THE ALLOCATION OF TALENT IN BRAZIL AND INDIA

Kanat Abdulla

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment of the requirements for the degree of Doctor of Philosophy in the Department of Economics

Chapel Hill
2018

Approved by:

Lutz Hendricks

Simon Alder

Luca Flabbi

Klara Peter

Toan Phan

ABSTRACT

Kanat Abdulla: The Allocation of Talent in Brazil and India
(Under the direction of Lutz Hendricks)

This dissertation is a collection of two independent essays on human capital in developing countries. In the first chapter, I investigate the labor market outcomes in Brazil and India and examine the effect of the frictions in the human capital accumulation and in the labor market on the aggregate output in these countries. The second chapter tests theories related to immigrant characteristics and their earnings by investigating immigrants in low-income countries.

ACKNOWLEDGEMENTS

I would like to express my gratitude to my advisor, Lutz Hendricks, for the continuous support throughout the research, for his motivation and immense expertise that greatly assisted the research. I am also grateful to Simon Alder, Luca Flabbi, Klara Peter, and Toan Phan for their valuable suggestions and comments that helped me improve the results of my work.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

THE ALLOCATION OF TALENT IN BRAZIL AND INDIA

1.1 Introduction

Hsieh et al. (2013) ask whether improved allocation of workers according to their talents was an important source of productivity growth in the U.S. This is motivated by substantial differences in occupational choices between men/women and blacks/whites. In particular, they document that the share of women and blacks in high-skill occupations was very low relative to that of white men. These differences in occupational distribution of women and blacks relative to white men declined over time, suggesting that misallocation has diminished. This change has positively affected the aggregate productivity growth in the United States. Hsieh et al. (2013) argue that better allocation of talent explains 15–20% of the economic growth in the U.S.

I use micro-level survey data from Brazil and India with detailed information on socio-economic and occupational characteristics and their earnings to investigate the role of allocation of talent in economic development of these countries. The analysis in these countries is motivated by the fact that there are substantial fractions of the population that are disadvantaged in terms of access to quality education and jobs. As a result there are large differences in occupational distribution and earnings between groups. This paper will argue that the allocation of talent affects the aggregate output in these countries.

Allocation of talent refers to the distribution of various groups across occupations, where the groups are categorized by race[1] and gender. Talent is misallocated when there is a difference in the occupational distribution between the groups. The main forces

[1] In India the groups are categorized by castes.

that produce the difference in occupational distribution across groups are frictions in accumulation of human capital and frictions in labor market. Frictions are estimated from the observed occupational distributions and the wage gap between groups. Given these frictions, workers choose occupations where they have the highest utility. An augmented Roy model of occupational choice developed by Hsieh et al. (2013) allows me to determine the potential gains to output from decreasing frictions in Brazil and India.

I investigate the occupational distribution and wage gaps of four groups (white men, white women, brown men and brown women) in Brazil and four groups (other men, other women, scheduled caste/tribe men, scheduled caste/tribe women) in India. The term "brown" is used to refer to Brazilians of mixed ethnic ancestries and sometimes known as "parda" in the Brazilian censuses. Browns make up 43% of the Brazilian population. Scheduled caste and tribe are terms recognized by the Indian constitution and refer to the most disadvantaged groups in India. They consist of 26% of the country's population.

I show that frictions are substantial, especially for brown women in Brazil and scheduled caste women in India. I conduct a counterfactual experiment which helps me to assess the role of misallocation of talent in productivity in these countries. First, I reduce frictions faced by the groups by half. I find that reducing frictions faced by various groups in Brazil and India by half increases the aggregate productivity by 10–20% in Brazil and by 14–22% in India. Second, I investigate the gain after eliminating frictions in these countries. Removing frictions increases the aggregate productivity of the countries by 21–42% in Brazil and by 36–46% in India.

This paper is organized as follows: Section 1.2 reviews the literature, Section 1.3 describes the census data obtained from the Integrated Public Use Microdata Series, Section 1.4 discusses the model, Section 1.5 provides an empirical evidence on earnings of various groups and their occupational distribution, Section 1.6 describes the results of the model and provides robustness checks, and Section 1.7 concludes.

1.2 Literature review

In most countries there are disadvantaged groups within population. They are disadvantaged because they face discrimination in the early stages of acquisition of human capital or later face unequal access to jobs or both. Brazil and India are among those countries.

There is significant evidence that in Brazil there is a gap in earnings between men and women and race groups. Men in Brazil earn about 25% more and are more likely to participate in the labor force than women in Brazil (Arabsheibani et al. (2003)). White people in Brazil earn 26% more than brown people with same human capital and labor market characteristics (Telles (2006)). A significant part of the racial wage gap in Brazil occurs because of discrimination (Lovell (1993)). The analysis of the returns to schooling for various groups shows that the returns to schooling for whites are higher than the returns to schooling for dark-skinned population (Loureiro et al. (2004)). The difference in occupational distribution between men and women in Brazil has an effect on wage gap between these groups (Madalozzo R. (2010)).

Caste- and gender-based discrimination in India produces significant gaps in terms of earnings and labor market participation. Scheduled caste and scheduled tribe workers earn 30% less than equally qualified others (Madheswaran and Attewell (2007)). The unconditional earnings gap of women relative to men in India was 55% in 1999–2000 and 49% in 2009–2010, and the gap persists even within the same education level and within most occupations and industries (Deshpande et al. (2018)). Occupational discrimination is more prevalent than wage discrimination. Some castes are discriminated against in terms of unequal access to jobs, especially in the private sector (Madheswaran and Attewell (2007)). Discrimination in hiring processes is a common practice in the urban labor market in India (Thorat and Attewell (2007)).

The observed gaps in earnings and unequal access to jobs force individuals from certain groups out of occupations for which they have necessary skills. This is called talent misallocation. Whether or not this misallocation has an effect on overall productivity has been the focus of a number of studies that have contributed to the understanding of the role of talent misallocation in economic development. One of the important factors in the allocation of talent is the relative rewards that different professions receive (Acemoglu (1995)). Rewards for entrepreneurship determine the allocation of productive versus unproductive entrepreneurship labor, which affects the aggregate output (Baumol (1990)). By analyzing the occupational distribution of women and blacks relative to white men in the period from 1960 to 2008, Hsieh et al. (2013) find that the share of women and blacks in high-skill occupations was very low relative to that of white men in the 1960s. This occupational gap shrank over time, affecting aggregate productivity growth in the United States. In particular, Hsieh et al. (2013) argue that better allocation of talent explains 15–20% of the economic growth in the U.S.

1.3 Data

I use Brazilian and Indian survey data available at Integrated Public Use Microdata Series (IPUMS). The Brazilian data spans the 1991, 2000, and 2010 survey years with a total sample size of about 5–10 million individuals per survey year. Indian data is a socio-economic survey conducted by the National Sample Survey Organization of India every 5–6 years with a sample size of 500–600 thousand individuals. The variable names, coding schemes, and documentation are consistent for most samples.

