



UNIVERSITY OF GOTHENBURG
SCHOOL OF BUSINESS, ECONOMICS AND LAW

Immigration, education and employment

Does education affect assimilation of migrants in the labor market?

Carl Tosteby & Malin Apler

Abstract:

The aim of this paper is to analyze how the level of education affects the probability of being employed for immigrants in the United States (U.S.) during the period 2000-2017. To be able to estimate the effect of education, we use census records from IPUMS USA that allow us to identify migrants, their employment status and level of education, together with other demographics. The analysis is made through regressions in STATA, using a Linear Probability Model. The results imply that education has a positive and significant effect on the probability of being employed. However, other factors such as the race or the birthplace of the immigrants, affect the likelihood of being employed to a greater extent than education. Additionally, we document the heterogeneous effects of education according to race and birthplace. That means, we find significant differences in the magnitude of the education effect *within* the group of immigrants. Finally, we present the returns of education for migrants. Our results are in line with previous research within the field of immigration economics.

Key words: Immigration Economics, Assimilation, Education

Bachelors thesis (15hp)
Department of Economics
School of Business, Economics and Law
University of Gothenburg
Supervisor: Melissa Rubio Ramos

Acknowledgement

We would like to give thanks and gratitude to our supervisor Melissa Rubio Ramos for all of her good advice, input, and encouragement. We could not have done this paper without her. Thank you!

Carl Tosteby

Malin Apler

Table of contents

| | |
|--|-----------|
| 1. Introduction | 3 |
| 2. Conceptual framework | 5 |
| 2.1 Heterogeneous effects | 6 |
| 2.2 Positive or negative selection..... | 6 |
| 2.3 Labor migration | 7 |
| 2.4 Combination of factors | 7 |
| 3. Literature review..... | 8 |
| 4. Data | 10 |
| 4.1 Descriptive statistics | 13 |
| 4.1.1 Employment | 16 |
| 4.1.2 Education | 17 |
| 4.1.3 Education and birthplace..... | 18 |
| 4.1.4 Education and race | 19 |
| 5. Empirical strategy..... | 20 |
| 6. Results | 21 |
| 6.1 Main results..... | 21 |
| 6.2 Heterogeneous effects | 25 |
| 6.3 The effect of education on wages | 28 |
| 6.4 Robustness check | 31 |
| 7. Discussion..... | 32 |
| 7.1 Strengths and limitations..... | 34 |
| 8. Conclusion..... | 36 |
| 9. References | 37 |
| 10. Appendices | 38 |
| 10.1 Graphs | 38 |
| 10.2 Tables | 41 |
| 10.3 STATA Commands | 47 |

1. Introduction

The introduction motivates the research question and purpose of this paper. This is followed by an explanation of why the paper is relevant. Last, we present the empirical method briefly and summarize the main results.

In 2017, 258 million people lived in a country they were not born in, so-called foreign-born (OECD, 2018). During the same year, 5 million people migrated to OECD countries, due to different reasons such as humanitarian migration, family migration, and labor migration (OECD, 2018). The United States (hereafter the U.S.) receives the largest number of immigrants per year compared to other OECD countries, 1.18 million. The second largest receiver, Germany, received about 1 million immigrants in 2016 and the third receiver, Great Britain, around 350,000 (OECD, 2018). However, migrants face difficulties in the arrival countries in terms of assimilation to a new culture, language, and especially to the labor market. In the U.S., some migrants face larger difficulties of assimilation to the labor market than others. For instance, the unemployment rate for migrants from South America is 5.2 %, whereas the unemployment rate for Europeans is 3.4 %. That is comparable with the unemployment rate for natives, 4.0 %. Given the dimension of the problem, this paper asks which factors contribute to the assimilation of migrants in the labor market. In particular, we look at the role of *education* in explaining the probability of finding a job, and its effect on labor income.

The definition of assimilation is broad and it has been used in different contexts. Fouka et al. (2018) and Abramitzky et.al (2018) focus on cultural assimilation, which has been defined as cultural practices such as food, accent, dress and names that help migrants to integrate in the arrival country. However, in this paper, we are interested in assimilation in economic terms. For this reason, we follow Borjas (2014) to define assimilation as how migrants perform in the host country in terms of the labor market. Thus, our measure for assimilation is the probability to find a job. Therefore, we focus on the employment status of immigrants in the U.S.

Having established our measure of assimilation, we ask the following question: *Does the level of education affect the probability of being employed for immigrants in the*

United States? In short, the paper aims to analyze if immigrants with a higher level of education find it easier to find employment compared to immigrants with a lower level of education, and if so, does the effect of education differ among subgroups of immigrants?

Analyzing these questions is relevant due to several reasons: First, the share of foreign-born in the U.S., i.e. people living in the U.S. who were born in another country, has increased rapidly the last decades. During 2016 the foreign-born rate reached an all-time high, 13 % (OECD, 2018). The unemployment rate for foreign-born is lower in the U.S. than the OECD average, 4 % (OECD, 2018). However, the performance on the labor market for foreign-born in the U.S. differs depending on the country of origin of the immigrant. For instance, the *employment* rate for migrants from Mexico was 65.7 % in the year of 2017, while the *employment* rate for migrants from Europe the same year was 70.7 % (OECD, 2018). A difference by 5 percentage points. Given the importance of employment to guarantee assimilation, it is relevant to ask what factors affect the employment status of immigrants and cause the differences in labor market performance.

Second, our study speaks to the current debate on migration and assimilation, particularly among political leaders. The U.S. President, Donald. J. Trump, has made immigration to a priority of his domestic agenda since he became president 2017. He wants to build walls towards Mexico, ban immigrants from Muslim countries and implement strict migration policies. Policies that have been followed by protests in the U.S. and around the world. During the spring of 2019, President Trump came up with a new immigration policy plan for the U.S. that aims to increase the educational and skills requirements for people who are migrating to the U.S. (Shear, 2019, May 15). The implementation of this, recently unveiled, immigration policy might change the composition of immigrants in the U.S. Presumably, it would increase the share of high skilled, well-educated immigrants. Hence, the relevance and timing of our study, that aims to analyze whether education has any impact on immigrants' performance on the U.S. labor market, could not be more precise.

For the purpose of this paper, we use the latest available data to test our research question. In particular, we are using census records from the U.S. for the period 2000 – 2017 that allow us to identify immigrants by place of birth and the employment status, together with other demographics. The research question is tested using an econometric approach. Briefly, we formulate an equation that describes the relationship between the probability of being employed and the level of education. Demographic controls are included as well.

The results of this study imply that education has a positive effect on the probability of being employed for immigrants on the U.S. labor market. However, the effect of one additional year of education is relatively small, between 0.15 and 1.01 percentage points. The magnitude of the effect varies depending on the immigrant's country of origin and race. To be able to compare our research with existing research we have extended our study to analyze the effect that education has on the wage of an immigrant.

The remainder of the paper is organized as follows. Next section presents the conceptual framework where we discuss different theories that help us to understand the relationship between education and labor outcomes. Section 3 presents the literature review to relate our question to previous studies. Section 4 describes the data and the construction of the variables. Section 5 investigates the effect of education on labor outcomes by presenting the empirical strategy. Section 6 shows the results, followed by a discussion and conclusion.

2. Conceptual framework

This section of the paper aims to summarize how we can think of the relationship between education and the probability of being employed. We have listed several possibilities which are based on recent studies on immigration economics in order to apply a conceptual framework to our research question. These theories will then be tested empirically.

There are different theories of how the relationship between educational background and probability of being employed looks like.

2.1 Heterogeneous effects

One is that returns to education, in literature often referred to as returns to skills (Borjas, 2014), are different for different groups of individuals. That means that the returns, in terms of income or probability of being employed, may differ within the group of immigrants due to the country of origin. For instance, Chiswick (1978) suggests that the returns of schooling on earnings are different for immigrants from English speaking countries compared to none speaking English countries: one year of schooling raises earnings with 6.6 % for immigrants from English-speaking countries but only 5.2 % for other immigrants. Applying Chiswick's findings to our study (assuming earnings and employment are correlated), the effect of education on probability on being employed in the U.S. may differ depending on the country of origin or whether the country of origin is an English-speaking country or not.

2.2 Positive or negative selection

Another theory is that the relationship between education and the probability of being employed depends on the selection of individuals who choose to migrate. This theory refers to the Roy Model of occupational choice which examines people's choice of occupation (Roy, 1951). More recently, the theory has been applied in immigration economics by Borjas (2014). He states there are two types of selections; positive and negative selection. Positive selection occurs when returns to skills (education or experience) are relatively lower in the sending country compared to the receiving country. That will result in a selection of highly skilled people who choose to migrate. Presumably, these individuals have a high educational level, are strongly motivated and will not struggle to find a job in the arrival country. In this case, most of the immigrants in the arrival country will both have a high education level and will probably be employed. However, there might be a case when immigrants are negatively selected. That occurs when the returns to skills are relatively higher in the sending country, resulting in the highly skilled individuals to stay and the low skilled to migrate to the host country (Borjas, 2014).

2.3 Labor migration

One third possibility is that the majority of the immigrants are labor immigrants, i.e. they are migrating to the U.S. because of changes due to employment conditions. For example, this may occur when factories are relocated to the U.S. and employees in the foreign country are forced to follow to keep their jobs. Another example when labor migration may occur is when governments establish guest worker programs, in order to attract foreign workers. For Instance, the U.S. implemented the Bracero Program in 1942, which was a bilateral agreement with Mexico that aimed to supply the American farm industry with workers (Schmitter Heslier, B. 2008). Between 1942 and 1947, 219 546 Mexicans were recruited as Braceros and migrated to U.S. for farm work. If a large share of the immigrants are labor immigrants, the relationship between the level of education and the probability of being employed is difficult to interpret since the migrants is already employed prior to the migration.

2.4 Combination of factors

It is also possible that there is a combination of all these theories. There may be a positively selection as well as different returns to education in different countries, and there might be a possibility that racial discrimination has a greater impact on probability on being employed than the education effect.

Last, it is possible that education does not affect the probability of being employed. That would mean that the probability of being employed in the arrival country does not depend on the individual's level of education. Instead, it would depend on other factors, such as race, gender or age. This theory is opposite the majority of the comprehensive studies we have found in the context of immigration and returns to education, but we consider it for our empirical analysis.

Clearly, the relationship between education and the probability of being employed is not obvious when looking at theories and recent studies applied in this area. For that reason, we will study this question empirically.

3. Literature review

This section presents the literature review in order to relate our question to previous studies.

Research in this area focuses mainly on two different issues: (1) How the recipient country's economy is affected by immigrants, and (2) how the immigrants perform or behave in the country they migrate to. In short, one could say there are two areas of research. One area that focuses mainly on the recipient country and the consequences of immigration on the economy. Another area that rather focuses on immigrants, i.e. the individuals and their performance in the arrival country and their assimilation into society.

Most of the research that has been made focus on the first (1) issue: How the host economy, particularly the labor market and the wages of the native workers, are affected by the migration¹. Brückner and Jahn (2011) study the labor market effects due to migration in Germany and conclude that the effect of immigration on the labor market is moderate: 1 percent increase in immigration increased unemployment rate for natives with 0.1 percent and decreased the wages for natives with 0.1 percent. These patterns are consistent with findings in Dustman et al. (2016) and Borjas (2003). On the other hand, Foged and Peri (2015) found positive effects of immigration on native wages in Denmark, particular native unskilled wages.

