive in any labour market, and some factors are specific to the home country-host country pair. These may be linked to selection effects that would explain that holding the host country fixed, immigrants from different home countries are drawn from different parts of the skill distribution of the home country. Differences in host countries' immigration policy are a factor among others driving immigrant selection. Similar characteristics of the home and host country labour markets may also make it easier for immigrants to adapt. This paper compares the wage gaps by country of origin in two host countries to make an attempt at distinguishing factors specific to the country of origin from factors linked to host country-home country pair.

Among the features of the country of origin that determine the workers' productivity, the literature has mainly focused on the importance of the quality of the country's educational system. Bratsberg and Ragan Jr. (2002) show that there is a wide dispersion in the returns to education acquired abroad for immigrant men in the US depending on where the education was acquired. In a related paper, Bratsberg and Terrell (2002) show in fact that there is a strong

[1] I am grateful to Alan Manning for suggesting the topic of this chapter and providing the data for the UK.

correlation between the returns to foreign education in the US labour market and the quality of the educational system of the country of origin, measured by inputs to education: the expenditures per pupil and the number of pupils per teacher. Sweetman (2004) uses country average scores in international standardized tests as a measure of de facto quality of the educational system and finds also for Canada that the home country school quality is highly correlated to the returns to immigrants' foreign education. Very similar results are also presented for the US in Hanushek and Kimko (2000) and more recently Hanushek and Woessmann (2009). The main objective in these two papers is not to study immigrant performance but to show a causal effect of school quality on productivity and growth. Comparing the performance in the US labour market of immigrants from the same home country who acquired their education either in the US or in the country of origin is one of the strategies presented to isolate the effect of the quality of education on productivity. Hanushek and co-authors use scores in standardized tests and show that these are positively related to immigrants' productivity in the US labour market. The home country effects hence partly reflect differences in the quality of the home countries' educational systems which are not taken into account in the commonly estimated wage regressions.

Although the quality of the educational system is the most widely studied factor, other characteristics of the country of origin are also likely to explain part of the differences in productivity across countries. Sweetman (2004) shows for instance that in Canada, among immigrants with a low educational attainment, those from countries with the highest quality of the educational system tend to perform worse. He argues this is due to selection into education in the country of origin. This effect is also indirectly linked to the educational system of the home country but is independent of the causal effect of schooling on individual productivity.

The second category of explanatory factors of the country of origin effect relates to the characteristics of the home country-host country pair. Firstly, the decisions to migrate and where to migrate to (that is selection and sorting) depend on factors such as the distance between the two countries, the relative returns to the workers' skills in different countries, the existing immigrant networks, etc. Borjas (1987) builds on a Roy model to study the selection into immigration. According to the model, workers make their decision to migrate by comparing the differences in expected returns to skills in the host versus the home country net of the migration costs. Migration costs are usually considered to be correlated to geographic distance between the two countries but also to cultural distance such as how different the languages spoken in both countries are or whether there were colonial ties between the country of origin and the host country in the past. Migrants are hence not randomly drawn from the skill distribution of the home country, and, similarly, different host countries attract immigrants with different skill levels. In a recent paper, Grogger and Hanson (2011) show using international migration flows into 15 OECD countries that there is positive sorting: the more skilled migrants (in the context of the paper, the more *educated* migrants) go to the host countries where the returns to skills are higher. Grogger and Hanson (2011) claim in particular that this mechanism explains why the US and Canada receive more educated immigrants than European countries of the

OECD such as the UK or Germany. One may imagine that this argument is valid not only for observable formal education but also for skills that are usually unobserved and that affect the workers' productivity. Even holding the educational level constant, there may be positive sorting between the home and the host country.

Putting aside immigrant selection, similarities between the two countries may make it easier for immigrant groups to integrate and make use of their skills in the host country labour market. If the productive systems are similar in the two countries, for e.g. the economy is mainly based on agriculture and the techniques used are the same, immigrants will be more productive in the host country. This argument relies on the complementarities between the characteristics of the country of origin and the host country and is independent of immigrant selection.

Most of the literature on immigrant assimilation focuses on one host country. When studying the differences among immigrant groups this makes it impossible to disentangle the assimilation factors that are specific to the characteristics of the home countries and those that are linked to the home country-host country pair. Matoo et al. (2008) focus on one host country, the US, and estimates the differences by country of origin in the likelihood to access skilled jobs for tertiary educated migrants. The estimated differences are then correlated to variables representing home country characteristics and variables measuring immigrant selection. The approach in this project is similar although it uses wages as the outcome of interest instead of occupational attainment. The major difference is that we compare immigrants from the same country of origin in two host countries.

Another strand of the immigration literature compares immigrant performance in different host countries to look at the impact of different immigration policies. However, these papers most often do not exploit the origin of immigrants in detail and do not focus on understanding the differences in assimilation of the different groups. The analysis in this project relates however to a sideline analysis made in Borjas (1991). The paper shows that the differences in immigration policy between the US and Canada after the mid-1960's did not act by selecting different immigrants from the same countries but by changing the country of origin mix of immigrants over time. As part of the argument, Borjas correlates the mean education and the wage gap by country of origin in the US and Canada. These two countries are very similar, performing similar analysis with other host countries allows to better identify the country of origin effect. Moreover, Borjas does not pursue the analysis to understand the country of origin effects.

This paper brings these two strands of the literature closer together by conducting a first descriptive analysis of immigrant wage gaps by country of origin in two major host countries: the UK and the US. The first step of the empirical analysis is to correlate the wage gaps by country of origin in the UK and in the US. The correlation is strong independently of the exact specification considered indicating that characteristics of the country of origin are important in explaining the wage gaps in the UK and the US. Allowing for country of origin specific returns to years of education in the wage regressions allows to show that the returns to years of education are also highly correlated in the two host countries and this accounts for all the observed correlation in the wage gaps. Average country scores in international standardized tests and whether the country is English speaking explain around one third of the differences in home country returns to years of education, both in the UK and the US. Variables usually used to measure immigrant selection, as distance between the home country and the host country, or the size of the immigrant community in the host country, explain a smaller part of the differences in wage gaps, specially in the UK. It remains that two thirds of the country of origin are specific to the host country and remain unexplained.

3.2 The Data and Descriptive Statistics

3.2.1 The Data

The micro data for the analysis come from the labour force surveys for the two countries: the UK Labour Force Survey (LFS) and the US Current Population Survey (CPS). The samples from 2000 to 2011 (2009 for the UK) are pooled in the analysis. Immigrants are defined as foreign-born and the home country is the declared country of birth. One of the first difficulties of the empirical analysis is to identify countries of origin common in both data sets over time. The classification of the country of birth used is the one in the UK LFS in 2000. All foreign-born from *individual* countries for which there are at least 25 observations in both data sets are kept for the analysis. Observations for which the country of birth was classified as a region, such as "other African countries" or "other Europe", are dropped, leaving a total of 72 home countries. Immigrants from these countries represent 83% of the immigrant population aged 15 to 64 in the UK LFS and 73% of that in the US CPS.[2]

An important variable in wage estimations is the educational attainment. The highest diploma obtained is not available for individuals who obtained their degrees abroad in the UK LFS. Instead, the survey has information on the age the individual left full-time education. This variable is used in the analysis as a proxy for years of schooling.[3] Individuals having left full-time education after age 28 are dropped from the UK LFS sample. In the CPS, there is information on the highest degree obtained but not on the years of schooling or on school leaving age. The information on the degree is recoded into years of schooling. Following Bratsberg and Ragan Jr. (2002), among others, the school leaving age is assumed to be six plus the equivalent number of years of schooling required for the degree obtained. In the cases the data are reported in intervals, the years of education are calculated as the midpoint of the interval. For professional and doctoral degrees, the school leaving age is set to 26 and 28.

An important distinction in the literature is whether the education was completed in the host country or abroad. $fgnedu$ is a dummy variable that equals one if the immigrant completed his education abroad. The UK LFS reports the year of arrival in the UK as a continuous variable,

[2] The country classifications of both surveys changed significantly in 2007. More detail on the country of origin has become available. However, pulling 5 years of data produces too small sample sizes. It would probably have been best to use Census data.

[3] Years of schooling are usually approximated by the age the individual left full time education minus 5. In the analysis, *years of education* are the age the individual graduated or left full-time education.

whereas the CPS reports the year of arrival in intervals. The year of arrival is compared to actual or estimated school leaving year to determine the origin of the education. In the US case, immigrants for whom the estimated graduation date falls in the *date of arrival* interval have a missing value for the $fgnedu$ variable[4].

In order to account for characteristics of the country of origin and those of the home country-host country pair, country level variables have also been collected from several sources:

English The home country is considered to be English speaking if the CIA World Factbook listed English as an official language or as the most widely spoken language in the country, following Hanushek and Woessmann (2009) among others.

Distance Data on distances between countries comes from the CEPII's GeoDist database.[5].

GDP per capita for the country of origin comes from the World Bank indicators. The data collected is for 2009 or the latest year available.

Test The measure of the quality of schooling used is one developed in Hanushek and Woessmann (2009). They use test scores in standardized international tests from 1964 to 2003 to construct cross-country comparable measures of cognitive skills for different age groups over time. The measure used in this project is the average test score in math and science, for all education levels from primary through the end of secondary school for all years available. The measure is scaled to the PISA test scores (mean 500 and standard deviation 100 across OECD countries) divided by 100.[6]

Size An approximation of the size of the immigrant community from a given home country in the host country is given by the sample size of immigrants from the home country in the labour force survey samples used.

3.2.2 Immigrants in the UK and in the US

Tables 3.1 to 3.4 present the means of the main variables for all individuals considered in the analysis by country of birth in the UK and in the US. The last two columns represent for each country of origin, the sample size and its share of the total immigrant population in the sample.

The shares of immigrants and their duration of stay give a rough picture of the main immigration waves into the US and the UK. Immigrants in the UK after the Second World War come traditionally from other states of the Common Wealth (India, Pakistan, Bangladesh, West Indies, Kenya) and from Western Europe (Ireland, Germany, France). More recently, the enlargement of the European Union brought a change in the most represented communities in the UK. According to OECD (2012), in 2010, the second largest inflow of migrants to the UK was from Poland, just after India. Immigrants to the US before the 1960's came mainly from

[4]The construction of this variable with the CPS data follows Bratsberg and Ragan Jr. (2002)

[5]For more details on different measures of distances available, please refer to Mayer and Zignago (2011)

[6]For more details on the construction of the variable, refer to Appendix B in Hanushek and Woessmann (2009).

North-Western Europe (UK, Ireland, Germany) but there was a shift in the country of origin of immigrants after the immigration policy change in 1965 emphasizing family reunification. The US immigration policy and its impact on the composition of the immigrant population over time has been much debated in the literature (ex. Borjas (1991), Jasso and Rosenzweig (1986) or LaLonde and Topel (1997)). The weight of immigrants from Asia has increased since the 1970's and that of Western Europe decreased. Mexico represents by far the largest community among the immigrant population.

Comparing mean educational attainment by country of birth in the UK and the US should give a first idea of how the immigrant selection process compares in the two countries. Is the educational attainment of immigrants from the same country of origin similar in the UK and the US? Graphic 3.1 shows the correlation between the years of education of immigrants by country of origin in the UK and the US. The correlation is 0.28 and is statistically significant at the 5 percent level.