The analysis uses a variable from IPUMS that indicates an individual's primary occupation, which is classified according to the system used by the respective census of countries. Brazilian and Indian surveys have different classification systems for occupations. Moreover, the Brazilian survey has varying classifications for different periods. To make data comparable across years and countries, I harmonize occupational coding to the 1990 Census occupational classification system used by Hsieh et al. (2013) and aggregate to 66 occupations.[2] Some related occupation categories were merged into one sub-heading. For instance, management-related occupations include some administrative support occupations, and the computer and communications equipment operator occupation consists of communication equipment operators and computer and peripheral equipment operators.

Other key variables used in the analysis are variables indicating an individual's earnings, hours worked, employment status, education, and race. For individual's earnings I use a variable that represents the total income from the labor (from wages, a business, or a farm) in the previous month or year.[3] A variable that indicates individual's social group or race in Brazilian census is named as "race" and in Indian census as "social group". Employment status of the person is defined by Emptat, which I use to identify employed individuals. Hrswork[4] shows a person's hours worked per week, and wkswork[4] shows person's weeks worked per year, which are used to compute hourly wages. A person's educational attainment is identified by the variable "edattaind" and shows the person's educational attainment in terms of the level of schooling completed, i.e. a person attending the final year of college receives the code for having completed secondary degree only. From this variable I construct a variable that indicates a person's number of schooling years completed, "educ". There is a limitation in constructing years of schooling from edattaind because it will show only the approximate number of years of schooling. For example, there is a discontinuity between 8 and 12 years of schooling, and it will not allow me to identify individuals with more than 16 years of schooling.

From the available data I construct hourly wages and experience. Hourly wages are constructed from income, weeks worked per year and hours worked per week. Experience is constructed from individual's age and years of schooling completed as age minus schooling minus 6. As I discussed previously, there is a problem in the recording years of

[2]The detailed occupational coding is provided in Appendix.

[3]The variable available in IPUMS for Brazil and India is called "incearn". I also use data from the US in analysis. For the US sample I use "incbus", income from business, "incwage", wage income, and "incfarm", income from farming.

[4]The variables "wkswork" and "hrswork" are available only for Brazil and US.

schooling correctly for some observations, which leads to difficulty in recording the potential experience for some observations. It may overstate the actual potential experience if actual years of schooling is higher and understate if the actual years of schooling is lower. Summary statistics for key variables used in the analysis are provided in Appendix Table A3.

1.3.1 Data from Brazilian household survey

The data are analyzed for the following sample periods: 1991, 2000, and 2010. The following restrictions are made to the data: 1) only brown[5] and white are chosen out of 5 possible race groups, 2) the analysis is restricted to individuals whose ages are between 25 and 60, 3) individuals who are on active military duty and unemployed individuals are excluded, 4) individuals who are unable to work due to disability, retired or at school are also excluded from the sample.

Table 1.1 reports the summary statistics of the restricted sample across years. The number of observations, as shown in the table, has increased considerably over time, with the sample size increasing two-fold from 969,000 observations to 1,530,000 observations over the 20-year period. The largest share of the sample belong to whites: the share of whites was 59% in 1991 and 55% in 2010. The shares of race groups have not changed much over the course of the period. White males and females constituted 28% and 30% of the population and 26% and 28% of the population in 1991 and 2010, respectively. The proportions of brown men and brown females have slightly increased from 21% to 23%, respectively, over the period. The education levels of these population have changed significantly over the 20-year period. Table A4 in Appendix reports the share of college-educated individuals by groups and survey years. In 1991 the share of college-educated individuals was only 5.5% of the total population, but in 2010 it had increased to 10.4%.

[5]In the analysis I use only brown and white because these races constitute the largest share of the population, and the US has also two largest race groups, namely black and white, which makes comparison with the US easier.

Among the race groups, brown men had the fewest college-educated individuals: in 1991 the share of college-educated brown men was only 1.8% of the total brown men, in 2010 it had increased only to 3.8%. White women in the sample show the highest increase in the share of college-educated individuals. In the total population the share of college-educated white women was 7.7% in 1991; by 2010 it had increased to 17.1%.

	1991	2000	2010
Sample size	969,833	1,204,718	1,531,081
white men	28%	29%	26%
white women	30%	32%	28%
brown men	21%	20%	23%
brown women	21%	20%	23%

Table 1.1: Sample statistics (Brazilian survey)

1.3.2 Data from Indian household survey

For India, IPUMS provides consistent data for the following sample periods: 1993, 1999, and 2004. There is a lack of data comparability across different survey periods in regard to caste identities. In the 1999 and 2004 surveys, other backward castes are treated separately; however prior to 1999 other backward castes and others were treated as one group. For the purposes of comparability across different periods, I treat other backward castes and others as one group in the 1999 and 2004 sample periods.

The following restrictions are made to the data: 1) the analysis is restricted to individuals aged 25–60, 2) individuals who are on active military duty and unemployed individuals are excluded, 3) individuals who are unable to work due to disability, retired, or at school are also excluded from the sample.

There are four main caste classifications: scheduled caste, scheduled tribe, other backward castes, and others. The most disadvantaged castes in socio-economic terms are scheduled castes and scheduled tribes. Table 1.2 reports the summary statistics of the samples. The sample size does not change much across periods: 244,514, 256,948 and 269,067 in 1993, 1999 and 2004, respectively. The majority of the Indian population in the sample are "others" with 72–75% of the total population. Scheduled tribes and scheduled castes are minority groups of the Indian population. They respectively comprise 13% and 16% of India's total population in the sample. The share of castes did not show appreciable change over time.

The majority of the Indian working population have education levels less than college degree. As shown in Table A4, only 6.8% and 7.8% of the total population had a college degree in 1999 and 2004, respectively. There is diversity in terms of education attainment among gender and castes. While 11.5% of other men had a college degree in 1999, the share of college-educated other women was only 5.5%. This is also true for other castes. Only 1.2 and 0.5% of scheduled tribe and scheduled caste women had college degrees in 1991 as compared to 3.5% and 2.5% of men from respective castes. We see an increase in college attainment for all castes. Overtime the groups experienced increase in the share of college-educated individuals. In particular, scheduled tribe and scheduled caste men show a noticeable increase in the share of college educated individuals, from 3.5% to 7% and from 2.5% to 5.1%, respectively, in 1993 and 2004.

	1993	1999	2004
Sample size	244,514	256,948	269,067
Other men	38%	37%	35%
Other women	38%	37%	36%
Scheduled tribe men	5%	5%	6%
Scheduled tribe women	5%	6%	7%
Scheduled caste men	7%	8%	8%
Scheduled caste women	7%	8%	8%

Table 1.2: Sample statistics (Indian survey)

1.3.3 Home sector and sample selection

A substantial part of the working population in developing countries is occupied in the informal sector. Taking into account this sector will greatly influence the results. IPUMS provides information about the employment status of individuals in the sample. Individuals not in the labor force are classified as being in housework, unable to work, at school, or retired and living on rents. Table A5 in Appendix shows the observation numbers in each category. The sample excludes individuals who are at school, unable to work or retired. So individuals in labor force and individuals not in labor force but those in housework are in the sample. As can be seen from Table 1.3, approximately 1/3 of the working age population in Brazil and India are classified as employed in housework. Most of the population occupied in housework in both countries are women: 40.8% of women in Brazil in 2010 and 60.5% of women in India in 2004 were classified as working in housework. For men this number is much lower: 15.7% in Brazil and 3.8% in India in the corresponding years.