Even if most of the research analyses the host country economy and the effects on the labor market, there are some exemptions. These exemptions focus on the assimilation of the immigrants in the host country. For instance, Bevelander and Veenman (2006) conclude that the probability of being naturalized, i.e. become a citizen of the Netherlands, is higher for immigrants with a higher educational background. In addition, the study shows that for the immigrants with an education that are certified in the Netherlands, i.e. not from the immigrant's country of origin, the probability was even higher to become a citizen. Becoming a citizen, in turn, increased the probability of getting a job and being employed in the Netherlands, according to the paper.

¹ For instance, see: East et al. (2008), Sameeksha et al. (2016) and Borjas et al. (1991)

Fouka, Mazumder, and Tabellini (2018) write in their paper that racial factors may affect adaptation and assimilation to the new society. The paper studies how the immigration of Blacks to the northern states in the U.S. during the great migration in the 1910s affected the *cultural* integration of groups that had previously been excluded, particularly European immigrants. They conclude that there was significant heterogeneity, i.e. when black immigrants arrived in the northern states, the effort of assimilation increased for all European immigrants, but the assimilation success was larger for immigrants from western and northern Europe than from southern and eastern Europe. The authors explain these findings by cultural and racial barriers, where the northern and western European immigrants faced lowered racial barriers compared to eastern and southern Europeans. However, they do not analyze how education affected assimilation.

Chiswick (1978) examines the effect of Americanization on earnings for foreign-born white men using the 1970 census of the U.S. population from IPUMS. His findings are related to our paper. For instance, differences in the effect of education were found, both for immigrants versus natives and *within* the population of immigrants. However, Chiswick's study differs compared to ours. In particular, he analyzed the education effect on wages only for *Whites and males* that were foreign-born and he used another period of studying than us.

Duncan and Trejo (2011) study the assimilation of low-skilled immigrants in the U.S. labor market. They conclude that low-skilled immigrants (migrants with low levels of schooling and English proficiency) do not face problems with finding paid employment. However, the earnings for these immigrants are commensurate with their skills. Applying their findings to this paper, it would mean that the level of education does not affect the probability of getting employed, but has effect on wages.

Aldén and Hammarstedt (2014) analyze the integration of immigrants in the Swedish labor market and find that immigrants from some specific areas (Africa and Asia) are more likely to be unemployed than immigrants from other areas (Europe). The paper also compares the unemployment rates between immigrants with different levels of education and present that education pays off for immigrants, i.e. the employment rate is higher for those with higher education, even though it has not the same effect on the employment rate as for natives.

The most cited research in this area is probably Borjas (2014). Borjas has written a large number of papers and books about labor economics and immigration economics. *Immigration Economics* (Borjas, 2014) summarizes his and other theories in immigration economics. Chapter 1 and 2 in this book are highly connected to this paper where he describes different theories of immigration selection (chapter 1) and economic assimilation (chapter 2). Most interesting is probably the section where Borjas presents a model of economic assimilation, which is the most closely related model to our study. Like other theories in this issue, Borjas uses wage as a measure of assimilation in contrast to our study where we use employment status. His findings about the determinants of economic assimilation are worth to mention: (1) the model implies that each additional year of education increases the rate of wage growth by 1.5 percentage point. (2) Borjas states there is a positive correlation between immigrant earnings and the sending countries GDP, which may be explained by the fact that it is easier to transfer “skills” such as education certificates between high-GDP countries.

All these papers analyze issues that are strongly connected to our research question. However, our question has not been addressed in the literature as of today. At least, not in the same context and the same period. Unlike Aldén and Hammarstedt (2014) we are studying the effect of education on employment at an individual level. Fouka et.al (2018) is closely related to ours. However, whilst they study cultural assimilation and factors that might affect that, we focus particularly on assimilation through employment and how education affects that process. In short, our research question is not that broad as recent research and puts the issue in another context. Therefore, our contribution to recent studies is to give a concrete picture of the effect of education on economic assimilation for immigrants, particularly the effect on the probability of being employed.

4. Data

In this section we provide an overview of the data used in this study, explaining how we have collected it and how variables were constructed.

The study is limited to the U.S. The limited time interval and budget limitations do not make it possible to extend the study to include European countries, for example

Sweden, as this requires too large costs to be able to provide the data at an individual level that the study requires.

IPUMS

We have used a dataset on individual level provided by The Integrated Public Use Microdata Series (IPUMS)². IPUMS collects data in the U.S. on individual and household level through yearly censuses, called the American Community Survey (hereafter called ACS). ACS is mailed to a random sample of individuals that are selected each month to represent each county in the U.S. The ACS contains questions about the individuals' characteristics, such as age, income, occupation, employment status, birthplace, race, etc. We decided to use this source of data for two reasons: (1) the database is reliable and is being used by other researchers to study the same kind of issues as we study, for instance, Borjas (2014) and Chiswick (1978), and (2) the dataset is useful since it provides us with the variables we need to test our hypothesis.

Size of dataset

The dataset contains more than 44 million observations (individuals) during 2000 to 2017. The sizes of the yearly samples are relatively big. The censuses from 2000, 2001, 2002, 2003 and 2004 contain 0.13 %, 0.43 %, 0.38 %, 0.42 % and 0.42 % of the population, respectively. From 2005 and forward the sample sizes are 1 % of the population in U.S. That is, each sample case (individual) in a sample represents 100 people in the population.

Variables requested from IPUMS

For this paper, the dependent variable is employment status and the variable of interest is years of education. These variables have been the starting point for the data we requested from IPUMS. In addition to these variables, other variables have been requested, for which we expect may have an impact on the probability of being employed, such as age, gender and language skills. Section 10.2, Table 1, in the appendix shows a brief summary of all variables requested from IPUMS and an explanation of each variable.

² <https://usa.ipums.org/usa/index.shtml>

Selection of data from the Dataset

The study focuses on immigrants in the U.S., their employment status and level of education. Thus, we had to filter the data and rule out individuals that were not of interest for the study. We decided that individuals of interest require two characteristics: (1) foreign-born and (2) within the labor force. Using STATA, we dropped individuals that are born in the U.S. and individuals that are not within the labor force. This resulted in a sample with 3 323 200 individuals that are (1) foreign-born and (2) within the labor force. See Table 4.1 for description of the selection process in STATA.

Table 4.1. Selection of data.

| | |
|--|----------------------|
| Number of observations retrieved from IPUMS | |
| 44 947 578 | |
| | |
| <i>Selection 1</i> | |
| Purpose | Code in Stata |
| Rule out individuals born in U.S. | drop if yimmig==0 |
| Number of observations filtered out | |
| 39 443 813 | |
| Observations left | |
| 5 503 765 (12%) | |
| | |
| <i>Selection 2</i> | |
| Purpose | Code in Stata |
| Rule out individuals that are not within the labor force | keep if labforce==2 |
| Number of observations filtered out | |
| 2 180 565 | |
| Observations left | |
| 3 323 200 (5 %) | |

Creating variables for regression

Most of the variables from IPUMS are coded in a way that makes them difficult to interpret when running a regression. Thus, to make the data and variables easier to manage and interpret, we have recoded and created new variables in STATA (see section 10.3 in the appendix for description of commands and recoding). These

variables are the ones we used in the regressions (see section 10.2, Table 2, in the appendix).

4.1 Descriptive statistics

In this section we present tables and graphs that explain the data used for this study. Note that there are 3 323 200 individuals within the sample, restricted to individuals that are foreign-born and within the labor force.

The variables of interest in the study are employment status and education. Thus, we have chosen to show graphs including only these variables in this section. There are more graphs and tables in the appendix. Table 4.2 (p.14) presents descriptive statistics of the dataset.

Panel A includes descriptive statistics of Age, Gender, Employment status, Years of education and Years stayed in the U.S. Age, Years of education and Years stayed in U.S. are continuous variables. Employment status and Male are binary variables. The variable of Employment status takes value 1 if individual is employed, 0 otherwise. The mean of the variable is 0.93, which means 93 % of the individuals in the dataset are employed. The variable of Male takes value 1 if individual is male, 0 if female. The mean of male is 0.553, which means that 55.3 % of the individuals in the dataset are males. The average age of the individuals in the dataset is 42.38 years, according to the mean of the age variable.

The number of years the individuals have stayed in U.S. since immigration is on average 20.79 years. However, the standard deviation is 13.5 years, indicating the variation in the dataset is high. The average years of education among the immigrants in our dataset is 12.39. That is comparable with the average year of education for natives in our data from IPUMS which is 13.34 years of education³.

³ To calculate the average years of education for natives, we used the data from IPUMS but included only individuals born in the U.S. and within the labor force (18 889 409 individuals) when running the “sum” command in STATA.

Table 4.2. Descriptive statistics: Composition of the migrant sample.

| Panel A | Observations | Mean | sd | min | max |
|--------------------------------|--------------|-------|-------|-----|-----|
| Age | 3,323,200 | 42.38 | 12.83 | 16 | 95 |
| Employment status (Employed=1) | 3,323,200 | 0.930 | 0.255 | 0 | 1 |
| Male | 3,323,200 | 0.552 | 0.497 | 0 | 1 |
| Years of Education | 3,323,200 | 12.39 | 3.757 | 0 | 18 |
| Years stayed in U.S. | 3,323,200 | 20.79 | 13.50 | 0 | 94 |

| Panel B: Birthplace | Observations (%) |
|----------------------------|---------------------|
| Europe | 489,237 (14.72) |
| Central America | 1,313,00 (39.50) |
| South America | 220,16 (6.625) |
| North America | 186,52 (5.613) |
| Asia | 966,351 (29.08) |
| Africa | 127,319 (3.831) |
| Oceania | 19,255 (0.579) |
| Other | 1,796 (0.0540) |
| Total | 3,323,200 |

| Panel C: English skills | Observations (%) |
|--------------------------------|----------------------|
| Does not speak English | 205,065 (6.17) |
| Speaks English, but not well | 524,27 (15.78) |
| Speaks English well | 709,175 (21.34) |
| Speaks very well | 1,205,660 (36.28) |
| Speaks only English | 679,030 (20.36) |
| Total | 3,323,200 |

| Panel D: Race | Observations (%) |
|------------------------------|----------------------|
| White | 1,592,468 (47.93) |
| Black | 269,000 (8.09) |
| Chinese | 205,676 (6.19) |
| Japanese | 24,464 (0.74) |
| Other Asian or Pacific | 628,499 (18.91) |
| Other Race or Combined Races | 602,913 (18.14) |
| Total | 3,323,200 |

Panel B, C, and D in table 4.2 present frequencies of individuals in each category. Panel B includes the variable birthplace, assigning all immigrants in subcategories depending on where they were born. Note that North America includes Canada, Puerto Rico, Guam, American Samoa, St. Pierre and Miquelon, U.S. Virgin Islands and Atlantic Islands. Central America includes Mexico, Central American countries, the Caribbean (including Cuba) and the West Indies. Europe includes all European countries and the Russian Empire. Oceania includes Australia, New Zealand and Pacific islands. South America includes countries from South America only. Africa includes African countries only. “Other” includes Antarctica, unknown or born at sea.

The largest share of immigrants in our dataset are born in Central America (39.5 %) followed by Asia (29.08 %), Europe (14.72 %), South America (6.625 %), North America (5.613 %), Africa (3.831 %) and Oceania (0.579 %). 0.054 % are born at sea, in Antarctica or do not know where they were born.