Borjas (1991) performs a similar exercise comparing 5 year cohorts of immigrants in the US and Canada between 1960 and 1980, using Census data (1971 and 1981 for Canada and 1970 and 1980 for the US). Borjas (1991) finds a stronger correlation than above. Regressing the mean education of immigrants in the US on that of Canada, he finds that the intercept and the slope are not statistically significantly different from 0 and 1 for most immigrant cohorts. Regressing the mean years of education of immigrants in the UK on that of immigrants in the US yields a constant of 14 and a slope of 0.30, both significantly different from 0 and 1. The R^2 of the regression is 6%, compared to an R^2 between 0.45 and 0.78 depending on the cohort considered in Borjas (1991).[7]. Although it is difficult to make comparisons as the data used in this exercise is much more recent than the one used in Borjas (1991) and mainly because all cohorts are pooled in the analysis, the difference indicates that the immigrant selection process may be more different between the UK and the US than between Canada and the US.

The average years of education is higher for immigrants in the UK than in the US. This is partly, or perhaps totally, due to the way the years of education are calculated in the two data sets. In the UK LFS, the years of education represent the years spent in the educational system whereas in the US, they are the equivalent in years to the highest degree obtained. Immigrants in the UK are also reported to have spent more years in education on average than natives, it is however not clear whether this reflects into higher educational achievement.

The mean education of an immigrant group is related to selection: self-selection of individuals choosing to migrate to the host country but also the selection imposed by the host country's immigration policy.

Bratsberg and Ragan Jr. (2002) show that the educational attainment of immigrants in the US who acquired their education abroad is positively correlated to the GDP of the country of origin, to distance to the host country and whether the country of origin is English speaking. OECD (2008) also shows for a larger number of home and host country pairs that the share of immigrants with tertiary education is negatively correlated to the size of the home country

[7]Dropping 5 countries that are strong outliers from the regression increases the correlation to 0.43. The coefficient of the regression becomes 9.3 and the slope 0.51 The R^2 is 0.17

community in the host country.

Table 3.5 presents, separately for the UK and the US, the coefficients from regressions of the mean years of education by country of origin on several variables believed to impact immigrant selection. One concern when analyzing differences between immigrant groups in the US is the weight of immigrants from Mexico. In all group estimations, Mexico is excluded from the analysis for the US. Table 3.16 in the Appendix presents the results including Mexico.

Using all variables in the regressions accounts for 50% of the cross country variation in immigrant mean education in the US but only for 30% in the UK.

As in the literature mentioned above, distance between the home and the host country is positively correlated to the education of immigrants in the UK and in the US.[8] The size of the home country community is negatively correlated to the years of education. However, this results only holds for the US when Mexico is included. GDP per capita has a different impact in both host countries: it is positively correlated with years of education in the US but not in the UK. Likewise, immigrants from English speaking countries are on avergae more educated than other immigrant groups in the US but less educated in the UK. Countries of origin from where a higher proportion of immigrants completes education before migration have lower mean education in the US but not in the UK.

3.3 Empirical Analysis

The empirical analysis is performed in two steps. First, the wage gap by country of origin is estimated using the UK and the US micro data separately. Then, the home country mean wage gaps are regressed on variables accounting for characteristics of the home country and for variables at the home country-host country level.

3.3.1 The Wage Gap in the UK and in the US

This first section presents preliminary results of the estimations of log wage regressions for the UK and the US, pooling all foreign-born workers of all countries of origin.

The equation estimated is:

$$\log(w_{i,t}) = X_{i,t}\beta + FB_i + \delta_t$$

where $\log(w_{i,t})$ is the log hourly wage of individual i in year t. FB is a dummy for foreign-born. $X_{i,t}$ are individual characteristics, such as gender, age, years of education, race, and years the immigrant has spent in the host country. Race is a dummy variable which indicates if the individual is the race of the majority of the native population, white. Years since migration are introduced as three dummy variables: 5 years or less (ysm1), 5 to 10 years (ysm2) and 10 years or more (ysm3). Indicators for regions and time (δ_t) are also included in all the specifications.

Tables 3.6 and 3.7 present the results for the UK and the US. The coefficients and R^2 of the regressions are very similar for the UK and the US, although the immigrant wage gap is

[8]This results is robust to excluding Australia and New Zealand in the UK estimations.

significantly different. Controlling for gender, race and a cubic in age, immigrants earn 8 log points less than natives in the UK and 24 log points less in the US. Not controlling for differences in race, the gaps are 12 log points in the UK and 25 log points in the US. Taking into account differences in education, the gap increases to 21 log points in the UK and decreases to 11 log points in the US.[9] This result reflects the lower mean education of immigrants in the US as shown in the previous section. Allowing for different experience profiles for men and women or restricting the sample only to men makes little difference to any of the results in the paper.[10]

An important explanatory factor of the immigrant wage penalty is the heterogeneity in returns to foreign education. This result is shown for instance in Bratsberg and Ragan Jr. (2002) or in Bratsberg and Terrell (2002) for the US. The last column introduces different returns to years of education for immigrants educated abroad. The returns to education in the host country are restricted to being the same independently of the country of birth of the worker. The wage gap decreases to 7 log points in the UK and +0.4 log points in the US, insignificantly different from 0. An extra year of education acquired abroad has on average a lower return than one years of education in the host country by 0.9 log points in the UK and 0.7 log points in the US.

3.3.2 The Country of Origin Effect in the UK and in the US

The Wage Gap by Country of Origin

In order to calculate the wage gaps by country of origin, the immigrant dummy is replaced by country of origin fixed effects in the specifications above. Graphics 3.2 and 3.3 show the correlation between the estimated home country effects in the two host countries for the first two specifications above, that is, with and without taking into account years in education. Tables 3.12 and 3.14 in the Appendix show the estimated coefficients and standard errors for the country of origin effects, for the UK and the US.

The graphs show a strong correlation between the home country effects in the two host countries. Three quarters of the countries in the first tier of countries that have higher average wages are the same in both host countries. Between a third and a half are countries from North-Western Europe and most others are English speaking countries such as Canada, New Zealand, Australia and South Africa. Among the worse performing countries are countries from Eastern Europe (more so in the UK), and also countries from Asia, such as Bangladesh, Pakistan, Vietnam, Thailand or the Philippines. Controlling for years of education changes the ranking of some countries but the overall picture remains the same. North-Western European countries and settlement countries are still at the top of the rankings. Some exceptions for individual countries are Mexico in the US: accounting for the low mean education of Mexican workers brings Mexico from the last position to about two-thirds of the ranking. The low wages of Portuguese immigrants are also strongly correlated to their educational background: accounting for education moves the Portuguese gap from rank 64, out of 72, to rank 11 in the

[9] Not accounting for differences in race, the gaps become 26 and 12 log points.

[10] The results are available on request

US.

The correlation between the wage gap by country of origin in the UK and the US estimated from the two specifications above is 0.49 and 0.53, both significant at the 5 percent level. Regressing the mean wage gaps in the UK on the wage gaps in the US, the slope is 0.62 and the constant is -0.02 for the first specification, and 0.91 and -0.11 for the second specification. The R^2 is 29% and 32%. In the main specification, which controls for years of education, the hypothesis that the slope is 1 can not be rejected at conventional levels.[11]

The negative constant indicates that the wage gap for a given country of origin is always more negative in the UK than in the US. This is not the case in the specification that does not control for education. On average immigrants in the UK have more years of education than in the US but receive lower returns to an extra year of education. As mentioned in the previous section, education is measured differently in the two labour force surveys, which can partly or fully drive the results. The CPS has information on the highest degree obtained which is likely to be more comparable in terms of productivity in the labour market than the age the workers left full-time education available in the UK LFS.

The specifications estimated mirror the one made in the previous section comparing the mean education in both host countries and has also been done in Borjas (1991) for Canada and the US. Borjas (1991) shows that the constant and the slope are not significantly different from 0 and 1 for all the 5 year cohorts of immigrants from 1960 to 1980. The R^2 increases a lot from 0.17 for the 1960-64 cohort to 0.89 for the 1975-80 cohort. Again, it is difficult to compare the Borjas (1991) results with the ones presented above since we have not introduced cohort effects in this exercise.

The second step of the analysis is to try to understand which factors explain the differences in labour market performance between countries of origin in the UK and the US. I regress the gaps by country of origin estimated with the micro data on variables that reflect either characteristics of the country of origin or variables linked to the selection of immigrants. Table 3.8 presents the results of the estimations for the UK and the US for the gaps estimated taking into account the years of education. The variables considered are first the distance between the country of origin and the host country, and the size of the home country community in the host country. These variables are associated with immigrant selection to the host country, as discussed above when analyzing the selection on immigrants' education. A bigger distance between home and host country increases the cost of migration and makes it only worthwhile for individuals with the highest expected returns to migration to actually migrate. Often these are believed to be drawn from the upper part of the skill distribution of the home country. A strong country of origin community in the host country on the one hand decreases the cost of migration, which may be an incentive for less productive workers to migrate; on the other hand, larger networks may be beneficial in terms of integration and having access to higher wages

[11]The results of the same analysis excluding race from the estimations are very similar. The correlation between the gaps by country of origin in the UK and in the US is nevertheless stronger, 0.60 in the first specification and 0.63 in the second. This indicates that part of the correlation in wage gaps by country of origin in the two host countries is explained by the race penalty. All results are available on request.

in the host country labour market. The effect is hence ambiguous. Coming from an English speaking country makes transferring skills between the two countries easier and is expected to have a positive sign on immigrant performance. The mean test scores on standardized tests are a measure of the quality of the educational system in the country of origin that should directly positively correlate with the immigrants' productivity. All estimations exclude Mexico for the US, as it has a very large weight on the regressions and is often an outlier. Results including Mexico are presented in the Appendix.

The larger the distance between the home and the host countries, the more positive is the gap for the US and the UK. This would indicate the usual selection mechanism. However, the result is driven by Australia and New Zealand in the UK. Excluding these two countries from the analysis changes the sign of the coefficient. Distance does not seem to have the same selection effect in both host countries. The size of the home country community is positively related to immigrant performance. However this variable is positively correlated with the English speaking country of origin variable. When using all variables in the same specification, the coefficient becomes negative in both the UK and the US. The negative selection effect documented for the mean educational attainment of immigrants by home country seems to be also at play in terms of labour market performance once education has been accounted for. Immigrants from English speaking countries suffer lower wage penalties as expected, and the mean test scores in standardized tests also have a positive impact in both host countries.

The characteristics of the home country play a similar role in explaining the country of origin effects in both host countries. The explanatory power of all variables taken together is nevertheless much lower in the UK than in the US, 0.28 versus 0.65.

The last step of the analysis is to compare the gaps in the UK and in the US in the same framework. The estimation now pools the wage gaps by country of origin in the two host countries and regresses them on country of origin fixed effects, host country fixed effects and selection variables. This estimation also allows to compare the relative effect of home country characteristics in the UK and the US by interacting the home country characteristics with the host country fixed effect. The results are presented in table 3.9. This estimation confirms that the distance effect is positive in both host countries but the coefficient is higher in the US. The size of the immigrant community is positive on average but this effect seems driven by the US. Speaking English brings higher benefits in the US than in the UK, whereas the home country mean test scores are more positively associated with immigrants' wages in the UK than in the US.

The Returns to Education

As mentioned in the previous section, differences in returns to education have been shown to be an important explanatory factor of the country of origin effect. Allowing for different returns to years of education acquired abroad in the wage estimations decreases the correlation between the wage gaps by country of origin in the UK and the US to -0.04, not significantly different from 0. The estimated returns to education acquired abroad by home country are highly correlated in the UK and the US. The correlation is 0.37, significant at the 5 percent level. Graphic 3.4 shows the correlation between the returns to foreign education in the two host countries.

The second step of the analysis is again to regress the estimated returns to education on the home country characteristics and on the variables at the home country-host country level. Considering first the home country characteristics, whether the country is English speaking and the mean test scores in international standardized tests, the R^2 obtained are similar for both host countries: 24 for the UK and 25 for the US. Adding distance and the size of the home country community increases the R^2 by approximately one fifth to 31 for both countries. The effect of distance is negative for both countries if Australia and New Zealand are again excluded in the UK estimation. On average, the two home country variables seem to have very similar effects in the UK and the US. However, more than two thirds of the cross-country differences in returns to education remain unexplained.