		1991	2000	2010
	Total	964,173	1,204,520	1,530,715
All	Employed	71%	70%	72%
	Housework	29%	30%	28%
Men	Employed	99.8%	88%	84%
	Housework	0.2%	11.7%	15.7%
Women	Employed	42.7%	51.8%	59.2%
	Housework	57.3%	48.2%	40.8%

(a) Brazil

		1993	1999	2004
	Total	251,690	266,404	275,405
All	Employed	66%	65%	68%
	Housework	34%	35%	32%
Men	Employed	96.2%	95%	96%
	Housework	3.8%	4.6%	3.8%
Women	Employed	36.0%	34.5%	39.5%
	Housework	64.0%	65.5%	60.5%

(b) India

Table 1.3: Sample data

In addition to the 66 occupation categories defined above, I create another occupational category for the home sector. An individual who is not in the labor force is considered to be working in the home sector. I impute wages for individuals in the home sector by assigning them the predicted wages of people in the market sector with the same observed characteristics. The observed characteristics include the region where an individual resides, the group to which individual belongs, schooling, and experience. Here I assume that the relationship between earnings and these characteristics are the same for the home and the market sectors.

Estimating the wage equation for individuals who are employed may not produce similar results to estimating it for the population as a whole. Those who are employed are the ones who made the decision to work, but this decision may not have been made randomly. If the ones who choose to work tend to have higher (lower) wages than those not in the labor force, then the sample of observed wages will be biased upward (downward). Thus this produces a biased result when estimating the returns to observable characteristics like education or experience. To assign wages to workers in the home sector, I implement selection bias correction, following Heckman (1979).

Thus the following model is analyzed:

$$\begin{aligned} log(wage) = \beta_1 + \beta_2 group + \beta_3 educ + \beta_4 exp \\ + \beta_5 exp^2 + \beta_6 year + \beta_7 region + \beta_8 marst + \varepsilon_1 \end{aligned} \tag{1.1}$$

and the earnings are observed if

$$\begin{aligned} \gamma_1 + \gamma_2 group + \gamma_3 educ + \gamma_4 exp + \gamma_5 exp^2 \\ + \gamma_6 year + \gamma_7 region + \gamma_8 marst + \gamma_9 numperson + \varepsilon_2 > 0 \end{aligned} \tag{1.2}$$

I assume that X_i includes education and experience, dummy variables for groups, census region, year, and marital status, and Z_i includes variables in X_i plus the number of people in the household.[6] The model is estimated on women. Using the estimated unbiased coefficients, I predict the earnings for women in the home sector.

[6]The studies use the number of children as an exclusion restriction (e.g. Mulligan and Rubinstein (2008)). For Brazil and India this variable is not available.

1.4 The Model

I use an augmented Roy (1951) model presented by Hsieh et al. (2013). There are an infinite number of individuals and a representative firm. Individuals consume goods, rent labor to maximize their utilities, and choose occupation that deliver the highest utility. A firm hires labor inputs and produces goods.

Demographics: There is a continuum of people, each belonging to a group g based on gender and race.

Preferences: Individuals maximize their utility:

$$U_{ig} = c_{ig}^{\beta}(1 - s_{ig}) \tag{1.3}$$

where i refers to occupation, c_{ig} is consumption, s_{ig} schooling, and β is a parameter showing the tradeoff between consumption and leisure.

Endowments: At birth, individuals are endowed with a random skill ϵ_i from a extreme value distribution as in McFadden (1974) and Eaton and Kortum (2002).

$$F_g(\epsilon_1, ..., \epsilon_N) = exp\{-[\sum_{i=1}^{N}(T_{ig}\epsilon_i^{-\theta})]^{1-\rho}\} \tag{1.4}$$

where θ determines the skill dispersion, ρ determines the correlation of skills across occupations, and T_{ig} defines occupation-group specific ability.

Technology: An individual accumulates human capital from education s and expenditure e according to the production function:

$$h(e, s) = \bar{h}_{ig} s_{ig}^{\phi_i} e_{ig}^{\eta} \tag{1.5}$$

The production function varies by group. The elasticity of human capital with respect to schooling, ϕ_i, differs by occupation. The parameter $\bar{h}_{ig}$, efficiency in human capital, differs across groups and occupations, which allows for differences in health and family background.

Markets: There is a market for labor rental.

1.4.1 Household Problem

Households maximize their utility by choosing consumption, schooling, and expenditure on goods:

$$U(\tau_{ig}^{w}, \tau_{ig}^{h}, \bar{h}_{ig}, w_i, \epsilon_i) = \underset{c,e,s}{max}\,(1 - s_{ig})c_{ig}^{\beta} \tag{1.6}$$

s.t.

$$c_{ig} = (1 - \tau_{ig}^{w})w_i\epsilon_i h(e_{ig}, s_{ig})e_{ig}(1 + \tau_{ig}^{h}) \tag{1.7}$$

Budget constraint relates consumption to income and expenditure. w_i is the wage per efficiency unit of labor paid by the firm, and ϵ_i is an idiosyncratic talent draw in the worker's chosen occupation. There are two additional variables: τ_{ig}^{h}, friction on accumulation of human capital, and τ_{ig}^{w} , friction in labor market. τ_{ig}^{h} acts like a tax on expenditure on human capital and τ_{ig}^{w} acts like a tax on wages in the labor market.

Household solution is $\{c_{ig}^{*}, e_{ig}^{*}, s_{ig}^{*}\}$ and U_{ig} that satisfy:

$$s_i^{*} = \frac{1}{1 + \frac{1-\eta}{\beta\phi_i}} \tag{1.8}$$

$$e_{ig}^{*} = (\frac{\eta w_i s_i^{\phi_i}\epsilon_i}{\tau_{ig}})^{\frac{1}{1-\eta}} \tag{1.9}$$

$$c_{ig}^{*} = \bar{\eta}(\frac{w_i s_i^{\phi_i}\epsilon_i}{\tau_{ig}})^{\frac{1}{1-\eta}} \tag{1.10}$$

$$U(\tau_{ig}, w_i, \epsilon_i) = (\frac{w_i s_i^{\phi_i}(1 - s_i)^{\frac{1-\eta}{\beta}}\epsilon_i\eta^{\eta}(1-\eta)^{1-\eta}}{\tau_{ig}})^{\frac{\beta}{1-\eta}} \tag{1.11}$$

Here, τ_{ig} summarizes the frictions such that:

$$\tau_{ig} = \frac{(1+\tau_{ig}^h)^\eta}{1-\tau_{ig}^w} \times \frac{1}{\bar{h}_{ig}} \tag{1.12}$$