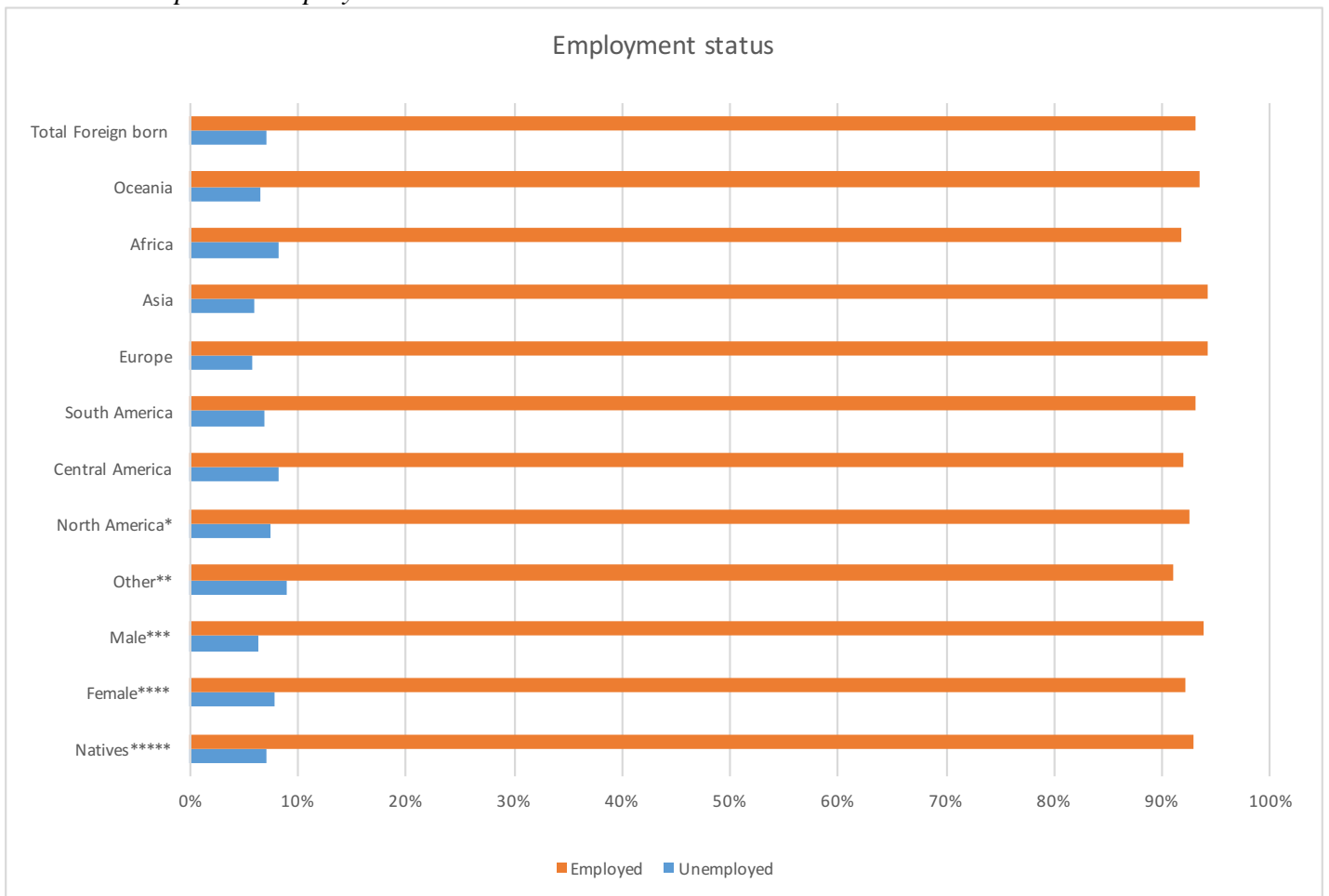
Panel C determines the English skills of the immigrants and divides the individuals into subcategories depending on their level of English. Note that “Speaks only English” corresponds to individuals that speak only English at home or has English as mother language. Most of the individuals in our dataset speak English well, very well or are fluent: 21.34 %, 36.08 % and 20.36 %, respectively.

Panel D determines the races of individuals. The individuals are assigned to a specific race group depending on their response in the ACS, in which there are questions considering which race the respondent belongs to. Note that the Census Bureau does not consider Hispanic/Latino to be a race group and therefore “Other race” is a common response for Hispanics. However, people of Hispanic/Latino origin have been re-coded by the Census Bureau and assigned to a race group that seems to be suitable due to their responses on other questions in the survey (country of origin, ancestry, etc.). That means Hispanics will be found in either White, Black or other races, depending on the Census Bureau’s assessment of their answers in the survey. Note that Whites constitute the largest share of immigrants, 47.93 %.

4.1.1 Employment

Graph 4.3 shows the employment status for the immigrants within the dataset, in total and by region of origin groups. Note that the share of unemployed among natives in this dataset is higher than the share of unemployed among immigrants. However, the share of unemployed differs between different groups of immigrants. The share of unemployed immigrants from Europe is 5.78 %, while the share of immigrants unemployed from Central America is 8.13 %. There are 3 323 200 observations, only foreign-born and within the labor force. Age between 16 to 95 years. The period is 2000 – 2017. The graph is made in Excel, based on census data from IPUMS.

Graph 4.3. Employment status.



**Canada, Puerto Rico, Guam, American Samoa, U.S. Virgin Islands, St. Pierre and Miquelon, Atlantic Islands.*

***Antarctica, born at sea, unknown.*

****Males within foreign born category (1 836 064 observations)*

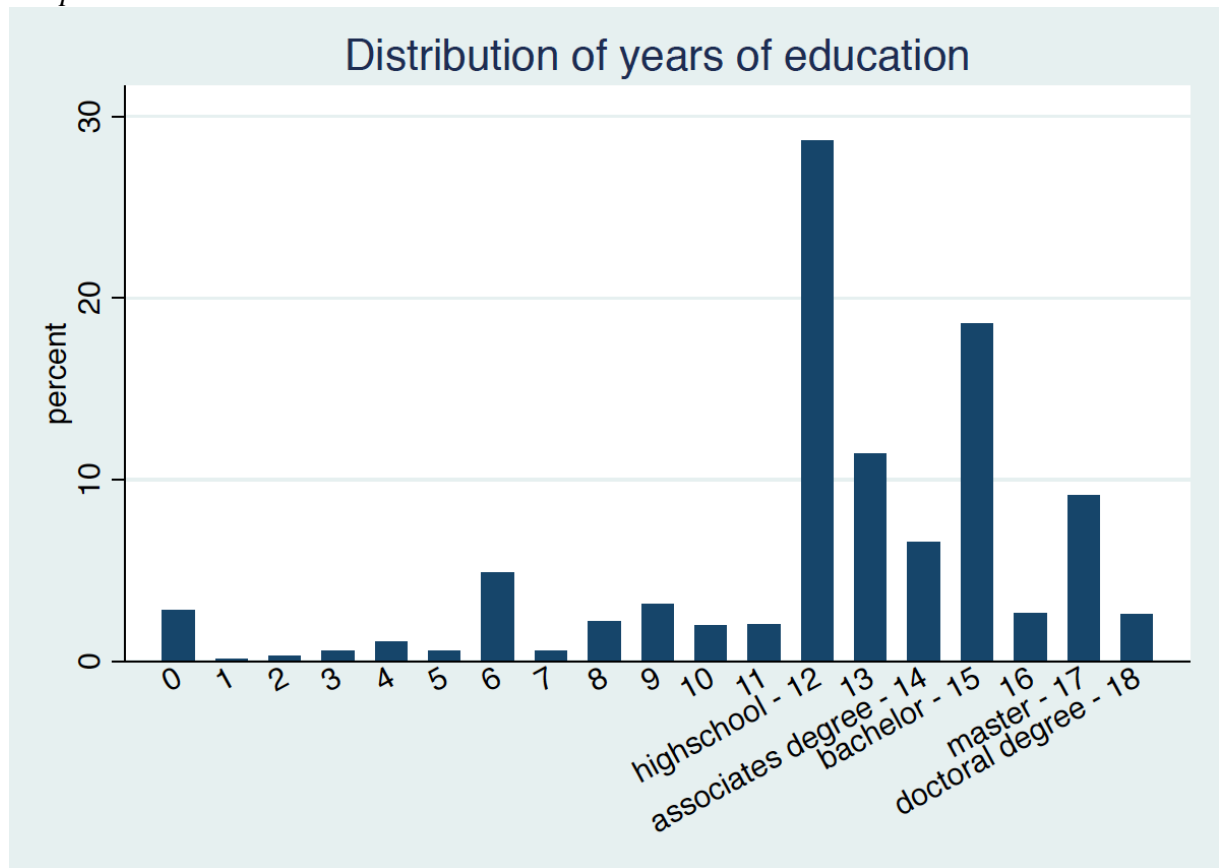
*****Females within foreign born category (1 487 136)*

******People born in U.S. and within labor force (1 889 409 observations)*

4.1.2 Education

Since we want to study the effect of education on employment status, it's relevant to present the average level of education among the immigrants in our dataset. Graph 4.4 shows the highest level of education among immigrants, as a percentage of the whole sample. There are 3 323 200 observations, including only foreign-born and individuals within the labor force. Age 16-95. The graph is made in STATA, based on data from IPUMS, period 2000 to 2017.

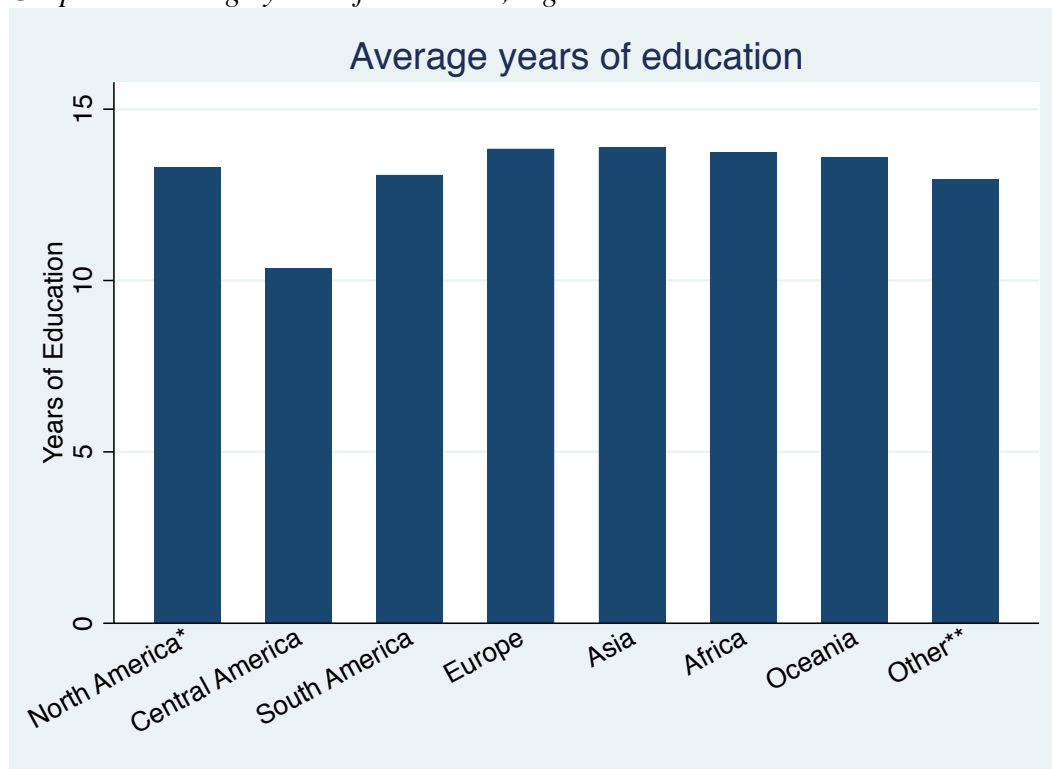
Graph 4.4. Education.