Pooling again the estimated gaps in the UK and the US in table 3.11 shows that the selection variables are more relevant in explaining differences in returns to education by country of origin in the US than in the UK. As for the estimated wage gaps, the country of origin's mean test scores in international tests are more important in explaining cross country differences in returns to education in the UK than in the US.

3.4 Conclusion and Further Work

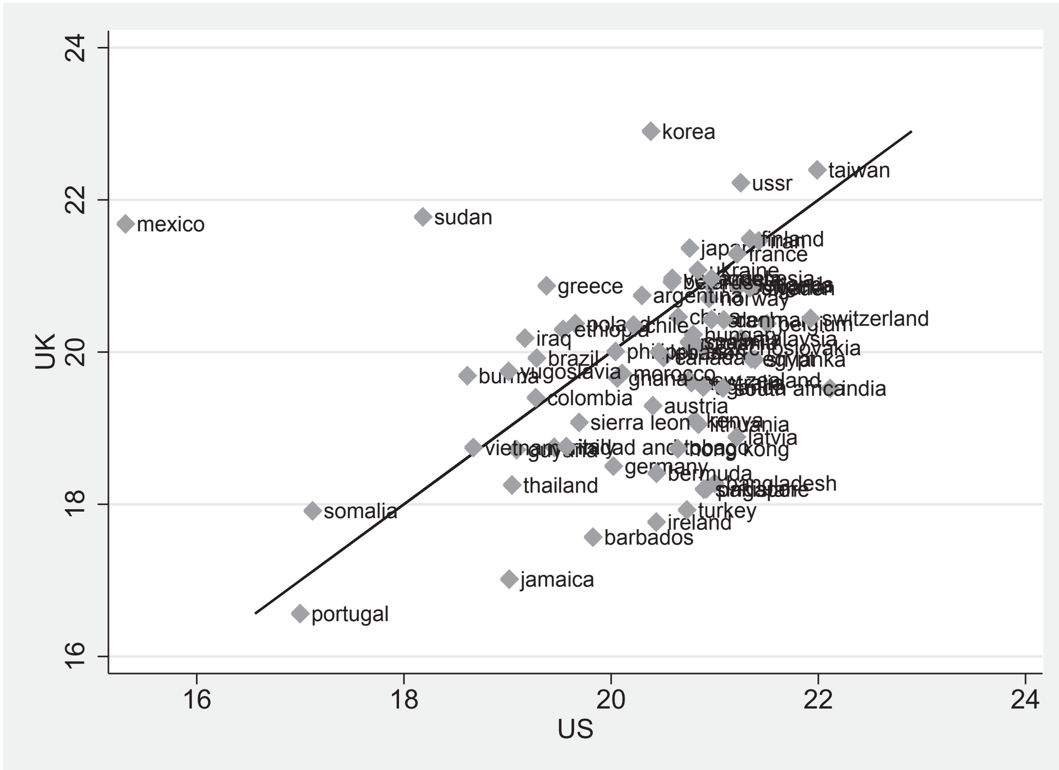
This paper compares the immigrant wage penalty by country of origin in two major host countries and tries to better understand the explanatory factors of the country of origin effect. The correlation between the wage gaps by country of origin in the UK and the US is very strong showing the importance of the country of origin effect. A large part of the country of origin effect is linked to differences in returns to years of education across countries of origin. Measures of the quality of the educational system of the home country and whether the country is English speaking are important predictors of the wage gaps and of the country specific returns to years of education. Variables used in the literature to measure immigrant selection are weaker predictors of immigrants' labour market performance in both host countries, but even more so in the UK. A large part of the country of origin effects are specific to the host country, to the UK and to the US, and remains unexplained.

A first step forward in the analysis would be to collect better data. The main concern with using labour force surveys is small samples. Not all countries of origin of immigrants were considered in the analysis because of very small samples; although the countries of more than 80% of immigrants in either host country are taken into account. Nevertheless, larger samples would allow the estimations to become more precise and to re-do the analysis by immigrant cohort. For now, all immigrant cohorts have been pooled together and the wage regressions control only for the years spent in the country. Being able to compare the same cohort of immigrants in the UK and the US would make the comparisons cleaner and would allow to exploit an extra source of variation: the changes over time in immigrant selection and home country characteristics. Using Census data instead of labour force surveys may allow to improve the analysis in this way.

Cross country differences in returns to years of education are shown to be very relevant in explaining the immigrant wage gap in both host countries. However it was mentioned throughout the analysis that the years of education are measured differently in both countries which may impact the results. Re-doing the analysis with the information on the highest degree obtained in the UK would allow a closer comparison of the estimations in the two countries.

An important aspect when studying the returns to education of immigrants and natives is to take into account the fact that the returns to years of education may be non linear. In general, there are lower returns at lower levels of education. If the years of education of immigrants are lower than the natives', then using years of education in a linear specification tends to over estimate the differences in returns to education between immigrants and natives, and hence to over estimate the wage gap. This is shown to be true for the US in Bratsberg and Ragan Jr. (2002). Once we have a comparable measure of education in both host countries, it will be possible to explore more flexible specifications and check whether the main results change in any way. A situation in which the results may change is if immigrants from a given country of origin, for example Mexico, have higher mean education in the UK than in the US. The wage gaps for Mexicans in the US is over-estimated and this may decrease the estimated correlation between the Mexican wage gap in the two countries.

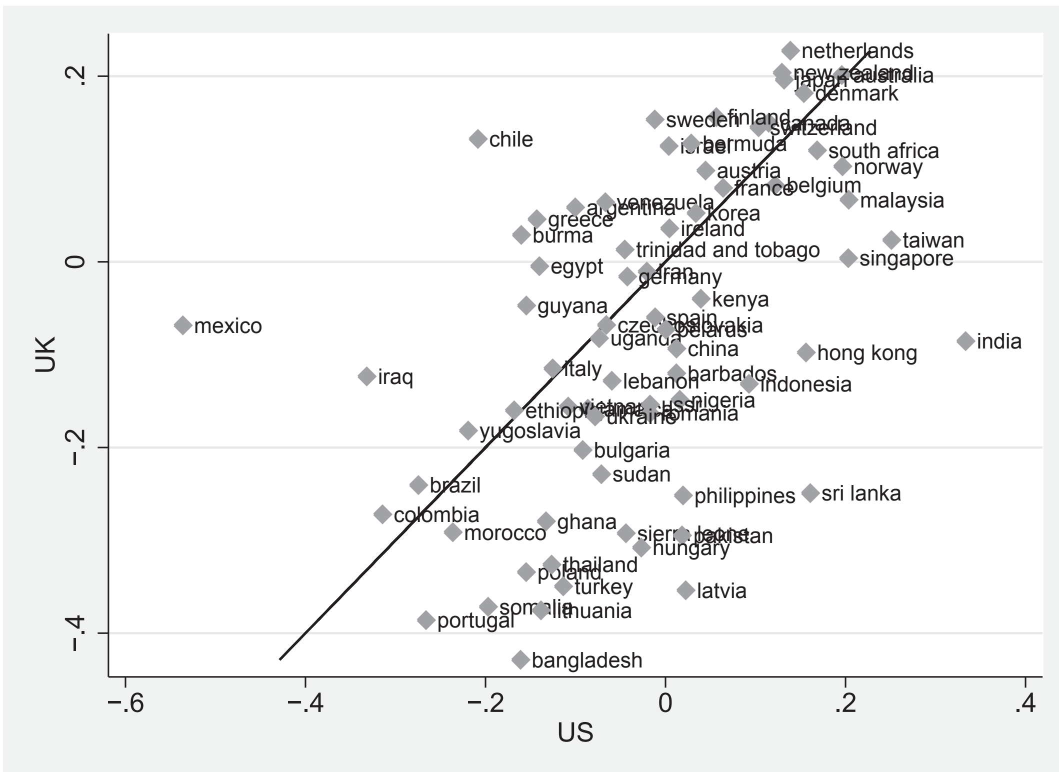
Figure 3.1: Immigrants' Education



Notes: The graphic shows the correlation between the mean age the immigrants left full-time education by country of origin in the UK and the US. The correlation is 0.28 and is significant at the 5 percent level.
The line represents the 45 degree line.

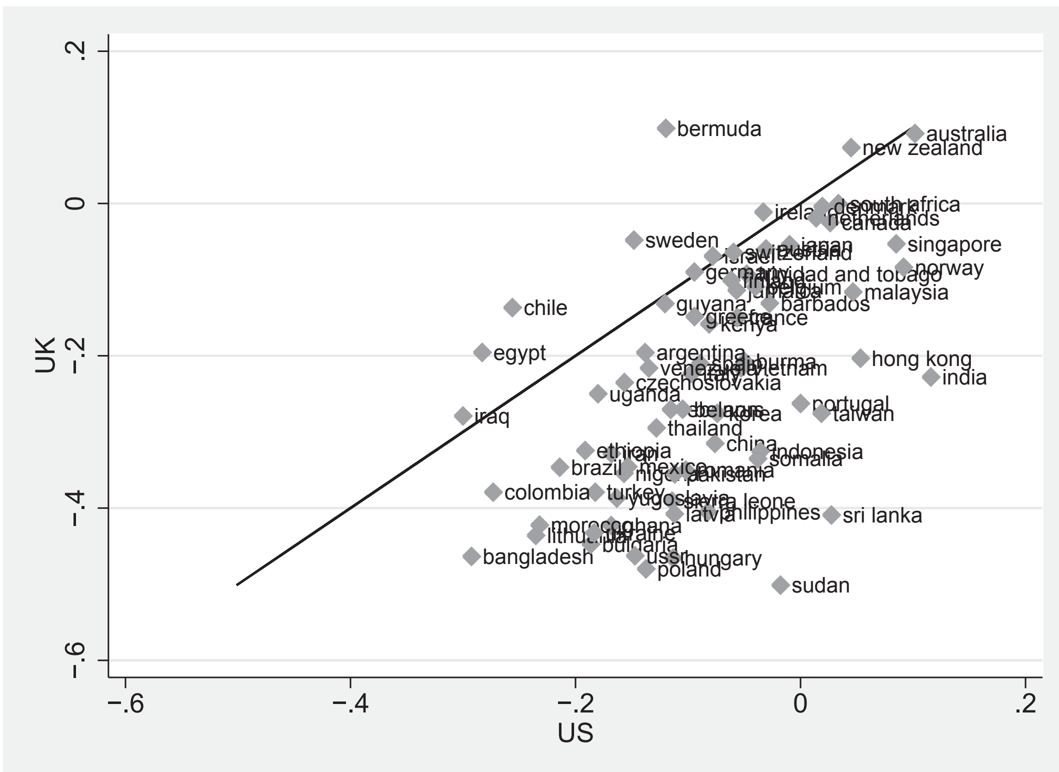
.