Occupational Sorting

Given skills, an individual will choose the occupation that yields the highest value of U_{ig} in equation (1.11). By aggregating the optimal occupation choices for all people, we arrive at the following equation, which is the overall occupational share of a group g[7]:

$$p_{ig} = \frac{\tilde{w}_{ig}^\theta}{\sum_{s=1}^N \tilde{w}_{sg}^\theta} \tag{1.13}$$

where $\tilde{w}_{ig} = \frac{T_{ig}^{1/\theta} w_i s_i^{\phi_i} (1-s_i)^{\frac{1-\eta}{\beta}}}{\tau_{ig}}$ and p_{ig} is the fraction of people in group g that work in occupation i. Equation (1.13) says that the occupational sorting depends on $\tilde{w}_{ig}$, which is the overall reward that someone from group g working in occupation i who has mean talents receives, relative to the power mean of $\tilde{w}$ for the group over all occupations. This means that the occupational distribution is driven by the relative reward, not the absolute reward, for working in an occupation. This sorting model generates an equation for average quality of workers in a given group working in a given occupation:

$$E[h_{ig}\epsilon_i] = \gamma[\eta^\eta s_i^{\phi_i} (\frac{w_i(1-\tau_{ig}^w)}{1+\tau_{ig}^h})^\eta (\frac{T_{ig}}{p_{ig}})^{\frac{1}{\theta}}]^{\frac{1}{1-\eta}} \tag{1.14}$$

where $\gamma = \Gamma(1 - \frac{1}{\theta(1-\rho)} \frac{1}{1-\eta})$ is related to the mean of the Frechet distribution for abilities. The average quality of worker in a group g and occupation i is inversely related to the share of that group in that occupation. This means that if the share of a group is small in a certain occupation, the workers representing that group working in that occupation will be of a higher quality on average than the workers representing other groups working in

[7] The derivation of the result can be found in Hsieh et al. (2013)

the same occupation. This can be explained by the fact that if a group faces high barriers in a certain occupation, the people from that group who succeed in that occupation must be highly skilled. Given that we have average quality of workers we can derive the average wage for a given group in a given occupation:

$$\bar{w}_{ig} = (1-\tau^w_{ig})w_i E[h_{ig}\epsilon_i] = (1-s_i)^{-1/\beta}\gamma\bar{\eta}(\sum_s^N \tilde{w}^\theta_{sg})^{\frac{1}{\theta}\frac{1}{1-\eta}} \tag{1.15}$$

where $\bar{w}_{ig}$ is the average earnings in occupation i by group g and $\bar{\eta} = \eta^{\frac{\eta}{1-\eta}}$.
The occupational wage gap between any two groups is given by:

$$\frac{\bar{w}_{ig}}{\bar{w}_{ig'}} = (\frac{\sum_s \tilde{w}^\theta_{sg}}{\sum_s \tilde{w}^\theta_{sg'}})^{\frac{1}{\theta}\frac{1}{1-\eta}} \tag{1.16}$$

Equation (1.16) shows that the wage gap between group g and group g', $\frac{\bar{w}_{ig}}{\bar{w}_{ig'}}$, is independent of occupations. Combining equation (1.13) and equation (1.16), we get the propensity of a group g to work in an occupation relative to group g':

$$(\frac{p_{ig}}{p_{ig'}}) = \frac{T_{ig'}}{T_{ig}}(\frac{\tau_{ig}}{\tau_{ig'}})^{-\theta}(\frac{\bar{w}_g}{\bar{w}_{g'}})^{-\theta(1-\eta)} \tag{1.17}$$

where $\bar{w}_g = (\sum_i^N \tilde{w}^\theta_{ig})^{\frac{1}{\theta}\frac{1}{1-\eta}-1}\sum_i^N \tilde{w}^\theta_{ig}(1-s_i)^{-\frac{1}{\beta}}\gamma\bar{\eta}$ is the average wage of the group. From equation (1.17) we can see that the propensity for a member of a group g to work in an occupation i compared to group g' is affected by three factors: the relative mean talent $\frac{T_{ig'}}{T_{ig}}$, the relative frictions $\frac{\tau_{ig}}{\tau_{ig'}}$, and the wage gap $\frac{\bar{w}_g}{\bar{w}_{g'}}$. The propensity for a group to work in an occupation is increasing in relative mean talent and decreasing in relative frictions and the relative wage gap.

1.4.2 Firm problem

A representative firm produces aggregate output Y from labor in N different occupations by hiring H_i, total efficiency labor units, in each occupation and taking A_i, exogenous productivity in occupation i, as given in order to maximize profits:

$$\max_{H_i} (Y - \sum_{i=1}^{N} w_i H_i) \tag{1.18}$$

where aggregate output, Y, is given by:

$$Y = (\sum_{i=1}^{N} (A_i H_i)^{\frac{\sigma-1}{\sigma}})^{\frac{\sigma}{\sigma-1}} \tag{1.19}$$

Firm solution:

$$H_i^{demand} = (\frac{A_i^{\frac{\sigma-1}{\sigma}}}{w_i})^{\sigma} Y \tag{1.20}$$

1.4.3 Market clearing

Wage per efficiency unit of labor, w_i, clears the labor market in each occupation:

$$H_i^{demand} = H_i^{supply} \tag{1.21}$$

where H_i^{supply}, aggregate supply is given by:

$$\begin{aligned} H_i^{supply} &= \sum_g q_g p_{ig} E[h_{ig}\epsilon_i] \\ &= \gamma\bar{\eta} w_i^{\theta-1}(1-s_i)^{(\theta(1-\eta)-1)/\beta} s_i^{\theta\phi_i} \sum_g q_g T_{ig} \frac{(1-\tau_{ig}^w)^{\theta-1}}{(1+\tau_{ig}^h)^{\eta\theta}} (\sum_{i=1}^{N} \tilde{w}_{sg}^{\theta})^{\frac{1}{\theta}\frac{1}{1-\eta}-1} \end{aligned} \tag{1.22}$$

where q_g is the total number of people in group g.

1.4.4 General equilibrium

General equilibrium consists of $\{p_{ig}, H_i^{supply}, H_i^{demand}, w_i\}$ and Y that:

1. p_{ig} satisfies equation 1.13;
2. H_i^{supply} satisfies equation 1.22;
3. H_i^{demand} satisfies equation 1.20;

4. w_i satisfies equation 1.21;
5. Y satisfies equation 1.19.

1.5 Empirical findings

As the model predicts in equation (1.17), frictions faced by each group can be derived from the wage gap and occupational distribution of the group relative to the privileged group. Here I estimate wage gaps between groups relative to the privileged group and occupational distribution of the groups. With the available information on these variables, I compute the frictions faced by each group in each occupation.

1.5.1 Occupational distribution across groups

I define four groups for Brazil: white women, white men, brown men, and brown women; and four groups for India: other men, other women, scheduled caste (SC) men, and scheduled caste (SC) women. I assume that white men in Brazil and other men in India face less frictions than other groups in these countries. This is a reasonable assumption based on occupational distributions that I will show below. Later wage gap estimations will also show that white men in Brazil and other men in India earn more than other groups with similar characteristics. Figure 1.1 shows the share of each group in highly skilled occupations[8] in 2010 for Brazil and in 2004 in India. From the figure we see that white men in Brazil and other men in India are more likely to work in highly skilled occupations. The most disadvantaged groups in terms of shares in these occupations are brown men and women in Brazil and other and scheduled caste women in India. All these groups are less likely than the privileged group to work as executives, architects, engineers, mathematicians, doctors, and lawyers.

[8] Highly skilled occupations are executives, architects, engineers, mathematicians, lawyers, and judges.