4.1.3 Education and birthplace

The level of education might differ depending on the region of origin. Graph 4.5 shows the average level of education for each region in our dataset. The average level of education is almost similar for all regions but Central America. The average years of education for immigrants from Central America are more or less ten years, while the average for Europeans is 14 years. There are 3 323 200 observations, including only foreign-born and individuals within the labor force. Age 16-95. The graph is made in STATA, based on data from IPUMS, period 2000 to 2017.

Graph 4.5 Average years of education, region.



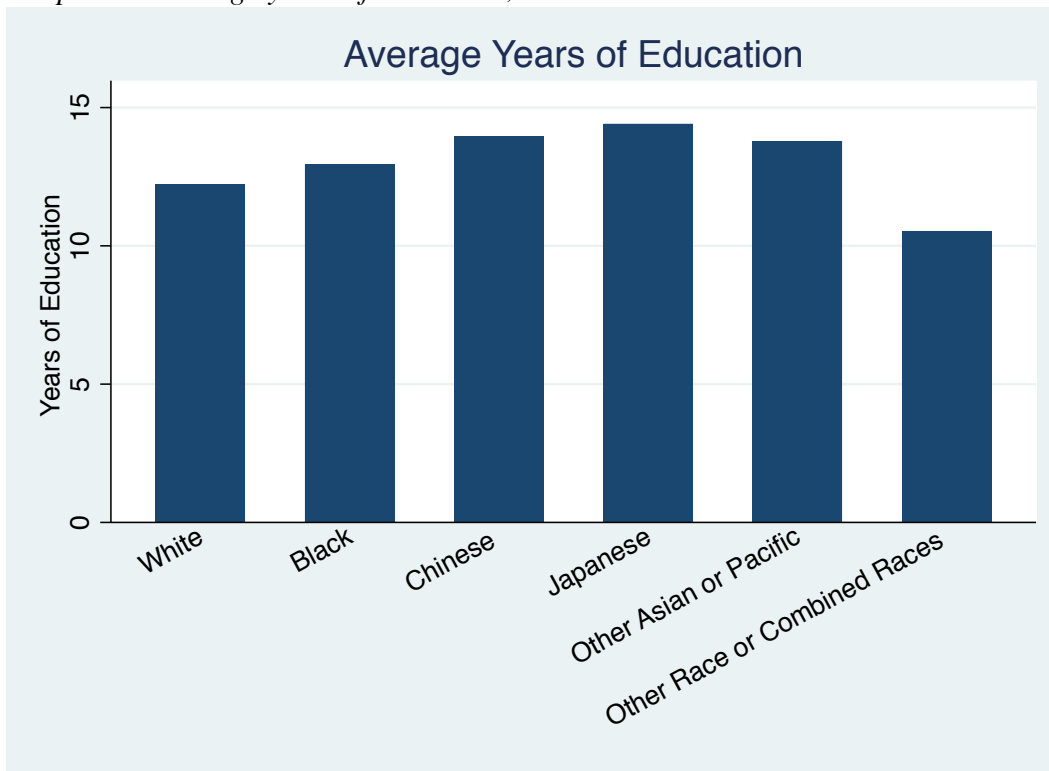
*Canada, Puerto Rico, Guam, American Samoa, U.S. Virgin Islands, St. Pierre and Miquelon, Atlantic Islands.

**Antarctica, born at sea, unknown.

4.1.4 Education and race

The level of education might differ depending on race as well. Graph 4.6 shows the average level of education for each race. Chinese and Japanese have the highest level of education on average, almost 14 years each, while Whites, Blacks, and Combined Races face the lowest level of education, on average. The low level of average education for Whites is probably caused by the fact the many Hispanics (Mexicans and Central American) are assigned to the race of Whites. There are 3 323 200 observations, including only foreign born and individuals within the labor force. Age 16-95. The graph is made in STATA, based on data from IPUMS, period 2000 to 2017.

Graph 4.6. Average years of education, race.



5. Empirical strategy

In this section, we present the strategy we have used in this paper to study our research question empirically.

The relationship between the level of education of the immigrant and the probability of getting employed can be formulated as an equation:

$$Prob(Y_{i,t,s} = 1) = \alpha + \beta Education_{i,t,s} + \lambda X_{i,t,s} + \delta_t + \gamma_s + \varepsilon$$

Where Y is a binary variable that takes the value 1 when individual i is employed and the value 0 when individual i is unemployed. $Education$ is the variable of interest and indicates the number of years of education for individual i , in state s , time t . X is a vector of control variables (birthplace, race, gender, age, age squared, number of years in U.S. and level of English skills).

δ indicates which state individual i lives in and controls for state fixed effects in the U.S. That is, controls for the fact that in some states it is easier for migrants to find a job. For instance, Florida is well-known for providing economic opportunities to Mexican workers because of the historical formation of migrant networks. On the other hand, finding a job in Ohio might be harder because people there might support migration policies less.

γ is the year the individual responded to the census. This variable controls for year fixed effects. That is, any macroeconomic shocks that occurred at year level and affect the return to education for all immigrants. For instance, the financial crisis or presidents with different economic politics.

To estimate the effect of education on employment status, i.e. estimate the coefficient β , we are doing a regression analysis in the statistical software program STATA. In particular, we are using a Linear Probability Model (LPM). In our case, we are estimating the probability that Y takes the value 1, i.e. that individual i is employed. Given that, the coefficients in our equation indicate the effect each variable has on the probability of Y taking value 1.

The strength of LPM is that the output is straightforward and easy to interpret. Additionally, LPM is commonly used for econometric analysis. However, LPM has its limitations. First, there is heteroscedasticity in the error term. To eliminate heteroscedasticity, we have used the command `robust` in STATA. Consequently, by using that command, we will not face heteroscedasticity. Second, LPM can give predictions that are less than 0 or greater than 1. Predictions like that are hard to interpret since the probability interval falls between 0 and 1. Luckily, we have not found any combinations of the independent variables that give a probability greater than 1 or less than 0 in our results. In the robustness section (section 6.4), we show that LPM does not differ from the marginal effects estimated in a Probit model.

6. Results

The result section is divided into four parts. The first section is the main result where we present the overall result based on the regression we have made in STATA. The second section analyses the heterogeneous effects by including interaction variables in the model. The third section expands the study to analyze the education effect on income. In the last section, we test the robustness of our model.

6.1 Main results

Overall, the effect of education on the probability of being employed is positive and significant. That is true for all immigrants, regardless of race and region of origin. Furthermore, the education effect is significant in each and every regression we have made, regardless of including control variables or not. Table 6.1 (p. 22) presents the main result of this paper. Model 1 includes only the variables of interest, Years of education. Model 2 controls for state fixed effects and year fixed effects. Model 3 includes control variables for age, age squared, gender, number of years the immigrant has stayed in the U.S., birthplace, English skills and race.

Table 6.1. Effect of education on probability of being employed.

| Dependent variable = (Employed=1) | (1) | (2) Fixed Effects | (3) Controls Effect |
|--|-----------------------|-----------------------|------------------------|
| Years of Education | 0.0040*** (0.0000) | 0.0040*** (0.0000) | 0.0030*** (0.0000) |
| Age | | | 0.0079*** (0.0001) |
| Age ² | | | -0.0001*** (0.0000) |
| Male | | | 0.0186*** (0.0003) |
| Years stayed in U.S. | | | 0.0002*** (0.0000) |
| English skills (Does not speak English = Benchmark group) | | | |
| Speaks, but not well | | | 0.0147*** (0.0008) |
| Speaks well | | | 0.0240*** (0.0008) |
| Speaks very well | | | 0.0279*** (0.0008) |
| Speaks only English | | | 0.0293*** (0.0009) |
| Birthplace region (Europe = Benchmark group) | | | |
| Central America | | | 0.0015*** (0.0005) |
| South America | | | -0.0015** (0.0007) |
| North America | | | -0.0126*** (0.0007) |
| Asia | | | -0.0067*** (0.0007) |
| Africa | | | -0.0047*** (0.0009) |
| Oceania | | | -0.0092*** (0.0018) |
| Other | | | -0.0254*** (0.0067) |
| Race (White = Benchmark group) | | | |
| Black | | | -0.0320*** (0.0007) |
| Chinese | | | 0.0147*** (0.0008) |
| Japanese | | | 0.0274*** (0.0013) |
| Other Asian or Pacific | | | 0.0078*** (0.0007) |
| Other Race or Combined Races | | | -0.0043*** (0.0004) |
| Constant | 0.8803*** (0.0005) | 0.9104*** (0.0025) | 0.7144*** (0.0031) |
| State Fixed Effects | | Yes | Yes |
| Year Fixed Effects | | Yes | Yes |
| Observations | 3,323,200 | 3,323,200 | 3,323,200 |
| R-squared | 0.0035 | 0.0092 | 0.0191 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

The output of model 1 implies a positive effect of education. In particular, one additional year of education for an immigrant increases the probability of being employed in the U.S. with 0.4 percentage points. The effect of education is the same in model 2, where we include variables which control for time trends and state differences. The results in model 1 and model 2 are highly significant. However, we cannot conclude that the education effect is positive by just observing the outcome of model 1 and 2. Presumably, there is a risk that other factors than education affect the probability of being employed. To take this into account, we include control variables in model 3. The outcome of model 3 yet implies a positive and significant effect of education, although we have control for other factors. The education effect is marginally lower than in model 1 and 2 though, 0.3 percentage points. It is worth to mention that when including control variables in Model 3, the R-squared increases compared to model 1 and 2. R-squared corresponds to the share of variance in the dependent variable that is explained by the independent variables. The increased R-squared in model 3 indicates that the control variables we have included are relevant for predicting the probability of being employed.

The fact that the education effect is between 0.3 and 0.4 percentage points might be seen as a relatively small rate of return to education. However, the returns to education seem larger considering the fact that the average numbers of years of education among immigrants in our dataset is 12 years. An immigrant with 12 years of education will face a 3.6 to 4.8 percentage points⁴ larger probability of being employed than one without any education at all, assuming everything else is the same.

To put the results of model 1, 2 and 3 in context we ran a regression with natives (individuals born in the U.S.) instead of foreign-born (see section 10.2, Table 5, in appendix). The output from that regression indicates that natives have a greater effect of education than foreign born: Natives face an increased probability of being employed with 1.25 percentage points for each additional year of education. That is more than 3 times greater effect than for foreign-born.

⁴ Calculated by using the predicted education effect interval in model 1-3, times the average years of education for immigrants in our dataset (12 years).

The positive effect of education is in line with other studies. Borjas (2014) found that one year of additional education increases the wage growth by 1.5 percentage points for immigrants in U.S. Aldén and Hammarstedt (2014) conclude that the employment rate in Sweden was higher for immigrants with education than for those without education and Bevelander's and Veenman's (2006) study implies there is a greater probability of being naturalized in the Netherlands for immigrants with higher education. The similarities between these results and ours are that the returns to education for immigrants are positive, regardless if it is measured in terms of wage growth, employment rate or probability of being employed.

Among the controls, almost all of them are significant, i.e. they have an impact on the probability being employed. The English skills variable shows an expected outcome. That is, the probability of being employed is higher for immigrants that speak English very well or speaks English fluently, compared to immigrants that do not speak English at all. For instance, immigrants that speak English very well are 2.79 percentage points more likely to be employed than those who not speak English at all.