Figure 3.2: The Wage Gap



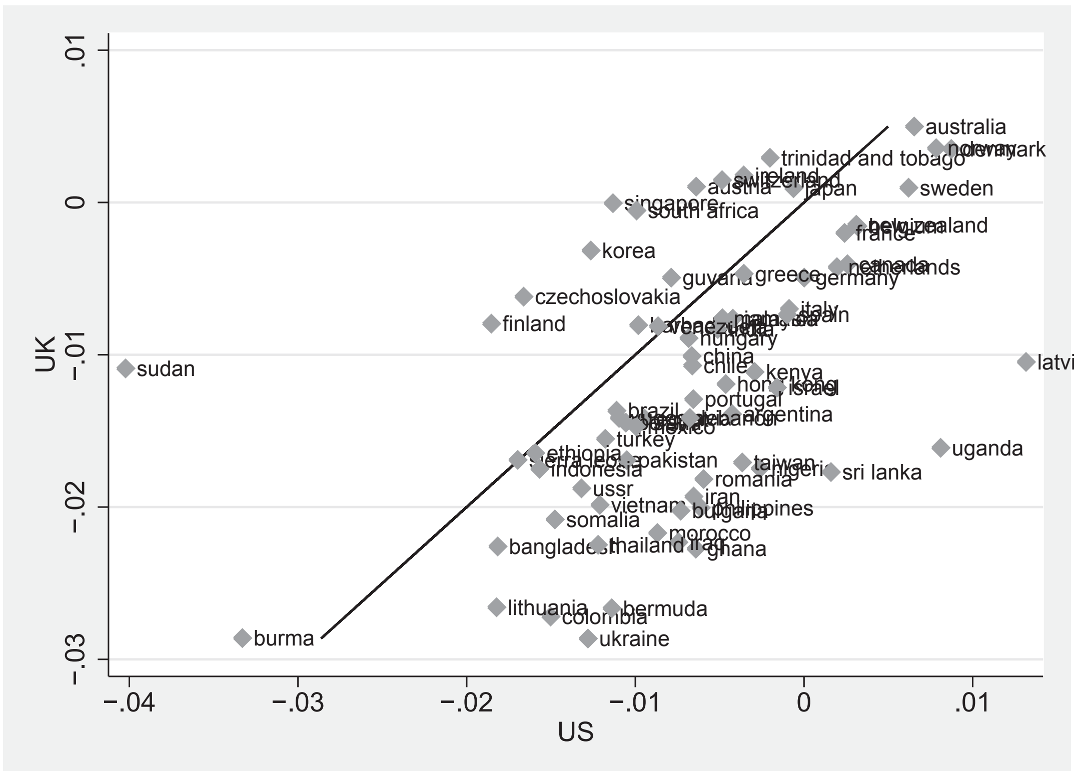
Notes: The wage gaps are estimated from a regression of log hourly wages on country of origin dummies, controlling for gender, race, a cubic in age, region and year fixed effects. The correlation is 0.49, significant at the 5% level.
The line represents the 45 degree line.

Figure 3.3: The Wage Gap (controlling for education)



Notes: The wage gaps are estimated from a regression of log hourly wages on country of origin dummies, controlling for gender, race, a cubic in age, *years of education*, region and year fixed effects. The correlation is 0.53, significant at the 5% level.
The line represents the 45 degree line.

Figure 3.4: The Returns to Education



Notes: The returns to education are estimated from a regression of log hourly wages on country of origin dummies and age left full-time education interacted with the country of origin dummies, and controlling also for gender, race, a cubic in age, age left full-time education, region and year fixed effects.
Belarus is excluded from the graphic as it is a strong outlier in the UK estimation.
The line represents the 45 degree line.

Table 3.1: Descriptive Statistics UK (1/2)

cb	edage	fgnedu	ysm1	ysm2	ysm3	sex	age	race	percent	size
ireland	17.79	0.62	0.12	0.08	0.80	0.48	45.39	0.10	8.80	6065
australia	19.62	0.69	0.43	0.16	0.41	0.49	35.65	0.04	3.30	2271
canada	20.02	0.40	0.24	0.15	0.61	0.46	39.42	0.06	1.84	1271
new zealand	19.65	0.80	0.40	0.17	0.42	0.53	37.46	0.06	1.59	1093
kenya	19.13	0.35	0.08	0.07	0.85	0.55	43.74	0.86	3.85	2655
uganda	19.58	0.37	0.08	0.09	0.83	0.56	43.73	0.91	1.52	1050
ghana	19.78	0.76	0.25	0.18	0.55	0.52	40.63	0.95	1.86	1279
nigeria	20.99	0.73	0.27	0.19	0.54	0.52	39.89	0.91	2.65	1829
sierra leone	19.17	0.70	0.16	0.27	0.57	0.42	40.56	0.93	0.33	229
barbados	17.57	0.44	0.06	0.07	0.87	0.54	48.43	0.92	0.39	267
jamaica	17.07	0.45	0.11	0.11	0.78	0.42	45.55	0.97	2.58	1776
trinidad and tobago	18.80	0.59	0.16	0.17	0.66	0.41	43.29	0.85	0.62	428
guyana	18.80	0.54	0.17	0.09	0.74	0.45	46.67	0.93	0.37	258
bangladesh	18.33	0.49	0.23	0.21	0.56	0.80	34.96	0.99	2.88	1982
india	19.59	0.66	0.27	0.13	0.60	0.58	41.60	0.94	11.52	7935
sri lanka	20.01	0.71	0.24	0.20	0.55	0.64	40.79	0.94	2.03	1397
hong kong	18.78	0.37	0.15	0.12	0.73	0.52	40.34	0.72	1.99	1371
malaysia	20.24	0.43	0.22	0.10	0.68	0.48	41.75	0.72	1.38	951
singapore	18.25	0.13	0.06	0.05	0.89	0.47	41.11	0.22	1.32	910
morocco	19.89	0.80	0.27	0.13	0.60	0.66	39.89	0.69	0.27	187
egypt	19.98	0.49	0.17	0.10	0.74	0.67	47.67	0.48	0.58	398
south africa	19.57	0.71	0.46	0.23	0.31	0.50	35.77	0.15	4.98	3432
pakistan	18.25	0.58	0.22	0.18	0.59	0.79	38.36	0.99	5.43	3739
burma	19.87	0.43	0.15	0.11	0.74	0.64	46.72	0.75	0.16	107
china	20.64	0.77	0.45	0.23	0.31	0.46	36.74	0.97	1.20	824
japan	21.49	0.82	0.47	0.19	0.34	0.51	39.55	0.90	0.57	391
philippines	20.04	0.95	0.50	0.21	0.29	0.32	38.53	0.98	2.28	1573
vietnam	18.89	0.54	0.12	0.18	0.70	0.60	37.61	0.97	0.33	230
iran	21.66	0.54	0.19	0.15	0.66	0.67	42.04	0.70	0.99	683
israel	20.59	0.65	0.30	0.11	0.59	0.64	40.53	0.17	0.25	175
belgium	20.51	0.50	0.32	0.18	0.50	0.54	36.85	0.05	0.48	329
denmark	20.58	0.68	0.34	0.22	0.44	0.42	38.21	0.05	0.42	292
france	21.39	0.71	0.42	0.22	0.36	0.42	35.61	0.09	2.50	1722
italy	18.88	0.68	0.25	0.16	0.59	0.57	41.92	0.04	2.18	1502
netherlands	21.01	0.71	0.35	0.16	0.49	0.51	39.48	0.06	1.08	744
germany	18.66	0.26	0.21	0.15	0.64	0.48	36.75	0.03	6.66	4588

Notes: The variables are the country of birth (cb); the age left education (edage); a dummy for whether the education was completed abroad (fgnedu); a dummy variable which indicates if the individual is white (race); years since migration are three dummy variables: 5 years or less (ysm1), 5 to 10 years (ysm2) and 10 years or more (ysm3); the proportion of immigrants from the same country of birth in the sample (percent); and the sample size (size).

Table 3.2: Descriptive Statistics UK (2/2)

cb	edage	fgnedu	ysm1	ysm2	ysm3	sex	age	race	percent	size
bulgaria	21.04	0.87	0.58	0.30	0.12	0.47	35.32	0.10	0.43	296
czechoslovakia	20.06	0.76	0.44	0.26	0.29	0.34	36.22	0.09	0.14	97
hungary	20.23	0.83	0.65	0.08	0.27	0.45	36.39	0.09	0.36	245
poland	20.40	0.92	0.83	0.09	0.08	0.55	31.37	0.09	5.41	3730
romania	20.13	0.92	0.69	0.20	0.11	0.55	32.44	0.09	0.45	310
austria	19.38	0.49	0.22	0.15	0.64	0.49	45.22	0.03	0.35	244
switzerland	20.59	0.63	0.31	0.15	0.54	0.43	40.54	0.05	0.27	186
greece	21.13	0.63	0.42	0.20	0.38	0.61	35.73	0.09	0.56	386
portugal	16.56	0.84	0.40	0.21	0.38	0.52	37.42	0.16	1.51	1043
spain	20.27	0.78	0.40	0.19	0.41	0.46	36.59	0.07	1.38	948
finland	21.61	0.77	0.48	0.19	0.33	0.26	36.12	0.02	0.29	197
norway	20.95	0.59	0.36	0.18	0.45	0.43	38.54	0.09	0.29	199
sweden	20.97	0.71	0.50	0.19	0.31	0.43	35.98	0.04	0.52	359
yugoslavia	19.75	0.83	0.25	0.37	0.39	0.65	36.43	0.09	0.37	256
turkey	17.97	0.79	0.30	0.26	0.44	0.71	35.57	0.30	1.05	724
ussr	22.41	0.83	0.62	0.27	0.11	0.44	35.02	0.13	0.07	45
ethiopia	20.30	0.70	0.19	0.26	0.56	0.53	37.87	0.86	0.20	135
somalia	18.02	0.71	0.37	0.38	0.25	0.62	32.70	0.99	0.34	237
mexico	21.94	0.71	0.40	0.22	0.38	0.41	37.31	0.40	0.15	105
argentina	21.00	0.67	0.38	0.21	0.42	0.53	39.87	0.13	0.20	141
brazil	20.00	0.80	0.62	0.15	0.23	0.46	33.53	0.32	0.75	515
chile	20.45	0.70	0.24	0.11	0.66	0.38	43.89	0.33	0.13	93
colombia	19.53	0.86	0.32	0.27	0.40	0.42	37.93	0.63	0.36	251
venezuela	21.09	0.53	0.35	0.16	0.49	0.55	39.36	0.28	0.10	69
iraq	20.36	0.68	0.28	0.24	0.48	0.79	38.72	0.77	0.58	402
lebanon	20.12	0.59	0.17	0.17	0.66	0.78	40.91	0.51	0.24	162
korea	23.03	0.85	0.53	0.26	0.21	0.52	38.42	0.97	0.16	109
belarus	20.92	0.96	0.64	0.29	0.07	0.43	33.04	0.00	0.04	28
lithuania	19.09	0.94	0.83	0.15	0.02	0.52	31.21	0.16	0.78	539
latvia	18.88	0.90	0.79	0.15	0.06	0.47	32.37	0.09	0.26	177
ukraine	21.20	0.87	0.53	0.32	0.15	0.36	35.35	0.15	0.29	197
sudan	21.84	0.75	0.25	0.30	0.45	0.74	39.94	0.87	0.17	120
indonesia	20.97	0.72	0.37	0.32	0.29	0.46	37.87	0.80	0.14	99
bermuda	18.41	0.11	0.11	0.29	0.61	0.32	39.39	0.11	0.04	28
taiwan	22.39	0.70	0.51	0.17	0.29	0.34	35.57	0.94	0.05	35
thailand	18.34	0.77	0.42	0.17	0.41	0.24	37.05	0.93	0.39	268
Natives	17.43	0.00	0.00	0.00	0.00	0.52	40.23	809154		
Foreign-born	19.35	0.62	0.30	0.15	0.54	0.54	39.25	68908		

Notes: The variables are the country of birth (cb); the age left education (edage); a dummy for whether the education was completed abroad (fgnedu); a dummy variable which indicates if the individual is white (race); years since migration are three dummy variables: 5 years or less (ysm1), 5 to 10 years (ysm2) and 10 years or more (ysm3); the proportion of immigrants from the same country of birth in the sample (percent); and the sample size (size).