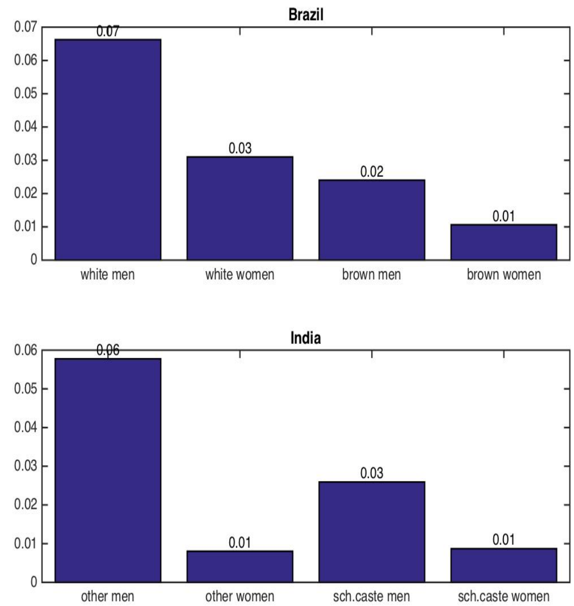


Figure 1.1: Share of groups in highly skilled occupations

Here I show another way of looking at occupational distributions across groups. I compute an index (Occupational similarity index[9]) that will show the similarity of occupational distributions of groups with respect to white men in Brazil and other men in India. The formula below captures the similarity in occupational distribution across groups relative to the privileged group:

$$\Psi_g = 1 - \frac{1}{2}\sum_{i=1}^{N} |p_{i,wm} - p_{i,g}| \tag{1.23}$$

Ψ_g is defined as the sum across occupations of the absolute value of the difference in the propensity of group g relative to white men.[10] The index shows the degree of difference in occupational distribution between groups. An index value of zero implies that the occupational distribution of the group is not similar to that of white men. A detailed distribution of the index across groups is presented in Table 1.4. Panels A, B,

[9] I borrow the index from Hsieh et al. (2013).

[10] Other men in India

and C present the occupational similarity index of the groups relative to the privileged group in Brazil, India, and the US, respectively. The value of 0.32 for white women in 1991 indicates that the occupational distribution of white women is not similar to that of white men. The value of 0.82 for brown men in 1991 shows that brown men were closer to white men in occupational distribution. As can be seen from the table, there is a slight increase in the index for all groups in Brazil. The index increased from 0.32, 0.82, and 0.28 in 1991 to 0.45, 0.85, and 0.37 in 2010 for white women, brown men and brown women, respectively.

The occupational distribution analysis for India shows a slightly different picture than for Brazil. We do not see as high an index value as for brown men in Brazil. The closest in occupational distribution to other men in India is scheduled caste men with 0.77 in 1993 and 0.81 in 2004. The most disadvantaged in terms of occupational distribution are other women, with the index of 0.33 in 1993 and 0.37 in 2004. Also the scheduled caste women did not experience any convergence in occupational distribution relative to other men, and their index value remained at 0.49.

Panel C of Table 1.4 shows occupational distribution of the groups relative to white men in the US. We can see that, as in Brazil and India, women in the US have less similar occupations than men. Women in the US have an occupational distribution closer to that of white men than do women in other countries. Especially, it is seen in 2010, the occupational similarity indexes for white and black women are 0.54 and 0.52 versus 0.45 and 0.37 for white and brown women in Brazil, and 0.37 and 0.49 for other and scheduled caste women in India. Black men in the US are less likely to work in similar occupations to those of the privileged group than are men in India and Brazil. In 2010 the similarity index for black men in the US was 0.73, whereas the indexes for brown men in 2010 and scheduled caste men in 2004 respectively were 0.85 and 0.81.

Panel A: Relative to white men in Brazil			
	1991	2000	2010
white women	0.32	0.36	0.45
brown men	0.82	0.83	0.85
brown women	0.27	0.31	0.37
Panel B: Relative to other men in India			
	1993	1999	2004
other women	0.33	0.31	0.37
scheduled caste men	0.77	0.76	0.81
scheduled caste women	0.49	0.49	0.49
Panel C: Relative to white men in the US			
	1990	2000	2010
white women	0.48	0.53	0.54
black men	0.71	0.72	0.73
black women	0.44	0.5	0.52

Table 1.4: Occupational similarity index

1.5.2 Wage gap estimations

As we saw in the previous section, there is a difference in occupational distribution between groups. Next, I examine if there are differences in wages between groups, their magnitudes, and if they change over time. From the available data on wages across occupations, I estimate the wage gaps of the groups relative to white men in Brazil and to other men in India. The general functional form of log wages can be summarized by the following equation:

$$\begin{aligned} log(wage_i) = \alpha + \textstyle\sum_g \beta_{1g} G_{ig} + \beta_2 Educ_i + \beta_3 Exp_i \\ + \beta_4 Exp_i^2 + \beta_5 Exp_i^3 + \beta_6 Exp_i^4 + \textstyle\sum_k \beta_{7k} O_{ik} + \epsilon_i \end{aligned} \tag{1.24}$$

where

wage - wage per hour;

G_g - dummy representing groups;

Exp_i - experience;

$Educ_i$ - years of schooling;

O_k - dummy referring to occupations.

Table 1.5 reports group dummies estimated using the equation 1.24. The value of -0.29 for white women in Brazil indicates that white women earned 0.29 log points less than Brazilian white men in 1991. Brown women face the highest disadvantage relative to white men in terms of wages, with a 0.51 log difference in 1991 and 0.46 in 2010. Wage gaps for brown men were -0.22, -0.25, and -0.19 in 1991, 2000, and 2010, respectively. During the 1991–2010 period white and brown women experienced 0.03 and 0.05 log points wage convergence, respectively.

Panel B reports the estimations of wage gaps relative to other men for groups in India. Wage gaps relative to other men faced by other women, scheduled caste men, and scheduled caste women in 1993 were 0.34, 0.21, and 0.49 log points, respectively. There was no noticeable change in wage gap over time.

Wage gaps of the groups relative to white men in the US show that women earn less than men with similar characteristics. The wage gaps of the white women and black men are closer to those of white women and brown men in Brazil. In 2010 white women and black men earned 0.24 and 0.12 log points lower than white men, whereas white women and brown men earned 0.26 and 0.19 log points lower than white men in Brazil. The earnings of black women in the US are closer to those of white men than the earnings of brown and scheduled caste women in Brazil and India, respectively.

Panel A: Relative to white men in Brazil			
	1991	2000	2010
white women	-0.29	-0.29	-0.26
	0.006*	0.005	0.005
brown men	-0.22	-0.25	-0.19
	0.005	0.005	0.005
brown women	-0.51	-0.50	-0.46
	0.006	0.006	0.005
Panel B: Relative to other men in India			
	1993	1999	2004
other women	-0.34	-0.33	-0.31
	0.003	0.003	0.003
scheduled caste men	-0.21	-0.19	-0.21
	0.004	0.003	0.003
scheduled caste women	-0.49	-0.49	-0.48
	0.004	0.004	0.003
Panel C: Relative to white men in the US			
	1990	2000	2010
white women	-0.32	-0.28	-0.24
	0.005	0.005	0.005
black men	-0.11	-0.13	-0.12
	0.009	0.009	0.009
black women	-0.28	-0.25	-0.26
	0.009	0.008	0.008

*standard errors

Table 1.5: Conditional log difference in wages

The model predicts that wage gaps are the same for all occupations (1.16) and independent of propensities. This means that changes in frictions faced by a group in one occupation, resulting in a change of relative propensities, does not affect the average wage of the group, because an increase (a decrease) of a friction will attract (deter) less qualified workers, thus lowering (increasing) the average quality of the group. Table 1.6 shows the results of the regression of the occupational wage gap and relative propensities. The regression was weighted by the share of the workers in the groups across occupations. As can be seen from the table, the slope and the R^2 from the regression of the wage gap on propensities are small for all three countries, which is an indication that there is little to no correlation between these variables, which supports the model version of the equation.