The birthplace variable indicates that immigrants from Europe and Central America face advantage in terms of likelihood of being employed, compared to other regions. On the other hand, North American immigrants are 1.26 percentage points less likely to get employed compared to Europeans.

There is an indication that Black immigrants face difficulties in the labour market compared to other races. Looking at the variable for race, Blacks have 3.2 percentage points less probability of being employed compared to Whites. Furthermore, comparing the coefficient for Blacks with the coefficient for Japanese, there's a significant difference of 5.9 percentage points. This suggests that Japanese have 5.9 percentage points higher probability of being employed in the U.S. than Blacks. This is consistent with the literature of racial discrimination in the labor market (Bayer & Kerwin, 2018).

Model 3 shows that the effect of race is greater than the education effect. For instance, being Black gives the immigrant 3.2 percentage points less probability of being employed compared to Whites, while one year of education increases the probability

with only 0.3 percentage points. In reality, this can be exemplified as follows: Suggest a Black immigrant and a White immigrant with 0 years of education, everything else the same, immigrating to the U.S. When entering the U.S., the Black immigrant has 3.2 percentage points less chance to get a job, compared to the White immigrant, according to the race variable in model 3. If the Black immigrant strives to achieve the same probability of being employed as the White immigrant, he or she needs to study approximately ten years more than the White immigrant, assuming everything else is the same, *ceteris paribus*. If the Black immigrant studies ten years, the likelihood of being employed will increase by 3.0 percentage points⁵ and the Black immigrant will face almost the same chance as the White, non-educated immigrant, to be employed.

Evidently, the likelihood of being employed differs depending on where the immigrant comes from and what race the immigrant belongs to. In a back of envelope calculation, it takes about ten years of studies for a Black immigrant to reach the same probability of being employed as a White immigrant. However, we have not taken into account the heterogeneous effects of education according to race and birthplace. Considering heterogeneity, one can assume that the effect of education is different within the group of immigrants. That implies, one year of education might not have the same magnitude for immigrants from Europe compared to immigrants from Africa. To test whether there is heterogeneity according to race or birthplace, we have constructed another model (4), including interaction variables. In the next section, we present the result of the heterogeneity model.

6.2 Heterogeneous effects

In this model (4), we have included interaction variables: Years of education times race and years of education times birthplace. This aims to determine heterogeneous effects among immigrants. The outcome of the interaction terms indicates if there are any differences in the effect of education *within* the group of immigrants. Table 6.2 (p.26) shows the regression output for model 4.

⁵ Calculated by using the predicted effect of one year of education, 0.3 (r. 1), times 10 years of education.

Table 6.2⁶. *Heterogeneous Effects.*

| Dependent variable = (Employed=1) | (4) Interaction effects |
|--|----------------------------|
| Years of Education | 0.0046*** (0.0001) |
| <i>Birthplace region * Years of education</i> <i>(Europe = Benchmark group)</i> | |
| Central America | -0.0031*** (0.0002) |
| South America | -0.0018*** (0.0002) |
| North America | 0.0055*** (0.0003) |
| Asia | -0.0009*** (0.0003) |
| Africa | -0.0013*** (0.0004) |
| Oceania | 0.0044*** (0.0008) |
| Other | 0.0021 (0.0021) |
| <i>Race * Years of education</i> <i>(White = Benchmark group)</i> | |
| Black | 0.0046*** (0.0002) |
| Chinese | -0.0010*** (0.0003) |
| Japanese | -0.0017*** (0.0006) |
| Other Asian or Pacific | -0.0003 (0.0002) |
| Other Race of Combined Races | 0.0000 (0.0001) |
| Constant | 0.6932*** (0.0035) |
| Year Fixed Effects | Yes |
| State Fixed Effects | Yes |
| Controls Effects | Yes |
| Observations | 3,323,200 |
| R-squared | 0.0198 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

⁶ Due to space limitations, Table 6.2 only shows the variable of interest, the interaction term between race and years of education and the interaction term between years of education and birthplace. However, same control variables as in model 1-3 have been used in model 4.

The output of the *interaction term* between years of education and birthplace region should be interpreted as follows: immigrants from Europe are the benchmark group. That is, their education effect is 0.0046 (0.46 percentage points), due to the coefficient of the “general” years of education variable. The coefficients for the other regions within the interaction variable category indicate the difference in education effect *compared* to European immigrants. For instance, the coefficient for Central America is negative (-0.0031). That means, one additional year of education does not increase the probability of being employed as much as for an immigrant from Europe. For an immigrant from Europe, the education effect is 0.46 percentage points (one year increases the probability of being employed by 0.46 percentage points), while the effect for an immigrant from Central America is just 0.15 percentage points (0.46 – 0.31).

The *education effect* for North American immigrants is, on the other hand, greater than for European immigrants, 1.01 percentage points (0.46 + 0.55). Hence, there are differences within the group of immigrants. The effect of education on the probability of being employed differ depending on region of origin. Note, the differences in education effect between the birthplace regions are all significant despite “Other”.

Looking at the other interaction term, the one with years of education and races, some of the differences between the races are not significant. However, the coefficient for Blacks is significant and indicates that the returns to education are greater for Blacks than for Whites (White is benchmark group). One additional year of education for Black immigrants increases the probability of being employed to a greater extent than for White immigrants. For Japanese and Chinese, it is the opposite. The effect of one year of additional education is lower for them than for Europeans and Blacks.

6.3 The effect of education on wages

In the previous sections, we have observed that education does have a positive impact on the probability of being employed, although the effect of one additional year of education seems to be relatively low. Additionally, we have found that the magnitude of the education effect differs depending on race and the country of origin. In this section, we expand our study to analyzing wage effects. This is relevant since, as described in the literature review, most of the research made in immigration economics uses wage as the dependent variable. In order to make our paper even more comparable with recent studies, we will use this approach in the extension of the paper.

Model 5 is in many aspects similar to the other models in this paper. The variable of interest is years of education, we control for state and year fixed effects and the control variables are exactly the same. However, the dependent variable is changed to Log (wage), which means the logarithm of wage. Note the wage corresponds to the income an individual receives from employment only. In this model, we are able to estimate the effect of one additional year of education on the relative change in wage for immigrants. Note that in model 5 there are 332 371 fewer observations than in the other models. That is because model 5 only analyzes the effect for those immigrants that have a wage greater than 0. Table 6.3 (p.29) summarizes the output from the regression of model 5.

The result implies that one additional year of education increases the wage with 6.89 %. That is significant at 1 % level. Obviously, the returns to education are positive in terms of wage and this confirms the results of other studies made in immigration economics. For instance, Borjas (2014); each year of additional schooling increases the wage growth with 1.5 %; and Chiswick (1978); one additional year of schooling increases the wage with 5.7 %.

Table 6.3. Effects on wage.

| Dependent variable = Log (wage) | (5) Wage effects |
|--|------------------------|
| Years of Education | 0.0689*** (0.0002) |
| Age | 0.1428*** (0.0003) |
| Age ² | -0.0015*** (0.0000) |
| Male | 0.4470*** (0.0011) |
| Years stayed in U.S. | 0.0052*** (0.0001) |
| English skills (Does not speak English = Benchmark) | |
| Speaks, but not well | 0.0094*** (0.0025) |
| Speaks well | 0.1162*** (0.0025) |
| Speaks very well | 0.3571*** (0.0026) |
| Speaks only English | 0.3583*** (0.0029) |
| Race (White = Benchmark group) | |
| Black | -0.1328*** (0.0025) |
| Chinese | 0.1121*** (0.0034) |
| Japanese | 0.1437*** (0.0073) |
| Other Asian or Pacific | 0.0621*** (0.0028) |
| Other Race or Combined Races | -0.0484*** (0.0015) |
| Birthplace region (Europe = Benchmark group) | |
| Central America | -0.0906*** (0.0020) |
| South America | -0.1132*** (0.0027) |
| North America | -0.0659*** (0.0029) |
| Asia | -0.0409*** (0.0029) |
| Africa | -0.1120*** (0.0038) |
| Oceania | 0.0242*** (0.0081) |
| Other | -0.2298*** (0.0245) |
| Constant | 5.2892*** (0.0129) |
| State fixed effects | Yes |
| Year fixed effects | Yes |
| Observations | 2,990,829 |
| R-squared | 0.2872 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

There are outcomes among the control variables that are worth to mention. For instance, the wage for a male is 44.7 % higher than for a female at a significant level of 1 %. The number of years the immigrant has stayed in the U.S. have a relatively small effect on the wage, 0.52 % per year. On the other hand, the level of English skills has a greater effect on the wage. Individuals that speak English fluently or very well earn approximately 35 % more than those who cannot speak English at all. Combining these findings gives us a conclusion that, in terms of wage, it is of greater importance having high English skills than having stayed in the U.S. for a long time.

Consistent with the results from previous models in our paper, model 5 implies a disadvantage of being Black on the labor market: Foreign-born Blacks earn 13.28 % less than foreign-born Whites. Moreover, Japanese and Chinese face disadvantages in the labor market, which is consistent with the results from the other models as well. They earn 14.37 % and 11.21 % more than Whites, respectively.

The birthplace variable shows an interesting result. In model 5, the effect on the wage of being born in Central America is negative and significant (-9.06 % compared to European immigrants). However, the result in model 3 suggests a *positive* effect of being born in Central America on the probability of being employed. Consequently, an immigrant that is born in Central America faces a greater probability of being employed but, on the other hand, receives less income, compared to other regions. Presumably, Central American immigrants get jobs in the low-paid sector, where the probability of being employed is high but where the wages are low. Concluding that Central American immigrants are drawn to the low paid/low skilled sector may be reasonable considering the fact that they are the least educated group of immigrants, on average (see table 4.5, p.18). This is consistent with Duncan and Trejo (2011) that found that low-skilled migrants do not face problems of finding employment. However, they also found that the wages for these immigrants are relatively low, reflecting the lack of skills.

As in model 1 to 3 we want to put this result in another context. Therefore, we have made a log(wage) regression on natives as well (see section 10.2, Table 6, in the appendix). Comparable to the 6.89 % effect of education on wage for immigrants, the

effect of education on wages is 15.86 % for natives. That means natives face more than 2 times greater effect of one year of education compared to foreign-born.

6.4 Robustness check

The result is in general robust. All of the coefficients are significant at 1 % level in model 1, 2 and 3. The coefficient for the variable of interest, years of education, is significant at 1 % level throughout all of the models we have tested. To analyze the robustness further, we tested whether the magnitude of the education effect was the same, regardless of the choice of a regression model. To test this, we regress model 1 again, but with a probit regression instead of the LPM and then calculating the marginal effects for years of education. Our result seems to be robust according to this test since the marginal effect for education after running the probit regression is equal or almost similar to the education effect our models suggest. The marginal effect of education is 0.38 percentage points, according to the probit regression, and the effect of education in our model (1) is, as shown, 0.40 percentage points. Table 6.4, shows the output from probit regression and the calculated marginal effects. (See appendix, section 10.3, for commands in STATA).

Table 6.4. Marginal effects from Probit Regression.

| Dependent variable = (Employed=1) | (1) Probit Regression | (2) Marginal effect |
|-----------------------------------|--------------------------|------------------------|
| Years of education | 0.0282*** (0.0002) | 0.0038*** (0.0000) |
| Constant | 1.1354*** (0.0031) | |
| Observations | 3,323,200 | |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

7. Discussion

This section aims to analyze the result of the paper and connect them to the theories presented in previous sections of the paper. This is followed by a section where the strengths and limitations of the study are discussed.