Table 3.3: Descriptive Statistics US (1/2)

cb	edage	fgnedu	ysm1	ysm2	ysm3	sex	age	race	percent	size
ireland	20.44	0.66	0.07	0.14	0.76	0.59	43.90	0.01	0.36	459
australia	20.77	0.55	0.21	0.18	0.58	0.62	38.70	0.05	0.24	306
canada	20.51	0.42	0.10	0.14	0.74	0.50	41.88	0.08	2.76	3555
new zealand	20.75	0.70	0.14	0.21	0.58	0.53	40.78	0.11	0.10	125
kenya	20.81	0.62	0.20	0.32	0.44	0.58	36.38	0.92	0.26	334
uganda	20.89	0.66	0.11	0.39	0.47	0.64	40.86	0.97	0.03	36
ghana	20.06	0.81	0.17	0.30	0.48	0.59	40.63	0.97	0.41	526
nigeria	21.41	0.69	0.19	0.23	0.53	0.61	40.53	0.95	0.59	765
sierra leone	19.69	0.87	0.10	0.36	0.48	0.55	40.05	0.98	0.03	42
barbados	19.82	0.55	0.03	0.09	0.87	0.39	44.89	0.96	0.20	262
jamaica	19.02	0.59	0.07	0.14	0.76	0.42	42.39	0.95	2.08	2674
trinidad and tobago	19.45	0.59	0.04	0.14	0.79	0.44	43.10	0.87	0.70	898
guyana	19.09	0.62	0.07	0.17	0.73	0.51	41.49	0.85	0.75	966
bangladesh	21.00	0.71	0.13	0.27	0.54	0.72	38.53	0.93	0.38	491
india	22.11	0.67	0.21	0.25	0.48	0.63	38.39	0.92	5.16	6643
sri lanka	21.39	0.75	0.10	0.15	0.73	0.53	41.11	0.98	0.05	62
hong kong	20.64	0.37	0.05	0.10	0.82	0.52	41.29	0.95	0.65	831
malaysia	21.26	0.55	0.11	0.22	0.62	0.49	38.97	0.94	0.21	274
singapore	20.89	0.64	0.13	0.17	0.63	0.52	39.63	0.88	0.08	104
morocco	20.11	0.71	0.22	0.23	0.51	0.69	39.63	0.16	0.17	217
egypt	21.35	0.71	0.11	0.22	0.61	0.68	42.56	0.14	0.35	456
south africa	21.08	0.64	0.14	0.26	0.57	0.50	39.35	0.25	0.25	317
pakistan	20.92	0.63	0.14	0.23	0.59	0.68	39.30	0.86	0.72	932
burma	18.61	0.78	0.25	0.27	0.47	0.58	39.52	0.96	0.22	288
china	20.65	0.71	0.15	0.23	0.57	0.50	41.50	0.98	4.11	5286
japan	20.76	0.42	0.13	0.09	0.76	0.46	41.50	0.74	1.59	2042
philippines	20.04	0.64	0.10	0.15	0.72	0.41	42.59	0.94	7.51	9663
vietnam	18.67	0.60	0.08	0.13	0.75	0.53	40.85	0.98	3.18	4097
iran	21.42	0.55	0.09	0.12	0.77	0.57	43.10	0.10	0.84	1078
israel	20.97	0.52	0.12	0.11	0.73	0.63	41.00	0.03	0.29	379
belgium	21.50	0.37	0.11	0.10	0.74	0.53	41.98	0.03	0.12	155
denmark	21.09	0.52	0.13	0.15	0.64	0.54	44.01	0.05	0.08	104
france	21.22	0.40	0.16	0.11	0.71	0.54	41.55	0.08	0.55	706
italy	19.57	0.32	0.06	0.06	0.87	0.61	45.48	0.02	0.97	1251
netherlands	20.96	0.48	0.09	0.10	0.79	0.58	45.85	0.10	0.28	360
germany	20.02	0.17	0.04	0.05	0.90	0.47	39.67	0.09	3.24	4175

Notes: The variables are the country of birth (cb); the age left education (edage); a dummy for whether the education was completed abroad (fgnedu); a dummy variable which indicates if the individual is white (race); years since migration are three dummy variables: 5 years or less (ysm1), 5 to 10 years (ysm2) and 10 years or more (ysm3); the proportion of immigrants from the same country of birth in the sample (percent); and the sample size (size).

Table 3.4: Descriptive Statistics US (2/2)

cb	edage	fgnedu	ysm1	ysm2	ysm3	sex	age	race	percent	size
bulgaria	21.22	0.85	0.16	0.32	0.51	0.52	36.89	0.00	0.05	63
czechoslovakia	20.90	0.59	0.20	0.16	0.61	0.53	40.49	0.01	0.20	262
hungary	20.79	0.60	0.10	0.19	0.69	0.55	43.76	0.01	0.16	201
poland	19.65	0.65	0.11	0.20	0.65	0.51	40.93	0.00	1.23	1577
romania	20.75	0.64	0.13	0.20	0.63	0.51	39.04	0.01	0.33	425
austria	20.40	0.26	0.08	0.07	0.82	0.51	46.86	0.04	0.13	169
switzerland	21.92	0.55	0.17	0.16	0.65	0.58	42.07	0.04	0.12	159
greece	19.38	0.40	0.02	0.05	0.91	0.62	46.39	0.04	0.35	450
portugal	17.00	0.44	0.04	0.06	0.90	0.51	43.65	0.05	0.86	1102
spain	20.79	0.43	0.08	0.12	0.77	0.51	39.80	0.04	0.42	537
finland	21.34	0.66	0.15	0.16	0.66	0.44	42.82	0.00	0.05	62
norway	20.95	0.41	0.13	0.07	0.77	0.54	44.17	0.03	0.09	110
sweden	21.34	0.61	0.16	0.25	0.55	0.53	40.21	0.02	0.13	165
yugoslavia	19.01	0.71	0.17	0.31	0.44	0.53	39.24	0.01	0.69	886
turkey	20.73	0.54	0.18	0.20	0.59	0.62	37.13	0.05	0.32	416
ussr	21.25	0.67	0.16	0.27	0.50	0.46	39.09	0.01	1.12	1447
ethiopia	19.54	0.75	0.23	0.25	0.46	0.54	37.16	0.93	0.50	641
somalia	17.12	0.74	0.17	0.45	0.32	0.50	32.06	0.93	0.16	204
mexico	15.31	0.71	0.14	0.20	0.62	0.66	36.35	0.04	42.43	54613
argentina	20.30	0.69	0.20	0.20	0.55	0.60	41.38	0.04	0.55	709
brazil	19.28	0.74	0.29	0.27	0.41	0.51	36.97	0.07	0.98	1262
chile	20.22	0.61	0.14	0.18	0.66	0.56	41.96	0.04	0.36	464
colombia	19.27	0.66	0.14	0.19	0.64	0.47	40.96	0.06	2.37	3049
venezuela	20.59	0.60	0.15	0.28	0.51	0.51	37.72	0.08	0.57	736
iraq	19.17	0.75	0.17	0.23	0.56	0.59	41.06	0.12	0.26	329
lebanon	20.46	0.44	0.10	0.13	0.74	0.67	38.32	0.09	0.39	499
korea	20.38	0.54	0.08	0.11	0.78	0.45	40.36	0.96	2.75	3535
belarus	20.58	0.35	0.09	0.21	0.70	0.64	33.02	0.00	0.04	53
lithuania	20.84	0.72	0.29	0.33	0.28	0.45	38.18	0.05	0.08	108
latvia	21.21	0.25	0.18	0.26	0.50	0.39	38.24	0.03	0.03	38
ukraine	20.84	0.72	0.20	0.28	0.47	0.51	39.05	0.01	0.55	711
sudan	18.18	0.81	0.20	0.33	0.44	0.72	36.22	0.88	0.06	82
indonesia	20.97	0.53	0.15	0.24	0.57	0.53	38.92	0.88	0.27	350
bermuda	20.44	0.19	0.06	0.06	0.88	0.54	41.83	0.29	0.04	48
taiwan	21.99	0.54	0.08	0.10	0.79	0.47	42.03	0.95	1.09	1405
thailand	19.04	0.48	0.08	0.11	0.78	0.40	38.17	0.91	0.76	980
Natives	19.53	0.00	0.00	0.00	0.00	0.50	38.97	936723		
Foreign-born	18.16	0.63	0.13	0.18	0.65	0.57	38.95	128706		

Notes: The variables are the country of birth (cb); the age left education (edage); a dummy for whether the education was completed abroad (fgnedu); a dummy variable which indicates if the individual is white (race); years since migration are three dummy variables: 5 years or less (ysm1), 5 to 10 years (ysm2) and 10 years or more (ysm3); the proportion of immigrants from the same country of birth in the sample (percent); and the sample size (size).

Table 3.5: Immigrants' Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance	5.76e-05*					3.76e-05	3.98e-05	2.60e-05					4.52e-05	3.04e-05
	(3.09e-05)					(3.45e-05)	(3.04e-05)	(3.04e-05)					(3.50e-05)	(3.78e-05)
English		0.431*				0.384	0.326		-0.647**				-0.503	-0.483
		(0.249)				(0.269)	(0.225)		(0.257)				(0.380)	(0.389)
Percent			5.101			8.093	14.73**			-8.331**			-1.809	-2.063
			(5.271)			(6.829)	(5.948)			(3.629)			(4.237)	(4.335)
Fgnedu				0.649		-0.971	2.445**				2.387***		1.749**	1.511**
				(0.859)		(0.827)	(1.090)				(0.648)		(0.682)	(0.725)
Sex						6.498***	6.238***						-1.378	-2.016
						(1.701)	(1.445)						(1.126)	(1.299)
Age						-0.0591	-0.150***						-0.0383	-0.0451
						(0.0601)	(0.0533)						(0.0387)	(0.0399)
Gdp					1.78e-06		4.09e-05***					-9.23e-06		-9.16e-06
					(7.99e-06)		(1.11e-05)					(7.13e-06)		(7.80e-06)
UK								Yes	Yes	Yes	Yes	Yes	Yes	Yes
US	Yes	Yes	Yes	Yes	Yes	Yes	Yes							
Constant	19.71***	20.09***	20.12***	19.89***	20.22***	19.17***	20.33***	19.14***	19.64***	19.65***	17.79***	19.45***	20.54***	21.55***
	(0.309)	(0.157)	(0.194)	(0.509)	(0.162)	(2.896)	(2.495)	(0.215)	(0.191)	(0.205)	(0.423)	(0.188)	(1.850)	(2.046)
Observations	70	71	71	71	66	70	66	71	72	72	72	67	71	67
R-squared	0.049	0.042	0.013	0.008	0.001	0.308	0.506	0.010	0.083	0.070	0.162	0.025	0.286	0.310

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
The same estimations including Mexico for the US are presented in table 3.16 in the Appendix.
The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the proportion of immigrants from the same country of birth in the sample (Percent); the proportion of immigrants from the country of origin who completed their education in the home country (Fgnedu); the proportion of women in the immigrant community (Sex); the GDP per capita of the country of origin in 2009 (or the latest date available).

Table 3.6: The Wage Gap in the UK

	(1)	(2)	(3)
Sex	0.231***	0.231***	0.232***
	(0.00134)	(0.00125)	(0.00125)
Ysm2	0.0686***	0.0642***	0.0431***
	(0.00859)	(0.00844)	(0.00837)
Ysm3	0.0737***	0.126***	0.0629***
	(0.00619)	(0.00612)	(0.00631)
Exp.		0.0762***	0.0767***
		(0.000447)	(0.000445)
Edage		0.0971***	0.0997***
		(0.000290)	(0.000295)
Fb	-0.0825***	-0.211***	-0.141***
	(0.00490)	(0.00493)	(0.00520)
Age	0.213***		
	(0.00153)		
Fgnedu			-0.0331***
			(0.00107)
Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Race	Yes	Yes	Yes
Constant	-1.437***	-0.489***	1.211***
	(0.0183)	(0.00743)	(0.00500)
Observations	570,066	550,115	550,115
R-squared	0.236	0.341	0.343

Notes: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The variables are defined in table 3.1.
In column (3), the immigrant dummy represents the effects at the mean years of education.