	1991			2000			2010		
	slope	st_dev	R^2	slope	st_dev	R^2	slope	st_dev	R^2
white women	-0.021	0.020	0.016	-0.023	0.020	0.019	0.021	0.015	0.031
brown men	0.009	0.026	0.002	0.026	0.031	0.010	0.015	0.022	0.007
brown women	-0.041	0.022	0.052	0.010	0.021	0.004	-0.003	0.019	0.000

(a) Brazil

	1993			1999			2004		
	slope	st_dev	R^2	slope	st_dev	R^2	slope	st_dev	R^2
other women	-0.014	0.025	0.005	-0.035	0.018	0.059	-0.018	0.021	0.011
sc.caste men	0.059	0.036	0.040	0.008	0.034	0.001	-0.026	0.034	0.009
sc.caste women	-0.014	0.028	0.004	-0.016	0.023	0.008	-0.030	0.026	0.019

(b) India

	1990			2000			2010		
	slope	st_dev	R^2	slope	st_dev	R^2	slope	st_dev	R^2
white women	0.005	0.024	0.001	-0.008	0.020	0.003	0.006	0.020	0.001
black men	0.032	0.034	0.065	0.036	0.026	0.066	0.007	0.021	0.002
black women	0.020	0.026	0.009	0.021	0.021	0.015	0.001	0.027	0.000

(c) USA

Table 1.6: Relationship of wage gaps and propensities

1.5.3 Estimation of Frictions

From the available data on the fraction of people in group g who work in occupation i (p_{ig}) and the wage of group g relative to privileged group ($\bar{w}_g/\bar{w}_{g'}$), I can estimate the relative frictions faced by groups in Brazil and India. So, by rearranging equation (1.17), I arrive at the following estimate of the composite friction $\hat{\tau}_{ig}$ for each group in each occupation:

$$\hat{\tau}_{ig} = \frac{\tau_{ig}}{\tau_{iwm}}(\frac{T_{iwm}}{T_{ig}})^{\frac{1}{\theta}} = (\frac{p_{ig}}{p_{iwm}})^{-\frac{1}{\theta}}(\frac{\bar{w}_g}{\bar{w}_{wm}})^{-(1-\eta)} \tag{1.25}$$

$\hat{\tau}_{ig}$ is called a composite friction because it is a function of both relative friction $\frac{\tau_{ig}}{\tau_{iwm}}$ and relative mean talent ($\frac{T_{iwm}}{T_{ig}}$). The composite friction will be high either because the relative propensity of the group $\frac{p_{ig}}{p_{iwm}}$ is low (the group is underrepresented in this occupation) or the group faces a low wage gap $\frac{\bar{w}_g}{\bar{w}_{wm}}$. The right-hand side of the equation is observed in the data, so we can use it to determine $\hat{\tau}_{ig}$ faced by each group in each occupation. The calculation of the friction by using the formula requires the estimates of θ (the parameter that governs the dispersion of talent) and η (the elasticity of human capital with respect to expenditure on human capital). I use the baseline parameter estimates from Hsieh et al. (2013) and conduct robustness checks later. The baseline parameter values are given in Table 1.7. With the baseline parameter value for θ equal to 3.44, and baseline parameter value for η equal to 0.25, I compute composite frictions.

	Parameter	Value
Elasticity of substitution	σ	3
Skill dispersion parameter	θ	3.44
Elasticity of human capital	η	0.25
Parameter in the utility	β	0.693

Table 1.7: Baseline parameter values

Table 1.8 shows the estimates of the mean and standard deviation of $\hat{\tau}_{ig}$ faced by the groups in all periods for Brazil, India, and the US. A value of the friction equal to one means a group faces no frictions relative to a privileged group. If the value is more than 1 then a group faces a friction, while a value less than 1 acts like a subsidy for that group in that occupation.

In Brazil the highest frictions are faced by brown women, the average friction for this group is 2.41 in 1991. The variance of frictions for the group is also the highest: in 1991 the standard deviation was 1.13. Over twenty years, the friction experienced by brown women in Brazil decreased: in 2010 the mean and standard deviation are 1.99 and 0.70, respectively. Of the three groups in Brazil, brown men face the least frictions. In 1991 the average friction for this group was 1.31, which only decreased by 0.10 to 1.20 in twenty years. The standard deviation of the frictions considerably decreased over the period from 0.25 to 0.14. The variance of frictions faced by white and brown women shows that frictions for these groups are highly dispersed across occupations. As shown by equation 1.17, dispersion of frictions across occupations causes misallocation of the talent.

For India we see that the frictions are the highest for scheduled caste women and other women. The average frictions are 2.24 and 2.79 in 1993, respectively for other and scheduled caste women. In 2004 these decreased slightly to 2.09 and 2.64, respectively. Scheduled caste women also face the higher dispersion of frictions than do other women. The dispersion is 1.26 for scheduled caste women versus 0.80 for other women in 1993, and these did not change much over the period. The lowest friction is faced by scheduled caste men: the average friction for the group is 1.34 and 1.26, in 1993 and 2004, respectively. The magnitude of frictions faced by the scheduled caste men in India is comparable to the that of brown men in Brazil.

The frictions faced by the groups in the US are lower than those of the groups in India and Brazil. The frictions faced by women are higher than the frictions faced by black men. Black women face slightly higher frictions than do white women. The dispersion of frictions for black women is also higher than that of white women. Overall, the table shows that frictions are higher for women than for men in all three countries.

	1991		2000		2010	
	mean	st_dev	mean	st_dev	mean	st_dev
white women	1.77	0.64	1.86	0.93	1.57	0.52
brown men	1.31	0.25	1.32	0.23	1.20	0.14
brown women	2.41	1.13	2.42	1.13	1.99	0.70

(a) Brazil

	1993		1999		2004	
	mean	st_dev	mean	st_dev	mean	st_dev
other women	2.24	0.80	2.29	1.10	2.09	0.73
sc. caste men	1.34	0.21	1.29	0.18	1.26	0.15
sc. caste women	2.79	1.26	2.54	0.95	2.64	1.33

(b) India

	1990		2000		2010	
	mean	st_dev	mean	st_dev	mean	st_dev
white women	1.55	0.58	1.46	0.52	1.45	0.53
black men	1.11	0.18	1.15	0.21	1.18	0.29
black women	1.62	0.85	1.53	0.76	1.59	0.73

(c) USA

Table 1.8: Summary stats of frictions across countries

1.6 Results

There are 8 exogenous parameters: A_i (technology by occupation), ϕ_i (elasticity of human capital with respect to schooling), τ_{ig} (frictions by occupation and group), q_g (total number of people by group), θ (the parameter that governs the dispersion of talent), η (the elasticity of human capital with respect to expenditure on human capital), σ (elasticity of substitution between occupations), and β (weight on consumption relative to time in the utility function). The baseline values of some parameters are given in Table 1.7. I check for robustness with different parameter values.