In the context of this paper, we measure assimilation in the host country as employment status for the immigrants. Given that, the results in this paper show that immigrants with higher education assimilate better than immigrants with lower education, although the effect of education is not very high.

One theory presented at the beginning of the paper, based on Chiswick's (1978) study about earnings and education among foreign-born, implies that the returns to education might differ depending on the country of origin. Our result confirms his theory. By observing the interaction term between years of education and birthplace region, we conclude that the effect of education is higher for immigrants from Europe, North America and Oceania than from other regions. This might be explained by Borjas (2014) and his findings about determinants of economic assimilation, where he implies that immigrants from High-GDP countries can easily transfer skills to the U.S. That explanation is suitable for our finding about the differences in education effect for different regions; regions that have the highest education effect for immigrants include countries with relatively high GDP.

Another theory presented in this paper, aimed to explain the relationship between education and employment, is the selection theory, implemented by Roy (1951) and Borjas (2014). That implies there might be either a positive or negative selection among immigrants, where positive selection results in only highly skilled and motivated individuals to migrate. This theory may explain some findings in our paper. For instance, in model 3 (variable Race), we can see a significant greater probability of being employed for Japanese compared to Whites. According to the result, the Japanese have 2.74 percentage points higher probability of being employed compared to Whites. Furthermore, Japanese have, on average, two years' higher education level than Whites (see table 4.6, p.19). These findings imply two things: first, the positive correlation between education and employment is distinct. Higher education on

average for Japanese induces higher probability of being employed, compared to Whites. Second, there might be a positive selection of Japanese immigrants, resulting in higher educated and motivated individuals from Japan than from elsewhere, which in turn might increase the Japanese probabilities of being employed in the U.S.

Another interesting finding is that the effect of how many years the individual has lived in the U.S. on the employment probability is relatively small (0.02 percentage points per year). An individual who has lived in the U.S. for ten years increases its probability of being employed by 0.2 percentage points, compared to one who just immigrated. To put this in context, the effect of being a male compared to female is 1.87 percentage points. Consequently, being born as a male is worth approximately 93 years in the U.S., in terms of the effect on the probability of being employed.

As stated in the introduction of the paper, the numbers of foreign-born in the world and in the U.S. have never been higher. That makes the discussion of assimilation relevant. How do we guarantee that all these people that migrate from one country to another, assimilate in the host country? This paper does not aim to answer a complex question like that. However, we can conclude that there are obviously no easy answers to that question. First, there are different definitions of assimilation. In this paper, we analyze economic assimilation and considering employment status as the most important aspect of assimilation, whilst others believe cultural aspects such as clothes, language and intermarriages are more important aspects when discussing assimilation. Second, even if we decide to focus at only one aspect of the assimilation, as we do in this paper, the answers of how we get people assimilated are still complex. Certainly, educated people assimilate better than non-educated people, according to our result. But so do Whites, Japanese and Males. Central Americans assimilate better than others, in terms of being employed, but they receive lower wages than other migrants. Blacks assimilate worst, but they have a greater effect of education than Whites. Obviously, it is not easy to answer how to improve the assimilation of foreign-born. However, considering the immigration policy President Trump suggested during spring 2019, described in the introduction of the paper, one may assume President Trump believes that by allowing only migrants with high level of education to enter the U.S., the problems of assimilation for foreign-born will be solved. According to our result, racial discrimination against Blacks on the U.S. labor market affects the

assimilation to a greater extent than the level of education. Thus, to improve the assimilation among foreign born in the U.S., we recommend policy makers in the U.S. to consider these issues when implementing new policies. Allowing only highly educated migrants to enter the U.S. will probably not solve the problems of assimilation for foreign born, at least not as long as racial discrimination still occurs on the U.S. labor market.

7.1 Strengths and limitations

The strengths of this study rely on the data. The source of the data is reliable and has been used before in other studies with a similar purpose. The sample size is big and represent the population in a proper way. This makes the output of the regression robust. Furthermore, the data provided contained almost all of the variables we required to make this study, such as age, birthplace, and race. In other words, the data used is very detailed which make the analysis robust. In addition, a study like this requires data on an individual level which the data used provided.

However, even if the data used is very robust and detailed there are some limitations. The education variable does not say anything about where the education was taken. Thus, we cannot say anything about how education taken in different regions affect the probability of being employed. It would have been desirable to determine if immigrants with an American education face a higher probability of being employed than immigrants with an education from abroad. Unfortunately, that is not possible within the context of this study due to the limitations of the data.

The data does not contain illegal immigrants. Certainly, it is not expected to take into account illegal immigrants in a study like this. However, it is still worth to mention that this group of immigrants, which we may assume is considerable, is not included in our study.

The study does not take into account the reason to why the immigrants in our dataset chose to migrate to the U.S. In future studies, it would be interesting to limit the study to a specific group of immigrants, for instance, humanitarian immigrants (refugees).

The study does not analyze if the marginal effects of education are different for different levels of education. For instance, the marginal effect of completing the twelfth year of education might be different compared to completing the fifth year of education. However, this study assumes that the marginal effect of education is equal for all levels of education.

We cannot completely exclude the risk for the result to be biased due to endogeneity. Endogeneity is a common problem when using statistical methods to estimate relationships between economic factors. In short, endogeneity occurs when the variables of interest, the explanatory variable, is correlated with the unobserved component in the equation. Endogeneity is a problem because the systematic correlation between the variable of interest and unobserved variables makes the estimation of the effect of the variable of interest biased. If changes in the variable of interest are systematically followed by changes in the unobserved variable, it is impossible to distinguish which one of these changes that affect the outcome of Y , and therefore it is impossible to estimate the true effect of the variable of interest. In our case, endogeneity arises if our variable of interest, years of education, is correlated with unobserved variables in ϵ , for instance, ability, motivation, and effort. To solve this problem, we have tried our best to find variables that we can observe and “pick them out” from the unobserved variable. Thus, we included age, gender, birthplace, race, English skills and years stayed in the U.S. in our models. However, we cannot include all factors as a control variable, since it is impossible to measure them. For instance, it is hard to measure ability or motivation and therefore we cannot include these variables in the model.

An ideal solution to endogeneity is to use an instrumental variable that is correlated with the variable of interest but uncorrelated with the unobserved variable. However, this solution is hard since it is complicated to find an instrumental variable that satisfies these conditions. In our case, we could not find a variable that was correlated with years of education but not correlated with the unobserved variable. Instead, we tried to include as many control variables as possible to control for endogeneity.

Last, the study is only applicable in the U.S. since the sample represents the U.S. population. Presumably, the distribution of foreign-born in other regions of the world might look different than in the U.S. as well as the labor market and how the assimilation works.

8. Conclusion

This section concludes the paper in short and answers the research question. In order to make the paper coherent, the section starts with a reminder of the research question.

The aim of this paper was to answer the research question: *Does the level of education affect the probability of being employed for immigrants in the United States?*

The result suggests that the level of education does have an impact on the probability of being employed for immigrants in the U.S. during the period of 2000-2017. In particular, one additional year of education has a positive effect on the probability of being employed. This is statistically significant, due to our study. The magnitude of the positive effect of education fluctuates between 0.15 and 1.01 percentage points, due to heterogeneous effects according to birthplace and race. Immigrants from Europe, North America and Oceania obtain a greater effect of education than immigrants from Central America, Africa, South America and Asia. Categorize immigrants into races, the result suggests Blacks having a larger effect of education than Whites. On the other hand, the Chinese and Japanese have less returns to education than both Whites and Blacks.

Certainly, the effect of education is positive for immigrants. However, we have found that other factors, in particular race and country of origin, affect the probability of being employed to a greater extent than education. Given the definition of economic assimilation in this paper, employment status, we conclude that higher educated immigrants indeed assimilate better than less educated, but that the assimilation is affected more by race and country of origin.

9. References

- Abramitzky, R., Boustan, L., Eriksson, K. (2018) *Cultural Assimilation during the Two Ages of Mass Migration*. Working Paper, No. 1013. Stanford Center on Global Poverty and Development. Retrieved from: https://siepr.stanford.edu/sites/default/files/publications/WP1013_0_58.pdf
- Aldén, L., Hammarstedt, M. (2014) *Integration of immigrants on the Swedish labour market – recent trends and explanations*. Centre for Labour Market and Discrimination Studies at Linnaeus University. Retrieved from: https://lnu.se/contentassets/729bb815ca184ffc9d1c33192c8b9a55/lmdswp20149_mall.pdf
- Bayer, P., Kerwin, C. K. (2018) *Divergent Paths: A New Perspective on Earnings Differences Between Black and White Men Since 1940*. Working Paper, No. 2018-45. Becker Friedman Institute. Retrieved from: <https://bfi.uchicago.edu/working-paper/divergent-paths-a-new-perspective-on-earnings-differences-between-Black-and-White-men-since-1940/>
- Bevelander, P., Veenman, J. (2006) *Naturalisation and Socioeconomic Integration: The Case of the Netherlands*. Discussion Paper, No. 2152. The Institute for the Study of Labor (IZA) Retrieved from: <http://ssrn.com/abstract=908239>
- Borjas, G. J. (2003) *The labor demand curve is downward sloping: Reexamining the impact of immigration on the labor market*. The quarterly journal of economics, Vol. 118, No. 4 (Nov., 2003), pp. 1335-1374. Retrieved from: <https://sites.hks.harvard.edu/fs/gborjas/publications/journal/QJE2003.pdf>
- Borjas, G. J. (2014) *Immigration economics*. Cambridge, Massachusetts: Harvard University Press.
- Borjas, G. J., Freeman, R.B., Katz, L.F. (1991) *Working paper on the labor market effects of immigration and trade*. Working Paper, No. 3761. National Bureau of Economic Research. Retrieved from: <https://www.nber.org/papers/w3761.pdf>
- Brückner, H., Jahn, T. E., (2011) *Migration and Wage-Setting: Reassessing the Labor Market Effects of Migration*. The Scandinavian Journal of Economics, Vol. 113, No. 2 (June 2011), pp. 286-317. Retrieved from: <https://www.jstor.org/stable/23016837>
- Chiswick, B. (1978) *The Effect of Americanization on the Earnings of Foreign-born Men* Source: Journal of Political Economy, Vol. 86, No. 5 (Oct., 1978), pp. 897-892.
- Duncan, B., Trejo, S. J. (2011) *Low-Skilled Immigrants and the U.S. Labor Market*. Discussion Paper, No. 5964. The Institute for the Study of Labor (IZA) Retrieved from: <http://ftp.iza.org/dp5964.pdf>
- Dustmann, C., Schonberg, U., Stuhler, J. (2016) *Labor supply shocks, native wages, and the adjustment of local employment*. Discussion Paper, No. 10114. The Institute for the Study of Labor (IZA). Retrieved from: <https://www.econstor.eu/bitstream/10419/145248/1/dp10114.pdf>
- East, C.N, Luck, P., Mansour, H., Velasquez, A. (2008) *The Labor Market Effects of Immigration Enforcement*. Discussion Paper, No. 11486. The Institute for the Study of Labor (IZA). Retrieved from: <http://ftp.iza.org/dp11486.pdf>
- Foged, M., Peri, G. (2016) *Immigrants' effect on native workers: New analysis on longitudinal data*. Discussion Paper, No. 2152. The Institute for the Study of Labor (IZA). Retrieved from: <http://ftp.iza.org/dp8961.pdf>
- Fouka, V., Mazumder, S., Tabellini, M. (2018) *From Immigrants to Americans: Race and Assimilation during the Great Migration*. Harvard Business School. Retrieved from: https://www.hbs.edu/faculty/Publication%20Files/19-018_b6563a83-4253-46ad-a437-a781894fecfe.pdf
- IPUMS USA (2019) Retrieved from: <https://usa.ipums.org/usa/index.shtml>
- OECD (2018) *International Migration Outlook 2018*. OECD Publishing, Paris. Retrieved from: https://doi.org/10.1787/migr_outlook-2018-en
- Roy, A.D. (1951) *Some thoughts on the distribution of earnings*. Oxford Economic Papers 3, pp. 135-146. Retrieved from: <https://faculty.smu.edu/millimet/classes/eco7377/papers/roy%201951.pdf>