Table 3.7: The Wage Gap in the US

	(1)	(2)	(3)
Sex	0.211***	0.225***	0.241***
	(0.00142)	(0.00132)	(0.00101)
Ysm2	0.0135*	0.0134**	0.0204**
	(0.00745)	(0.00670)	(0.00591)
Ysm3	0.0894***	0.0925***	0.120***
	(0.00607)	(0.00546)	(0.00482)
Exp		0.0714***	0.0692***
		(0.000426)	(0.000358)
Edage		0.104***	0.110***
		(0.000273)	(0.000262)
Fb	-0.238***	-0.106***	-0.1663
	(0.00550)	(0.00503)	(0.00454)
Age	0.211***		
	(0.00160)		
Fg Education			-0.0364***
			(0.000505)
Year FE	Yes	Yes	Yes
Region FE	Yes	Yes	Yes
Race	Yes	Yes	Yes
Constant	-0.985***	-0.177***	1.832***
	(0.0198)	(0.00793)	(0.00553)
Observations	1,065,429	1,065,429	1,053,531
R-squared	0.212	0.324	0.338

Notes: Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1
The variables are defined in table 3.1.
In column (3), the immigrant dummy represents the effects at the mean years of education.

Table 3.8: The Country of Origin Effect

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance	5.80e-06				-5.82e-07	2.97e-06	4.76e-06*				4.38e-06	7.57e-06***
	(4.43e-06)				(5.17e-06)	(5.80e-06)	(2.85e-06)				(2.95e-06)	(2.70e-06)
Size		1.03e-05			-1.90e-06	-4.67e-06		9.06e-06**			-9.27e-07	-4.17e-06
		(7.89e-06)			(9.43e-06)	(1.08e-05)		(3.61e-06)			(4.42e-06)	(4.15e-06)
English			0.124***		0.130**	0.154**			0.0900***		0.0893***	0.172***
			(0.0363)		(0.0503)	(0.0639)			(0.0206)		(0.0243)	(0.0245)
Test				0.0656		0.0991**				0.0412*		0.110***
				(0.0393)		(0.0371)				(0.0234)		(0.0175)
US							Yes	Yes	Yes	Yes	Yes	Yes
UK	Yes	Yes	Yes	Yes	Yes	Yes						
Constant	-0.235***	-0.235***	-0.272***	-0.479**	-0.267***	-0.712***	-0.115***	-0.106***	-0.109***	-0.250**	-0.145***	-0.683***
	(0.0312)	(0.0307)	(0.0270)	(0.181)	(0.0406)	(0.184)	(0.0285)	(0.0171)	(0.0130)	(0.107)	(0.0268)	(0.0914)
Observations	71	72	72	50	71	50	70	71	71	49	70	49
R-squared	0.024	0.024	0.143	0.055	0.143	0.278	0.040	0.084	0.217	0.062	0.240	0.650

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
The same estimations including Mexico for the US are presented in Table 3.17 in the Appendix.
The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the sample size of immigrants from the same country of birth (Size); the mean test score in international standardized tests in the country of origin (Test).

Table 3.9: The Country of Origin Effect 2

	(1)	(2)	(3)	(4)	(5)	(6)
Distance		1.53e-05***			8.55e-06	8.74e-06
		(4.66e-06)			(6.21e-06)	(8.06e-06)
Distance*US					1.07e-05**	8.80e-06
					(4.63e-06)	(6.00e-06)
US	0.171***	0.120***	0.139***	0.672***	-0.00989	0.232
	(0.0160)	(0.0215)	(0.0214)	(0.162)	(0.0336)	(0.186)
Size		1.20e-05*			-1.01e-06	-2.38e-06
		(6.18e-06)			(7.77e-06)	(9.39e-06)
Size*US					1.48e-05**	1.41e-05
					(7.19e-06)	(9.33e-06)
English*US			0.0693**		0.0389	0.0356
			(0.0312)		(0.0394)	(0.0534)
Test*US				-0.108***		-0.0492
				(0.0355)		(0.0358)
Constant	-0.224***	-0.351***	-0.222***	-0.214***	-0.277***	-0.261***
	(0.0105)	(0.0396)	(0.0103)	(0.0123)	(0.0527)	(0.0671)
Observations	142	140	142	98	140	98
R-squared	0.863	0.884	0.872	0.870	0.918	0.910
CO FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the sample size of immigrants from the same country of birth (Size); the mean test score in international standardized tests in the country of origin (Test). CO FE are country of origin fixed effects and US is a dummy indicating gaps estimated for the US.

Table 3.10: The Returns to Education

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
English	0.00493***				0.00452*	0.00578*	0.00311**				0.00243*	0.00539***
	(0.00183)				(0.00249)	(0.00288)	(0.00129)				(0.00145)	(0.00170)
Test		0.00445**				0.00570***		0.00229*				0.00347***
		(0.00175)				(0.00168)		(0.00115)				(0.00121)
Distance			1.35e-09		-1.70e-07	3.84e-08			-3.97e-07**		-5.04e-07***	-3.12e-07
			(2.20e-07)		(2.56e-07)	(2.62e-07)			(1.59e-07)		(1.75e-07)	(1.87e-07)
Size				9.33e-07**	4.64e-07	2.79e-07				1.73e-07	2.51e-07	-6.85e-08
				(3.76e-07)	(4.68e-07)	(4.87e-07)				(2.16e-07)	(2.63e-07)	(2.88e-07)
UK	Yes	Yes	Yes	Yes	Yes	Yes						
US							Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.0118***	-0.0276***	-0.00905***	-0.0119***	-0.0120***	-0.0375***	-0.00787***	-0.0160***	-0.00284*	-0.00726***	-0.00377**	-0.0202***
	(0.00136)	(0.00809)	(0.00155)	(0.00146)	(0.00201)	(0.00832)	(0.000814)	(0.00523)	(0.00159)	(0.00102)	(0.00159)	(0.00634)
Observations	71	50	70	71	70	50	71	49	70	71	70	49
R-squared	0.095	0.118	0.000	0.082	0.127	0.310	0.078	0.078	0.084	0.009	0.176	0.312

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
The same estimations including Mexico for the US are presented in Table 3.18 in the Appendix.
The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the sample size of immigrants from the same country of birth (Size); the mean test score in international standardized tests in the country of origin (Test).

Table 3.11: The Returns to Education 2

	(1)	(2)	(3)	(4)	(5)	(6)
Distance		3.34e-07			5.10e-07	3.86e-07
		(2.01e-07)			(3.18e-07)	(3.28e-07)
Distance*US					-1.10e-07	-4.02e-07
					(2.37e-07)	(2.44e-07)
US	0.00518***	0.00397***	0.00565***	0.0217***	0.00479***	0.0234***
	(0.000697)	(0.000928)	(0.000960)	(0.00641)	(0.00172)	(0.00755)
Size		9.61e-07***			1.23e-06***	1.24e-06***
		(2.67e-07)			(3.98e-07)	(3.82e-07)
Size*US					-3.24e-07	-4.35e-07
					(3.68e-07)	(3.79e-07)
English*US			-0.00101		0.000997	0.00151
			(0.00140)		(0.00202)	(0.00217)
Test*US				-0.00359**		-0.00339**
				(0.00140)		(0.00146)
Constant	-0.0105***	-0.0155***	-0.0105***	-0.00924***	-0.0172***	-0.0155***
	(0.000459)	(0.00171)	(0.000464)	(0.000486)	(0.00270)	(0.00273)
Observations	142	140	142	98	140	98
R-squared	0.883	0.902	0.884	0.895	0.903	0.923
CO FE	Yes	Yes	Yes	Yes	Yes	Yes

Notes: Standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the sample size of immigrants from the same country of birth (Size); the mean test score in international standardized tests in the country of origin (Test). CO FE are country of origin fixed effects and US is a dummy indicating gaps estimated for the US.

3.A Appendix

The tables below present the estimated mean wage gap (and corresponding standard error) by country of origin for the UK and the US. The first specification does not include education, whereas the second one does.

3.A.1 More Tables

The tables below present the estimated mean wage gap (and corresponding standard error) by country of origin for the UK and the US. The first specification does not include education, whereas the second one does.

Table 3.12: The Country of Origin Effect in the UK (1/2)

cb	Gap 1	SE Gap 1	Gap 2	SE Gap 2
ireland	0.036	0.011	-0.011	0.010
australia	0.201	0.014	0.092	0.014
canada	0.150	0.020	-0.025	0.019
new zealand	0.204	0.021	0.073	0.019
kenya	-0.040	0.016	-0.158	0.015
uganda	-0.082	0.026	-0.250	0.026
ghana	-0.280	0.020	-0.423	0.020
nigeria	-0.149	0.019	-0.354	0.020
sierra leone	-0.292	0.047	-0.391	0.047
barbados	-0.120	0.035	-0.131	0.034
jamaica	-0.158	0.017	-0.114	0.016
trinidad and tobago	0.013	0.034	-0.093	0.036
guyana	-0.047	0.043	-0.131	0.040
bangladesh	-0.429	0.022	-0.464	0.021
india	-0.085	0.010	-0.228	0.010
sri lanka	-0.249	0.022	-0.409	0.021
hong kong	-0.098	0.023	-0.203	0.022
malaysia	0.067	0.024	-0.116	0.024
singapore	0.004	0.021	-0.053	0.020
morocco	-0.291	0.059	-0.422	0.064
egypt	-0.005	0.041	-0.196	0.040
south africa	0.120	0.011	0.000	0.011
pakistan	-0.295	0.015	-0.355	0.015
burma	0.029	0.090	-0.208	0.089
china	-0.093	0.026	-0.315	0.028
japan	0.196	0.039	-0.054	0.037
philippines	-0.252	0.014	-0.405	0.016
vietnam	-0.156	0.062	-0.216	0.055
iran	-0.011	0.033	-0.328	0.035
israel	0.125	0.070	-0.069	0.069
belgium	0.083	0.034	-0.109	0.031
denmark	0.182	0.038	-0.004	0.038
france	0.080	0.016	-0.150	0.016
italy	-0.115	0.020	-0.223	0.019
netherlands	0.228	0.025	-0.019	0.025
germany	-0.016	0.011	-0.090	0.010

Notes: The country of origin effects are estimated from a regression of log hourly wages on country of origin dummies, controlling for gender, race, a cubic in age, region and year fixed effects. The second specification controls also for *years of education*.

Table 3.13: The Country of Origin Effect in the UK (2/2)

cb	Gap 1	SE Gap 1	Gap 2	SE Gap 2
bulgaria	-0.203	0.044	-0.448	0.045
czechoslovakia	-0.068	0.070	-0.235	0.070
hungary	-0.308	0.049	-0.466	0.046
poland	-0.334	0.009	-0.480	0.009
romania	-0.163	0.047	-0.350	0.046
austria	0.098	0.046	-0.060	0.044
switzerland	0.145	0.051	-0.064	0.052
greece	0.046	0.034	-0.149	0.032
portugal	-0.386	0.019	-0.263	0.018
spain	-0.060	0.021	-0.210	0.020
finland	0.156	0.047	-0.101	0.050
norway	0.103	0.046	-0.084	0.044
sweden	0.153	0.034	-0.048	0.031
yugoslavia	-0.182	0.052	-0.387	0.048
turkey	-0.350	0.034	-0.379	0.031
ussr	-0.154	0.102	-0.463	0.109
ethiopia	-0.160	0.068	-0.324	0.065
somalia	-0.372	0.041	-0.335	0.054
mexico	-0.068	0.072	-0.346	0.077
argentina	0.059	0.055	-0.196	0.060
brazil	-0.240	0.030	-0.346	0.032
chile	0.132	0.086	-0.137	0.083
colombia	-0.272	0.053	-0.379	0.056
venezuela	0.064	0.103	-0.216	0.103
iraq	-0.123	0.037	-0.279	0.041
lebanon	-0.128	0.077	-0.270	0.065
korea	0.053	0.089	-0.276	0.083
belarus	-0.073	0.067	-0.270	0.073
lithuania	-0.375	0.028	-0.436	0.032
latvia	-0.354	0.041	-0.408	0.044
ukraine	-0.167	0.053	-0.433	0.060
sudan	-0.229	0.067	-0.501	0.070
indonesia	-0.131	0.067	-0.326	0.069
bermuda	0.127	0.097	0.099	0.115
taiwan	0.024	0.096	-0.275	0.099
thailand	-0.326	0.035	-0.295	0.038

Notes: The country of origin effects are estimated from a regression of log hourly wages on country of origin dummies, controlling for gender, race, a cubic in age, region and year fixed effects. The second specification controls also for *years of education*.