The number of people in each group q_g is taken from the data. Assuming that τ^h_{ig} captures the efficiency in human capital accumulation, I set $\bar{h}_{ig}$ to one. I normalize mean talent across groups for each occupation as $T_{ig} = 1$. The normalization $T_{ig} = 1$ assumes that there are differences in mean talent between men and women but that it is the same across occupations within groups. From the equation on average wage gaps and equilibrium condition for schooling and matching the wage gap in the data I estimate ϕ_i for each occupation. The technology parameter across occupations A_i is estimated from equations 1.22, 1.20, and 1.21. Values for the price of efficiency units of human capital w_i are obtained by using equation 1.13 and matching to the data.

1.6.1 Model fit

Given these parameters I can compare the results produced by the model with the data. In particular, I compare the model and data version of mean earnings and occupational shares across groups and occupations. In the model, equation 1.13 produces the occupational shares across groups and occupations and equation 1.15 produces mean earnings across groups and occupations.

The model is calibrated to the occupational shares of white men in each period. Table 1.9 compares the occupational shares produced by the model with the data for the five occupational categories with the highest shares for each group. For example, according to the data, the share of white men in Brazil working as farm non-managers is 0.102. The model counterpart of the data is also 0.102. For other groups in Brazil the model produces close results. According to the data, 40.6% of white women and 49.7% of brown women work in home sector. The model shows that 35.6% of white and 36.7% of brown women work in home sector.

In India most men work as farm non-managers and most women work in the home sector. The data shows that in 2004 33% of other men and 42% of scheduled caste men were occupied in farming. The model versions of these shares are 33% and 40%, respectively for other and scheduled caste men. In the same period 63.4% of other women and 48.6% of scheduled caste women were occupied in home sector. The model predicts that 61.5% of other women and 46% of scheduled caste women work in home.

	Data	Model
white men		
farm non-managers	0.102	0.102
construction	0.101	0.101
motor vehicl op.	0.096	0.096
sales	0.086	0.086
home	0.072	0.072
white women		
home	0.406	0.356
sales	0.075	0.050
private occupations	0.067	0.098
teachers	0.060	0.061
farm non-managers	0.034	0.031
brown men		
construction	0.148	0.154
farm non-managers	0.113	0.068
motor vehicle op.	0.089	0.109
related agriculture	0.087	0.085
home	0.065	0.043
brown women		
home	0.497	0.367
private occupations	0.098	0.156
sales	0.057	0.037
teachers	0.047	0.051
cleaning	0.031	0.064

(a) Brazil

	Data	Model
other men		
farm non-managers	0.330	0.330
sales	0.148	0.148
executives	0.050	0.050
construction	0.044	0.044
motor vehicle op.	0.040	0.040
other women		
home	0.634	0.615
farm non-managers	0.219	0.225
teachers	0.022	0.022
sales	0.021	0.024
precision, textile	0.019	0.016
scheduled caste men		
farm non-managers	0.422	0.400
sales	0.071	0.066
freight handler	0.067	0.062
construction	0.067	0.056
teachers	0.046	0.060
scheduled caste women		
home	0.486	0.460
farm non-managers	0.354	0.351
sales	0.025	0.024
teachers	0.018	0.025
private occupations	0.016	0.004

(b) India

Table 1.9: Occupational shares in Brazil and India (data vs model)

Table 1.10 shows the results of regressing the earnings data on the model version of earnings for each group and period in Brazil and India. For Brazilian white men in 1991, a value of 1.551 indicates that a 1 percent increase in mean earnings of white men in 1991 produced by the model corresponds to a 1.551 percent increase in mean earnings given by the data. Overall, the model produces less earnings than data. As can be seen from the table in Brazil the model produces the highest fit for white men in 2010 with an R^2 of 0.894. The lowest fit corresponds to the earnings of brown women in 1991 with an R^2 of 0.550.

The mean earnings for India produced by the model fits better the data in terms of the slope than for Brazil. Overall, 1 percent increase in mean earnings produced by the model corresponds to 0.8–1.15 percent increase in mean earnings given by the data. However, the percentage of the variation in the the data that the model explains is lower in India than in Brazil. The lowest fit belongs to scheduled caste women in 1991 with an R^2 of 0.230 and other men in 2004 an R^2 of 0.669.

	slope	st.error	R^2
1991			
white men	1.551	0.107	0.765
white women	1.412	0.149	0.585
brown men	1.613	0.117	0.747
brown women	1.532	0.176	0.550
2000			
white men	1.623	0.080	0.864
white women	1.591	0.129	0.704
brown men	1.552	0.093	0.812
brown women	1.716	0.128	0.744
2010			
white men	1.323	0.057	0.894
white women	1.418	0.063	0.886
brown men	1.299	0.058	0.885
brown women	1.440	0.081	0.831

(a) Brazil

	slope	st.error	R^2
1993			
other men	0.884	0.105	0.524
other women	1.096	0.182	0.364
SC men	0.729	0.110	0.401
SC women	0.906	0.220	0.230
1999			
other men	0.993	0.102	0.592
other women	1.150	0.165	0.441
SC men	0.926	0.117	0.496
SC women	1.098	0.196	0.346
2004			
other men	0.983	0.086	0.669
other women	1.156	0.128	0.563
SC men	0.959	0.100	0.584
SC women	1.164	0.169	0.445

(b) India

Table 1.10: Mean earnings across groups in Brazil and India (data vs model)

1.6.2 Output gain

Since I have all the exogenous parameters, I can compute aggregate output from the model. Then I can investigate how changing frictions affects the output. Since I have data only for aggregate frictions $\tau_{ig} = \frac{(1+\tau_{ig}^h)^\eta}{1-\tau_{ig}^w}$, I can not separately identify the effects from τ_{ig}^w and τ_{ig}^h. So, I do the analysis for two different cases: a case in which I allow frictions only in acquisition of human capital τ_{ig}^h, and a case in which only frictions in the labor market τ_{ig}^w are allowed.

I explore several counterfactuals. In a baseline case, I compute aggregate output in each period by using estimated frictions of each period, setting frictions to one period, eliminating all frictions, and using US frictions. Earlier I showed frictions faced by the groups in US. In this section I check if replacing frictions faced by the groups in Brazil and India with those of the US affects the aggregate productivity in these countries. In a robustness check section, I test the counterfactual output gain due to zero frictions with different parameter values.

Counterfactual output gain in Brazil

Table 1.11 presents output gain in Brazil with various frictions. The top panel of the table shows output gain due to frictions in the labor market and the bottom panel shows output gain due to frictions in accumulation of human capital. The output gain is higher if frictions were replaced by the 2010 frictions in Brazil as shown in the first row of the table. In the case of frictions in the labor market if the 1991 and 2000 frictions were replaced by the 2010 frictions, the output would increase by 9.4% and 4.4%, respectively. The second row shows the gain with the US frictions in 2010. The gains are 20.8%, 13.7% and 11.1%. From the previous sections, I showed that the 2010 frictions in Brazil are lower than in other periods, and that the frictions in the US are also lower than those in Brazil in corresponding periods. Thus, the analysis shows that output increases with the reduction of frictions. The last two rows show the counterfactual output gain from halving the frictions in corresponding years and removing them. Halving the frictions faced by the groups across occupations increases the output by 15.4%, 10.7%, and 9.3%. Removing them entirely increases the output even more by 42%, 31.1%, and 27.5%.[11] Output gain due to acquisition of human capital shows a similar pattern in output gain, but the gain is lower than with frictions due to the labor market.The gain increases both with τ_{ig}^{w} and τ_{ig}^{h} cases. In the τ_{ig}^{h} case eliminating frictions has a smaller effect compared to the τ_{ig}^{w} case.