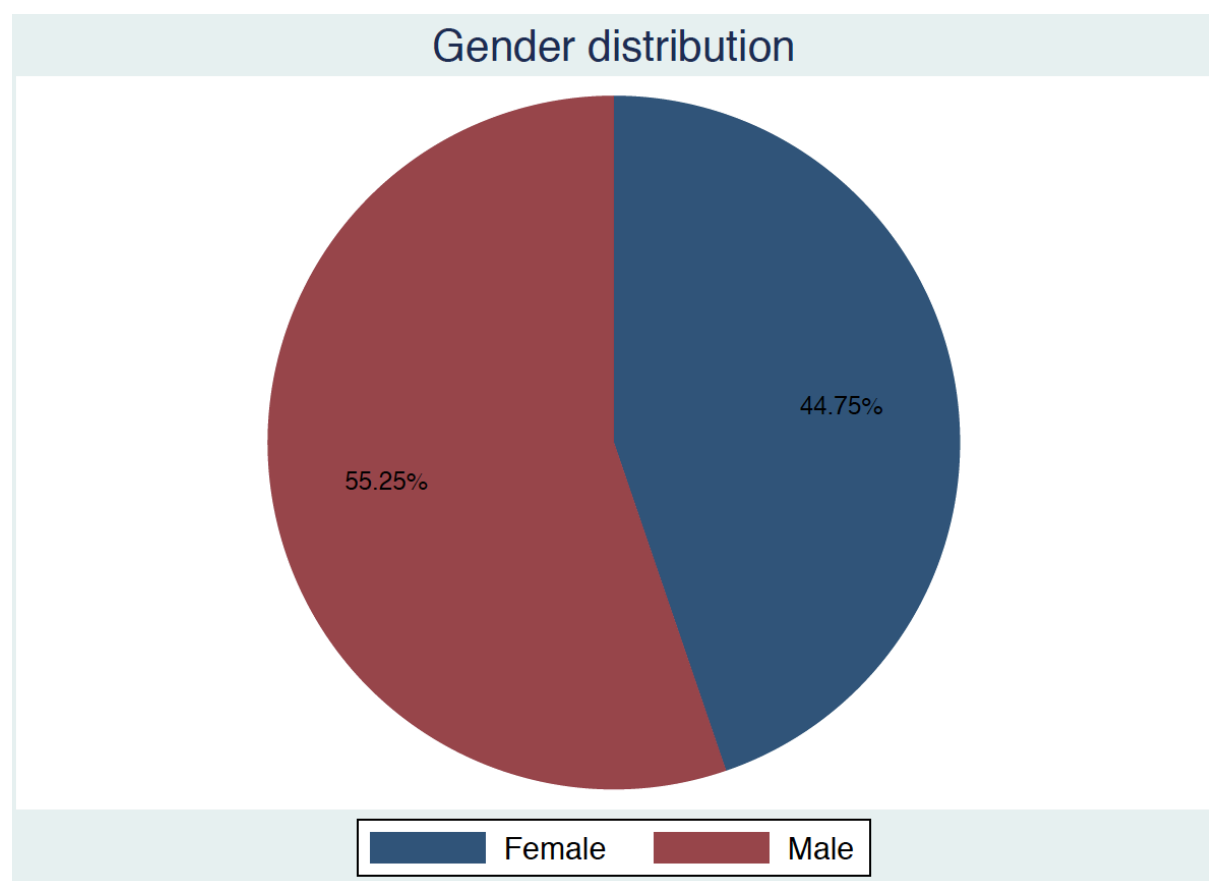
Sameeksha, D., Eklund, J., Nabavi, P. (2016) *Labor market outcomes of immigrants and natives: Evidence on wages in Sweden, 2001-2012*. Working Paper, No. 2016:43. Swedish Entrepreneurship Forum. Retrieved from: https://entreprenorskapsforum.se/wp-content/uploads/2016/09/WP_43.pdf

Schmitter Heslier, B. (2008) *The Bracero Program and Mexican Migration to the United States*. Journal of the West. Retrieved from: https://www.researchgate.net/publication/281321058_The_Bracero_Program_and_Mexican_Migration_to_the_United_States

Shear, M. (2019, 15th May) *Trump Immigration Proposal Emphasizes Immigrants' Skills Over Family Ties*. New York Times. Retrieved: <https://www.nytimes.com/2019/05/15/us/politics/trump-immigration-kushner.html>

10. Appendices

10.1 Graphs



Observations: 3 323 200

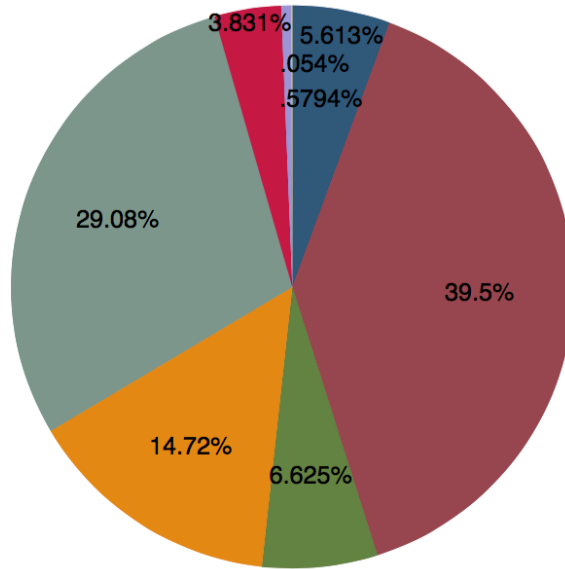
Foreign born individuals within the labor force in the U.S.

Age: 16-94

Period: 2000-2017

Based on census data from IPUMS. Made in STATA.

Birthplace region distribution



| | |
|----------------|-----------------|
| North America* | Central America |
| South America | Europe |
| Asia | Africa |
| Oceania | Other** |

*Canada, Puerto Rico, Guam, American Samoa, U.S. Virgin Islands, St. Pierre and Miquelon, Atlantic Islands.

**Antarctica, born at sea, unknown.

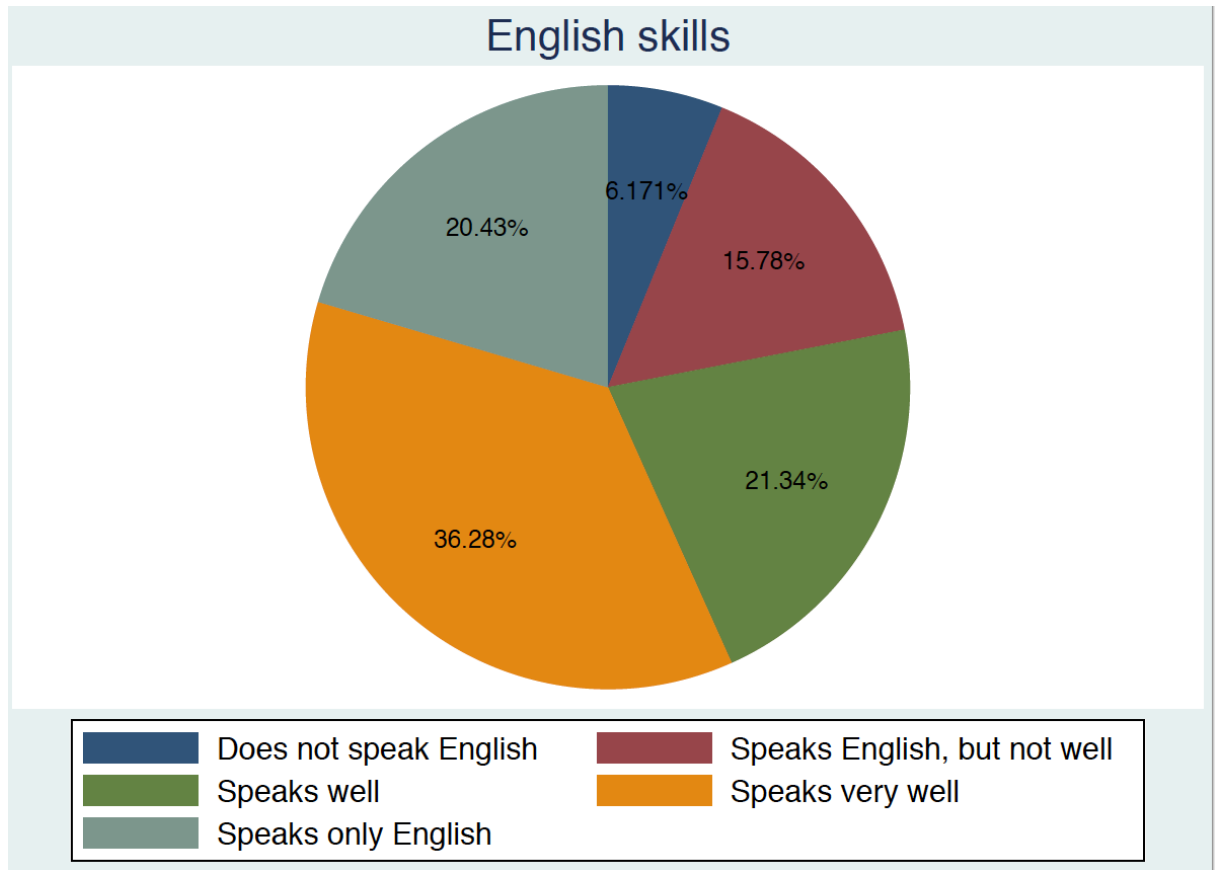
Observations: 3 323 200

Foreign born individuals within the labor force in the U.S.

Age: 16-94

Period: 2000-2017

Based on census data from IPUMS. Made in STATA.



Observations: 3 323 200

Foreign born individuals within the labor force in the U.S.

Age: 16-94

Period: 2000-2017

Based on census data from IPUMS. Made in STATA.