Table 3.14: The Country of Origin Effect in the US (1/2)

cb	Gap 1	SE Gap 1	Gap 2	SE Gap 2
ireland	0.004	0.040	-0.033	0.039
australia	0.196	0.045	0.102	0.042
canada	0.115	0.014	0.026	0.014
new zealand	0.129	0.063	0.045	0.056
kenya	0.039	0.034	-0.081	0.032
uganda	-0.074	0.110	-0.180	0.121
ghana	-0.133	0.031	-0.168	0.030
nigeria	0.016	0.028	-0.157	0.027
sierra leone	-0.044	0.117	-0.114	0.119
barbados	0.012	0.039	-0.028	0.036
jamaica	-0.086	0.014	-0.057	0.013
trinidad and tobago	-0.045	0.022	-0.048	0.021
guyana	-0.155	0.020	-0.120	0.019
bangladesh	-0.161	0.035	-0.292	0.034
india	0.333	0.011	0.116	0.010
sri lanka	0.161	0.083	0.027	0.080
hong kong	0.156	0.027	0.053	0.024
malaysia	0.204	0.050	0.047	0.046
singapore	0.203	0.074	0.085	0.063
morocco	-0.236	0.054	-0.232	0.050
egypt	-0.140	0.039	-0.283	0.036
south africa	0.168	0.044	0.034	0.041
pakistan	0.018	0.025	-0.112	0.023
burma	-0.160	0.044	-0.050	0.046
china	0.012	0.012	-0.076	0.011
japan	0.132	0.017	-0.010	0.016
philippines	0.019	0.009	-0.082	0.009
vietnam	-0.108	0.012	-0.053	0.011
iran	-0.021	0.026	-0.168	0.023
israel	0.004	0.039	-0.078	0.037
belgium	0.122	0.067	-0.040	0.062
denmark	0.154	0.062	0.019	0.056
france	0.064	0.030	-0.056	0.027
italy	-0.125	0.021	-0.097	0.020
netherlands	0.139	0.042	0.014	0.040
germany	-0.042	0.012	-0.094	0.011

Notes: The country of origin effects are estimated from a regression of log hourly wages on country of origin dummies, controlling for gender, race, a cubic in age, region and year fixed effects. The second specification controls also for *years of education*.

Table 3.15: The Country of Origin Effect in the US (2/2)

cb	Gap 1	SE Gap 1	Gap 2	SE Gap 2
bulgaria	-0.092	0.092	-0.186	0.081
czechoslovakia	-0.066	0.050	-0.157	0.044
hungary	-0.027	0.038	-0.113	0.034
poland	-0.155	0.018	-0.137	0.017
romania	-0.017	0.038	-0.102	0.036
austria	0.044	0.054	-0.031	0.053
switzerland	0.103	0.061	-0.060	0.061
greece	-0.143	0.036	-0.094	0.034
portugal	-0.266	0.024	0.000	0.022
spain	-0.012	0.030	-0.089	0.028
finland	0.056	0.089	-0.062	0.091
norway	0.197	0.080	0.092	0.070
sweden	-0.012	0.066	-0.148	0.070
yugoslavia	-0.219	0.026	-0.163	0.024
turkey	-0.114	0.038	-0.182	0.035
ussr	-0.017	0.019	-0.147	0.018
ethiopia	-0.168	0.028	-0.191	0.027
somalia	-0.197	0.056	-0.038	0.056
mexico	-0.536	0.005	-0.153	0.005
argentina	-0.100	0.030	-0.138	0.027
brazil	-0.275	0.022	-0.214	0.022
chile	-0.208	0.034	-0.256	0.029
colombia	-0.314	0.014	-0.273	0.013
venezuela	-0.067	0.029	-0.135	0.027
iraq	-0.332	0.038	-0.300	0.036
lebanon	-0.060	0.033	-0.115	0.031
korea	0.034	0.014	-0.074	0.013
belarus	0.000	0.074	-0.105	0.068
lithuania	-0.138	0.059	-0.235	0.052
latvia	0.022	0.116	-0.112	0.116
ukraine	-0.078	0.027	-0.183	0.026
sudan	-0.071	0.118	-0.018	0.117
indonesia	0.093	0.040	-0.036	0.037
bermuda	0.029	0.074	-0.120	0.063
taiwan	0.251	0.020	0.019	0.019
thailand	-0.127	0.024	-0.128	0.024

Notes: The country of origin effects are estimated from a regression of log hourly wages on country of origin dummies, controlling for gender, race, a cubic in age, region and year fixed effects. The second specification controls also for *years of education*.

3.A.2 Including Mexico in the US specifications

The tables below present the estimated mean wage gap (and corresponding standard error) by country of origin for the UK and the US. The first specification does not include education, whereas the second one does.

Table 3.16: Immigrants' Education, including Mexico

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)	(13)	(14)
Distance	0.000437***					9.65e-05***	0.000115***	2.60e-05					4.52e-05	3.04e-05
	(5.23e-05)					(2.96e-05)	(2.96e-05)	(3.04e-05)					(3.50e-05)	(3.78e-05)
English		3.066***				0.822***	0.880***		-0.647**				-0.503	-0.483
		(0.631)				(0.237)	(0.220)		(0.257)				(0.380)	(0.389)
Percent			-12.36***			-11.73***	-12.39***			-8.331**			-1.809	-2.063
			(0.511)			(0.975)	(0.918)			(3.629)			(4.237)	(4.335)
Fgnedu				-9.345***		-1.089	1.592				2.387***		1.749**	1.511**
				(2.133)		(0.874)	(1.244)				(0.648)		(0.682)	(0.725)
Sex						4.319***	3.405**						-1.378	-2.016
						(1.618)	(1.514)						(1.126)	(1.299)
Age						-0.0326	-0.105*						-0.0383	-0.0451
						(0.0629)	(0.0607)						(0.0387)	(0.0399)
Gdp					4.54e-05*		3.15e-05**					-9.23e-06		-9.16e-06
					(2.59e-05)		(1.26e-05)					(7.13e-06)		(7.80e-06)
UK								Yes	Yes	Yes	Yes	Yes	Yes	Yes
US	Yes	Yes	Yes	Yes	Yes	Yes	Yes							
Constant	15.20***	17.46***	20.59***	24.06***	17.59***	19.10***	20.41***	19.14***	19.64***	19.65***	17.79***	19.45***	20.54***	21.55***
	(0.412)	(0.303)	(0.142)	(1.372)	(0.417)	(3.062)	(2.890)	(0.215)	(0.191)	(0.205)	(0.423)	(0.188)	(1.850)	(2.046)
Observations	71	72	72	72	67	71	67	71	72	72	72	67	71	67
R-squared	0.503	0.252	0.893	0.215	0.045	0.927	0.943	0.010	0.083	0.070	0.162	0.025	0.286	0.310

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the proportion of immigrants from the same country of birth in the sample (Percent); the proportion of immigrants from the country of origin who completed their education in the home country (Fgnedu); the proportion of women in the immigrant community (Sex); the GDP per capita of the country of origin in 2009 (or the latest date available).

Table 3.17: The Country of Origin Effect, including Mexico

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Distance	5.80e-06				-5.82e-07	2.97e-06	9.25e-06***				4.20e-06*	5.99e-06**
	(4.43e-06)				(5.17e-06)	(5.80e-06)	(2.11e-06)				(2.47e-06)	(2.44e-06)
Size		1.03e-05			-1.90e-06	-4.67e-06		-1.50e-06***			-4.31e-07	1.29e-06***
		(7.89e-06)			(9.43e-06)	(1.08e-05)		(3.44e-07)			(4.19e-07)	(4.73e-07)
English			0.124***		0.130**	0.154**			0.115***		0.0878***	0.153***
			(0.0363)		(0.0503)	(0.0639)			(0.0189)		(0.0203)	(0.0203)
Test				0.0656		0.0991**				0.0759***		0.113***
				(0.0393)		(0.0371)				(0.0215)		(0.0175)
UK	Yes	Yes	Yes	Yes	Yes	Yes						
US							Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.235***	-0.235***	-0.272***	-0.479**	-0.267***	-0.712***	-0.168***	-0.0692***	-0.133***	-0.431***	-0.144***	-0.697***
	(0.0312)	(0.0307)	(0.0270)	(0.181)	(0.0406)	(0.184)	(0.0166)	(0.0123)	(0.00906)	(0.0925)	(0.0262)	(0.0915)
Observations	71	72	72	50	71	50	71	72	72	50	71	50
R-squared	0.024	0.024	0.143	0.055	0.143	0.278	0.218	0.212	0.345	0.206	0.425	0.738

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$
The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the sample size of immigrants from the same country of birth (Size); the mean test score in international standardized tests in the country of origin (Test).

Table 3.18: The Returns to Education, including Mexico

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
English	0.00493***				0.00452*	0.00578*	0.00426***				0.00343***	0.00532***
	(0.00183)				(0.00249)	(0.00288)	(0.00116)				(0.00122)	(0.00138)
Test		0.00445**				0.00570***		0.00392***				0.00348***
		(0.00175)				(0.00168)		(0.00105)				(0.00119)
Distance			1.35e-09		-1.70e-07	3.84e-08			7.97e-08		-3.83e-07**	-3.18e-07*
			(2.20e-07)		(2.56e-07)	(2.62e-07)			(1.28e-07)		(1.48e-07)	(1.66e-07)
Size				9.33e-07**	4.64e-07	2.79e-07				-6.34e-08***	-8.28e-08***	-4.71e-08
				(3.76e-07)	(4.68e-07)	(4.87e-07)				(1.96e-08)	(2.52e-08)	(3.22e-08)
UK	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No
US	No	No	No	No	No	No	Yes	Yes	Yes	Yes	Yes	Yes
Constant	-0.0118***	-0.0276***	-0.00905***	-0.0119***	-0.0120***	-0.0375***	-0.00902***	-0.0244***	-0.00852***	-0.00644***	-0.00410**	-0.0202***
	(0.00136)	(0.00809)	(0.00155)	(0.00146)	(0.00201)	(0.00832)	(0.000555)	(0.00450)	(0.00101)	(0.000701)	(0.00158)	(0.00622)
Observations	71	50	70	71	70	50	72	50	71	72	71	50
R-squared	0.095	0.118	0.000	0.082	0.127	0.310	0.162	0.226	0.006	0.130	0.282	0.502

Notes: Standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

The variables are the distance between the host and the home country (Distance); a dummy if English is an official language of the country of origin; the sample size of immigrants from the same country of birth (Size); the mean test score in international standardized tests in the country of origin (Test).

Bibliography

Abbott, M. and Beach, C. (1994), 'Wage changes and job changes of canadian women', *Journal of Human Resources* **29**(2), 429–460.

Abowd, J. M., Kramarz, F. and Margolis, D. N. (1999), 'High wage workers and high wage firms', *Econometrica* **67**(2), 251–333.

Allison, P. D. (1982), 'Discrete-time methods for the analysis of event histories', *Sociological Methodology* **13**, 61–98.

Andersson, F., Garca-Prez, M., Haltiwanger, J., McCue, K. and Sanders, S. (2010), 'Workplace concentration of immigrants', *NBER Working Paper* (16544).

Aydemir, A. and Skuterud, M. (2008), 'The immigrant wage differential within and across establishments', *Industrial and Labor Relations Review* .

Bartel, A. P. (1989), 'Where do the new us immigrants live?', *Journal of Labor Economics* **7**(4), 371–391.

Bartel, A. P. and Borjas, G. (1981), Wage growth and job turnover: An empirical analysis, *in* S. Rosen, ed., 'Studies in Labor Markets', Chicago, University of Chicago Press, pp. 65–84.

Bartelsman, E. J. and Doms, M. (2000), 'Understanding productivity: Lessons from longitudinal microdata', *Journal of Economic Literature* **38**(3), 569–594.

Bauer, P. and Riphan, R. (2007), 'Heterogeneity in the intergenerational transmission of educational attainment: evidence from switzerland on natives and second generation immigrants', *Journal of Population Economics* **20**(1).

Black, S., Devereux, P. and Salvanes, K. (2005), 'the more the merrier? the effect of family size and birth order on children's education', *The Quarterly Journal of Economics* **May**.

Blau, F., Kahn, L., Liu, A.-H. and Papps, K. (2008), 'The transmission of women's fertility, human capital and work orientation across immigrant generations', *IZA Discussion Paper* (3732).

Borjas, G. (1989), 'Immigrant and emigrant earnings: A longitudinal study', *Economic Inquiry* **27**(1), 21–37.

Borjas, G. J. (1987), 'Self-selection and the earnings of immigrants', *American Economic Review* **77**(4), 531–553.

Borjas, G. J. (1991), 'Immigration policy, natioanl origin and immigrant skills: A comparison of canada and the united states', *NBER Working Paper* (3691).

Borjas, G. J. (2000), The economic progress of immigrants, *in* 'Issues in the Economics of Immigration', University of Chicago Press.

Bratsberg, B. and Ragan Jr., J. F. (2002), 'The impact of host-country schooling on earnings, a study of male immigrants in the united states', *Journal of Human Resources* **37**(1), 63–105.

Bratsberg, B. and Terrell, D. (2002), 'School quality and returns to education of us immigrants', *Economic Inquiry* **40**(2), 177–198.

Burdett, K. (1978), 'A theory of employee job search and quit rates', *American Economic Review* pp. 212–20.

Card, D., Dinardo, J. and Estes, E. (2000), 'The more things change: Immigrants and the children of immigrants in the 1940s, the 1970s, and the 1990s', *NBER Working Paper* (6519).

Cardoso, A. R. (2005), 'Big fish in small pond or small fish in big pond? an analysis of job mobility', *IZA Discussion Paper* (1900).

Carneiro, A., Fortuna, N. and Varejo, J. (2006), 'Immigrants' earnings and workplace characteristics', *Working Paper* .

Carneiro, A., Fortuna, N. and Varejo, J. (2009), 'Immigrants at new destinations: How they fare and why?', *IZA Discussion Paper* (4892).

Carrington, W. J. and Lima, P. (1996), 'The impact of 1970's repatriates from africa on the portuguese labor market', *Industrial Labor and Relations Review* **49**(2), 330–347.

Carrington, W. J. and Troske, K. R. (1997), 'On measuring segregation in samples with small units', *Journal of Business and Economic Statistics* **15**(4), 402–409.

Chiswick, B. R. (1978), 'The effect of americanization on the earnings of foreign-born men', *Journal of Political Economy* **86**(5).

Chiswick, B. R., Lee, Y. L. and Miller, P. W. (2005), 'A longitudinal analysis of immigrant occupational mobility: A test of the immigrant assimilation', *International Migration Review* **39**(2), 332–353.

Chiswick, B. R. and Miller, P. W. (1995), 'The endogeneity between language and earnings: International analyses', *Journal of Labor Economics* **13**(2), 246–88.

Corak, M. (2006), 'Do poor children become poor adults? lessons from a cross country comparison of generational earnings mobility', *IZA Discussion Paper* (1993).

Duleep, H. O. and Dowhan, D. J. (2002), 'Insights from longitudinal data on the earnings growth of us foreign-born men', *Demography* **39**(3), 485–506.

Dustmann, C., Fabbri, F. and Wadsworth, J. (2003), 'Labour market performance of immigrants in the uk labour market', *Home Office Online Report* **05/03**.

Dustmann, C., Glitz, A. and Schnberg, U. (2011), 'Job search networks and ethnic segregation in the workplace', *IZA Discussion Paper* (5777).

Eckstein, Z. and Weiss, Y. (2004), 'On the wage growth of immigrants: Israel, 1990-2000', *Journal of the European Economic Association* **2**(4), 665–695.

Effron, B. (1988), 'Logistic regression, survival analysis, and the kaplan-meier curve', *Journal of the American Statistical Association* **83**(402), 414–425.

Farber, H. S. (1999), Mobility and stability: The dynamics of job change in labor market, *in* 'Handbook of Labor Economics', Vol. 3, Elsevier B. V., pp. 2439–2483.

Farber, H. S. and Gibbons, R. (1996), 'Learning and wage dynamics', *The Quarterly Journal of Economics* **111**(4), 1007–1047.

Fernandez, R. and Fogli, A. (2009), 'Culture: An empirical investigation of beliefs, work, and fertility', *American Economic Journal* **1**(1).

Firpo, S., Fortin, N. M. and Lemieux, T. (2009), 'Uncondtional quantile regressions', *Econometrica* **77**(3), 953–973.

Friedberg, R. M. (1992), 'The labor market assimilation of immigrants in the united states: The role of age at arrival', *Brown University, Working Paper* .

Friedberg, R. M. (2000), 'You cant take it with you? immigrant assimilation and the portability of human capital.', *Journal of Labor Economics* **18**(2), 221–51.

Furtado, D. and Theodoropoulos, N. (2009), 'Intermarriage and immigrant employment: The role of networks', *Cream Discussion Paper* (06/09).

Gang, I. and Zimmermann, K. (2000), 'Is child like parent? educational attainment and ethnic origin', *The Journal of Human Ressources* **35**(3).

Ganzeboom, H., De Graf, P. and Treiman, D. (1992), 'A standard international socio-economic index of occupational status', *Social Science Research* (21).

Gibbons, R., Katz, L. F., Lemieux, T. and Parent, D. (2005), 'Comparative advantage, learning and sectoral wage determination', *Journal of Labor Economics* **23**(4).

Green, D. A. (1999), 'Immigrant occupational attainment: Assimilation and mobility over time', *Journal of Labor Economics* **17**(1), 49–79.

Groes, F., Kircher, P. and Manovskii, I. (2010), 'The u-shapes of occupational mobility', *Working Paper* .

Grogger, J. and Hanson, G. H. (2011), 'Income maximization and the selection and sorting of international migrants', *Journal of Development Economics* **95**, 42–57.

Guimaraes, P. and Portugal, P. (2009), 'A simple feasible alternative procedure to estimate models with high-dimensional fixed effects', *IZA Discussion Paper* (3935).

Hammarstedt, M. and Palme, M. (2006), 'Intergenerational mobility, human capital transmission and the earnings of second-generation immigrants in sweden', *IZA Discussion Paper* (1943).

Hanushek, E. A. and Kimko, D. D. (2000), 'Schooling, labor force quality and the growth of nations', *The American Economic Review* **90**(5), 1184–1208.

Hanushek, E. A. and Woessmann, L. (2009), 'Do better schools lead to more growth? cognitive skills, economic outcomes, and causation', *NBER Working Paper* **14633**.

Hartog, J. (1981), 'Wages and allocation under imperfect information', *De Economist* **129**(3).

Heath, A., Rothon, C. and Kilpi, E. (2008), 'The labour market assimilation of immigrants in the united states: The role of age at arrival', *Annual Review of Sociology* (34).

Hu, W.-Y. (2000), 'Immigrant earnings assimilation: Estimates from longitudinal data', *The American Economic Review* **90**(2), 368–372.

Jasso, G. and Rosenzweig, M. R. (1986), 'What's in a name? country-of-origin influences on the earnings of immigrants in the united states', *Research in Human Capital and Development* (4), 75–106.

Keith, K. and McWilliams, A. (1999), 'The returns to job mobility and job search by gender', *Industrial and Labor Relations Review* **52**(3), 460–477.

Korn, E. L. and Graubard, B. I. (1995), 'Examples of differing weighted and unweighted', *The American Statistician* **49**(3).

LaLonde, R. J. and Topel, R. H. (1997), Economic impact of international migration and the economic performance of migrants, *in* 'Handbook of Population and Family Economics', Vol. 1, North Holland, pp. 799–850.

Lange, F. (2007), 'The speed of employer learning', *Journal of Labor Economics* **25**(1).

Lubotsky, D. (2007), 'Chutes or ladders? a longitudinal analysis of immigrant earnings', *Journal of Political Economy* **115**(15).

Manning, A. (2000), 'Movin' on up: Interpreting the earnings-experience profile', *Bulletin of Economic Research* **52**(4), 261–296.

Matoo, A., Neagu, I. C. and zden, a. (2008), 'Brain waste? educated immigrants in the us labor market', *Journal of Development Economics* **87**, 255–269.

Mayer, T. and Zignago, S. (2011), 'Notes on cepii's distances measures: The geodist database', *CEPII Working Paper* **2011**(25).

Mincer, J. (1986), Wage changes and job changes, *in* R. G. Ehrenberg, ed., 'Research in Labor Economics: A Research Annual', pp. 171–197.

Neal, D. and Rosen, S. (2000), 'Theories of the distribution of earnings', *The Handbook of Income Distribution* **1**, 379–427.

OECD (2005), 'Trends in international migration'.

OECD (2007), 'Jobs for immigrants - vol. 1: Labour market integration in australia, denmark, germany and sweden'.

OECD (2008*a*), 'International migration outlook'.

OECD (2008*b*), 'Jobs for immigrants - vol. 2: Labour market integration in belgium, france, the netherlands and portugal'.

OECD (2012), 'International migration outlook'.

Pendakur, K. and Woodcock, S. (2009), 'Glass ceilings or glass doors? wage disparities within and between firms', *IZA Discussion Paper* (4626).

Riphahn, R. (2003), 'Cohort effects in the educational attainment of second generation immigrants in germany: An analysis of census data', *Journal of Population Economics* **16**(5).

Rothschild, M. and Stiglitz, J. E. (1982), 'A model of employment outcomes illustrating the effect of the structure of information on the level and distribution of income', *Economics Letters* **10**, 231–236.

Sattinger, M. (1993), 'Assignment models of the distribution of earnings', *Journal of Economic Literature* **31**, 831–880.

Solon, G. (1999), 'Intergenerational mobility in the labour market', *Handbook of Labor Economics* **29**.

Sweetman, A. (2004), 'Immigrant source country educational quality and canadian labour market outcomes', *Research Paper, Statistics Canada* .

Topel, R. H. and Ward, M. P. (1992), 'Job mobility and the careers of young men', *The Quarterly Journal of Economics* .

Van Ours, J. and Veenman, J. (2003), 'The educational attainment of second-generation immigrants in the netherlands', *Journal of Population Economics* **16**(4).

Woodcock, S. D. (2008), 'Wage differentials in the presence of unobserved worker, firm and match heterogeneity', *Labour Economics* (15), 772–794.