10.2 Tables

Table 1. Variables requested from IPUMS.

| Variable | Label | Type | Description |
|-----------------|--|-------------|---|
| YEAR | Census year | Numeric | Reports when the household/individual was included in the census |
| DATANUM | Data set number | Numeric | Identifies the particular sample from which the case is drawn in a given year. 1 = ACS sample |
| SERIAL | Household serial number | Numeric | 8-digit variable that assigns a unique identification number to each household in a given sample |
| CBSERIAL | Original Census Bureau household serial number | Numeric | 8-digit variable that assigns a unique identification number to each household in a given sample |
| HHWT | Household weight | Numeric | 6- digit numeric variable indicates how many households in the U.S. population are represented by a given household in an IPUMS sample |
| STATEICP | State (ICPSR code) | Categorical | Identifies the state in which the household is located (using the ICPSR coding scheme) |
| STATEFIP | State (FIPS code) | Categorical | Identifies the state in which the household is located (using the FIPS coding scheme) |
| COUNTYICP | County (ICPSR code) | Categorical | Identifies the county in which the household is located (using the ICPSR coding scheme) |
| COUNTYFIP | County (FIPS code) | Categorical | Identifies the county in which the household is located (using the FIPS coding scheme) |
| GQ | Group quarters | Categorical | Classifies all housing units into one of three categories: Households, group quarters or vacant units |
| PERNUM | Person number in sample unit | Numeric | Specific number to all persons within a household. |
| PERWT | Person weight | Numeric | 6- digit numeric variable indicates how many persons in the U.S. population are represented by a given person in an IPUMS sample. A value 010461 should be interpret as 104.61. |

| | | | |
|----------------------|---|-------------|---|
| SEX | Sex | Dummy | Report whether the respondent is male or female |
| AGE | Age | Numeric | Reports age in years as of the last birthday |
| FERTYR | Children born within the last year | Dummy | Women, ages 15-50, regardless of marital status reports if they had given birth to a child the past 12 months |
| RACE (general) | Race – general version | Categorical | Reports race of the respondent |
| RACED (detailed) | Race – detailed version | Categorical | Reports race of the respondent (detailed) |
| HISPAN (general) | Hispanic origin – general version | Dummy | Reports if respondent is Hispanic or not |
| HISPAND (detailed) | Hispanic origin – detailed version | Categorical | Reports if respondent is Hispanic or not |
| BPL (general) | Birthplace – general version | Categorical | Indicates the state or foreign country the person was born |
| BPLD (detailed) | Birthplace – detailed version | Dummy | Indicates the state or foreign country the person was born |
| YRNATUR | Year naturalized | Numeric | Reports the year in which the person became naturalized in U.S. |
| YRIMMIG | Year of immigration | Numeric | Reports the year the respondent immigrated to U.S. |
| YRSUSA1 | Years in the United States | Numeric | Reports how long a person who was born in foreign country has lived in U.S. |
| LANGUAGE (general) | Language spoken – general version | Categorical | Reports the language respondent speaks at home |
| LANGUAGED (detailed) | Language spoken – detailed version | Categorical | Reports the language respondent speaks at home |
| SPEAKENG | Speaks English | Categorical | Indicates if the respondent speaks English at home and if yes, how well the respondent speaks |
| SCHOOL | School attendance | Dummy | Indicates whether respondents attended school during the specific period |
| EDUC (general) | Educational attainment – general version | Categorical | Indicates the respondent educational attainment, measured as the highest year of school completed |
| EDUCD (detailed) | Educational attainment – detailed version | Categorical | Indicates the respondent educational attainment, measured as the highest year of school completed |
| EMPSTAT (general) | Employment status – general version | Dummy | Indicates whether the respondent was a part of the labor force and if so, whether the person was currently unemployed |
| EMPSTATD (detailed) | Employment status – detailed version | Dummy | Indicates whether the respondent was a part of the labor force and if |

| | | | |
|----------|------------------------|---------|---|
| | | | so, whether the person was currently unemployed |
| LABFORCE | Labor force status | Dummy | Indicates whether the person is a part of the labor force or not |
| INCWAGE | Wage and salary income | Numeric | Reports respondents total pre-tax wage and salary received as an employee |

Table 2. Variables used in the regressions of this paper.

| Variable | Label | Type | Description |
|-----------------|----------------------------|-------------|---|
| EMPSTATUS | Employment status | Dummy | 0=unemployed, 1 = employed |
| YEARS EDUC | Years of education | Numeric | 0-18 years |
| SPEAKSENG | Level of English skills | Ordinal | 0 = does not speak English 1 = does not speak well 2 = speaks well 3 = speaks very well 4 = Speaks only English |
| BPLACEREGION | Birthplace region | Categorical | 0 = Europe 1 = Central America 2 = South America 3 = North America 4 = Asia 5 = Africa 6 = Oceania 7 = Other |
| RACEOFIND | Race of individual | Categorical | 0 = White 1 = Black 2 = Chinese 3 = Japanese 4 = Other Asian Race 5 = Other Race or Combined Races |
| GENDER | Gender | Dummy | 0=female, 1=male |
| AGE | Age | Numeric | Age in years |
| AGE_2 | Age ² | Numeric | Age squared |
| YRSUSA1 | Years in the United States | Numeric | Number of years the individual has lived in U.S. |

| | | | |
|-----------------------------|---|-------------------------|---|
| YEARESDUC x RACEOFIND | Years of education times Race of individual <i>i</i> | Interaction variable | The premium effect of one additional year of education due to Race |
| YEARESDUC x BPLACEREGION | Years of education times Birthplace region of individual <i>i</i> | Interaction variable | The premium effect of one additional year of education due to region of birth |
| LOG (WAGE) | The logarithm of the income from employment of individual <i>i</i> | Logarithm variable | Indicate the relatively change (%) in income when the independent variable increases with one unit. |

Table 3. Employment status, observations.

| Category | Unemployed | Employed | Total |
|--------------------|------------|------------|------------|
| Female**** | 117 291 | 1 369 845 | 1 487 136 |
| Male*** | 115 125 | 1 720 939 | 1 836 064 |
| North America* | 13 873 | 172 647 | 186 520 |
| Central America | 106 721 | 1 205 841 | 1 312 562 |
| South America | 15 264 | 204 896 | 220 160 |
| Europe | 28 269 | 460 968 | 489 237 |
| Asia | 56 385 | 909 966 | 966 351 |
| Africa | 10 489 | 116 830 | 127 319 |
| Oceania | 1 255 | 18 000 | 19 255 |
| Other** | 160 | 1 636 | 1 796 |
| Total Foreign Born | 232 416 | 3 090 784 | 3 323 200 |
| Natives***** | 1 330 430 | 17 558 979 | 18 889 409 |

Table 4. Employment status, percent (%)

| Category | Unemployed | Employed | Total |
|--------------------|------------|----------|-------|
| Female**** | 7,89% | 92,11% | 100% |
| Male*** | 6,27% | 93,73% | 100% |
| North America* | 7,44% | 92,56% | 100% |
| Central America | 8,13% | 91,87% | 100% |
| South America | 6,93% | 93,07% | 100% |
| Europe | 5,78% | 94,22% | 100% |
| Asia | 5,83% | 94,17% | 100% |
| Africa | 8,24% | 91,76% | 100% |
| Oceania | 6,52% | 93,48% | 100% |
| Other** | 8,91% | 91,09% | 100% |
| Total Foreign born | 6,99% | 93,01% | 100% |
| Natives***** | 7,04% | 92,96% | 100% |

Table 5. Regression (6) output.

Natives - Education effect on the probability on being employed

| Dependent variable = (Employed=1) | (6) Natives |
|---------------------------------------|------------------------|
| Years of education | 0.0125*** (0.0000) |
| Age | 0.0081*** (0.0000) |
| Age^2 | -0.0001*** (0.0000) |
| Male | -0.0075*** (0.0001) |
| Race (White = Benchmark group) | |
| Black | -0.0682*** (0.0003) |
| Chinese | 0.0053*** (0.0011) |
| Japanese | 0.0044*** (0.0011) |
| Other Asian or Pacific | -0.0036*** (0.0007) |
| Other Race or Combined Races | -0.0342*** (0.0004) |
| Constant | 0.6325*** (0.0047) |
| State fixed effects | Yes |
| Year fixed effects | Yes |
| Observations | 18,889,409 |
| R-squared | 0.0440 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 6. Regression (7) output

Natives – Education effect on wages

| Dependent variable = Log (wage) | (7) Native wages |
|---------------------------------------|------------------------|
| Years of education | 0.1586*** (0.0001) |
| Age | 0.1788*** (0.0001) |
| Age^2 | -0.0018*** (0.0000) |
| Male | 0.4365*** (0.0005) |
| Race (White = Benchmark group) | |
| Black | -0.2156*** (0.0009) |
| Chinese | 0.0546*** (0.0049) |
| Japanese | 0.0145*** (0.0052) |
| Other Asian or Pacific | -0.0307*** (0.0030) |
| Other Race or Combined Races | -0.1091*** (0.0013) |
| Constant | 3.8223*** (0.0182) |
| State fixed effects | Yes |
| Year fixed effects | Yes |
| Observations | 17,251,234 |
| R-squared | 0.3570 |

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

10.3 STATA Commands

Creating variables

```
-gen empstatus=0
-replace empstatus=1 if empstat==1
-label define emplabel 0 "unemployed" 1 "employed"
-label values empstatus emplabel

- gen yearseduc=0
- replace yearseduc=1 if educd==15 (Continuing with
replace up to educ==116)
- label define yearseduclabel 12 "highschool - 12" 14
"associates degree - 14" 15 "bachelor - 15" 17 "master -
17" 18 "doctoral degree - 18"
- label values yearseduc yearseduclabel

- gen speakseng=0
- replace speakseng=1 if speakeng==6 (Continuing with
replace up to speakeng==1)
- label define englabe1 0 "does not speak eng - 0" 1
"yes, but not well - 1" 2 "yes, speaks well - 2" 3 "speaks
very well - 3" 4 "speaks only -English - 4"
- label values speakseng englabe1

-gen hispanic=0
-replace hispanic=0 if hispan=1 / hispan=2 / hispan=3 / hispan=4
-label define hispaniclabel 0 "not Hispanic - 0" 1 "Hispanic - 1"
-label values hispanic hispaniclabel

- gen bplaceregion=0
- replace bplaceregion=1 if bpl>199 & bpl<300 (Continuing with replace bpl<800)
- label define bpllabel2 0 "Europe - 0" 1 "Central America - 1" 2 "South America - 2" 3 "North America - 3" 4 "Asia - 4" 5 "Africa - 5" 6
"Oceania - 6" 7 "Other - 7"
- label values bplaceregion bpllabel2

- gen raceofind=0
- replace raceofind=1 if race==2 (Continuing with replace up to race==9)
- label define racelabel 0 "White - 0" 1 "Black - 1" 2 "Chinese - 2" 3 "Japanese - 3" 4 "Other Asian or Pacific - 4" 5 "Other Race or Combined
Races - 5"
- label values raceofind racelabel

-gen gender=0
-replace gender=1 if sex==1
-label define sexlabel 0 "female" 1 "male"
-label values gender sexlabel
```

Regressions

```
reg empstatus yearseduc, r
outreg2 using "table 5.xls", replace dec(4) keep (yearseduc) label
```

```
reg empstatus yearseduc i.year i.statefip, r
outreg2 using "table 5.xls", dec (4) keep (yearseduc) label
```

```
reg empstatus yearseduc age age_2 gender yrsusa1 i.speakseng i.bplaceregion i.raceofind i.year i.statefip,r
outreg2 using "table 5.xls", dec (4) keep (yearseduc age age_2 gender yrsusa1 i.speakseng i.bplaceregion
i.raceofind), label
```

```
Gen log_wage=log(wage)
reg log_wage yearseduc age age_2 gender yrsusa1 i.speakseng i.raceofind i.bplaceregion i.statefip i.year, r
outreg2 using "regincome.xls", replace dec(4) keep (yearseduc age age_2 gender yrsusa1 i.speakseng
i.raceofind i.bplaceregion)
```


Natives

Drop if yrimmig>0

Keep if labforce==2

reg empstatus yearseduc age age_2 gender i.speakseng i.raceofind i.year i.statefip,r

reg log_wage yearseduc age age_2 gender i.speakseng i.raceofind i.year i.statefip,r

Probit regression and marginal effects

Probit empstatus yearseduc, r

Mfx