[11] Hsieh et al. (2013) has output gain of 14.3% in 2008 for the US if frictions were reduced to zero.

Frictions in labour market			
	1991	2000	2010
Brazil 2010 friction	9.4%	4.4%	0.0%
US 2010 friction	20.8%	13.7%	11.1%
Frictions halved	15.4%	10.7%	9.3%
No friction	42.0%	31.1%	27.5%
Human capital frictions			
Brazil 2010 friction	12.9%	10.2%	0.0%
US 2010 friction	20.6%	14.8%	5.9%
Frictions halved	20.2%	17.3%	11.4%
No friction	35.5%	28.2%	21.4%

Table 1.11: Counterfactuals: Output gain in Brazil

Counterfactual output gain in India

For India I first show the results of the model with the limited number of caste categories but detailed occupational categories. Then I investigate if results change with more detailed caste categories but broader occupational categories. I do this because the data size is small with detailed caste and detailed occupation categories. So there is a trade-off between the number of caste categories and the number of occupation categories.

Broader caste categories. Table 1.12 presents counterfactuals output gain in India due to labor market frictions (τ_{ig}^{w}) on the top and frictions in human capital (τ_{ig}^{h}) on the bottom panel. The following four cases are investigated: output gain if frictions were replaced by 2004 Indian frictions, gain with US 2010 frictions, gain if frictions were halved, and gain if frictions were removed. Replacing the 1993 and 1999 frictions in India with 2004 frictions increases production in 1993 and 1999 by 4% and 2.7%, respectively, meaning that frictions in 2004 were slightly less than in 1993 and 1999. If frictions faced by the groups were replaced by those of the groups in the US in 2010 the output would increase by 23.6%, 22.5%, and 19.9%, respectively in 1993, 1999, and 2004. Cutting frictions to half in all groups across all occupations increases the output even more, by 17.1%, 14.9%, and 14.1% in the corresponding years. We observe higher gains than in Brazil when frictions are reduced to zero. Removing all frictions increases aggregate output by 45.9% in 1993 and by 44.2% and 41.2% in 1999 and 2004. The model predicts that output increases with reducing frictions. The gain increases both with τ_{ig}^h and τ_{ig}^w cases. In the τ_{ig}^h case eliminating frictions has a smaller effect compared to the τ_{ig}^w case.

Frictions in labour market			
	1993	1999	2004
Indian 2004 friction	4.0%	2.7%	0.0%
US 2010 friction	23.6%	22.5%	19.9%
Frictions halved	17.1%	14.9%	14.1%
No friction	45.9%	44.2%	41.2%
Human capital frictions			
Indian 2004 friction	5.6%	4.9%	0.0%
US 2010 friction	21.2%	22.9%	19.6%
Frictions halved	22.4%	22.3%	19.8%
No friction	39.4%	40.0%	36.1%

Table 1.12: Counterfactuals: Output gain in India

Detailed caste categories. Here I show the results generated by using detailed caste categories. The categories available for all periods are “other”, “scheduled tribe”, and “scheduled caste”. I use only 19 broad occupation categories as opposed to the 67 occupation codes used in the previous analysis. The broader categories are aggregated by using 67 occupations. These 19 occupation categories are shown in Appendix Table A2.

Table 1.12 shows the effect of reducing frictions faced by different groups on aggregate production in India. The column headings refer to the number of caste categories. The column 2 shows the output gain with 3 caste categories and column 3 shows the output gain with 2 caste categories. As can been seen from the table the output gain is close in both cases. The counterfactual output gain from removing all frictions with detailed castes increases the aggregate output by 34%, 31.1%, and 28.7% in 1993, 1999, and 2004, respectively. The counterfactual output gain from removing all frictions with broad castes increases the aggregate output by 33.6%, 30.9%, and 28% in 1993, 1999, and 2004, respectively.

The gain due to removing frictions in the case of τ_{ig}^h is also substantial. In the case of detailed caste categories removing frictions in all occupations faced by the groups results in increase of the output by 33.9%, 31.7%, and 28.9% in 1993, 1999, and 2004, respectively. The counterfactual gain in the case of broad caste categories is also significant: the output goes up by 33.1%, 31.1%, and 27.9% in the respective years.

	more castes	less castes
	due to labor market	
1993	34.0%	33.6%
1999	31.1%	30.9%
2004	28.7%	28.0%
	due to human capital	
1993	33.9%	33.1%
1999	31.7%	31.1%
2004	28.9%	27.9%

Table 1.13: Counterfactuals: Output gain in India with detailed caste categories

Gains in Brazil vs. India

The output gains from reducing frictions are larger in India than in Brazil. According to the model, there are three forces that vary across countries and that affect output gains from removing frictions: occupational shares, wage gaps, and population shares. Here I investigate which of these three forces is most important for larger gains in India than in Brazil. To do that I compute output in India by replacing each of the three items in India with that of Brazil. Then I compare counterfactual output gains in India by removing frictions. Table 1.14 shows the results in the case of frictions in labor market and friction in human capital, respectively. The first row shows the baseline case with occupational shares, wage gaps and population shares in India where the gain in output is due to removing frictions.

The second row of Table 1.14 illustrates how changes in wage gaps in India affect the output of the country. This is a counterfactual in which wage gaps of the groups in India are replaced by the wage gaps of the groups in Brazil. The effects of changing the wage gaps in the case of frictions in human capital and in the labor market are similar in all years. This indicates that wage gaps faced by the groups in Brazil and India are similar in corresponding years.

The third row of the table shows the counterfactual gain if the Indian population shares were replaced by Brazilian population shares. That is, the population shares of the four groups in each period in India are replaced with the population shares of the four groups in Brazil in corresponding periods, holding everything else fixed. This will produce the output gain from removing frictions in the case of frictions in labor market of 53.5%, 52.7%, and 52.2%, and in case of frictions in human capital of 44%, 45.8%, and 43% in 1993, 1999, and 2004, respectively. The gain is larger with Brazilian population shares than with Indian population shares. This is not surprising since the share of disadvantaged groups in India is smaller than the share of disadvantaged groups in Brazil, and reducing frictions for groups with larger population share will have a larger effect on output.

The last row shows the productivity effects of replacing the occupational shares in India with the occupational shares in Brazil, holding everything else fixed. Removing frictions will result in 34.7%, 31.1%, and 28.6% increase in output in the case of the frictions in labor market and in 27%, 25.7%, and 15.8% increase in output in the case of the frictions in human capital, in corresponding years. The gains are smaller with Brazilian occupational shares than with Indian occupational shares. This shows that in India the groups are misallocated more than the groups in Brazil.

	1993	1999	2004
Baseline	45.9%	44.2%	41.2%
with Brazilian wage gaps	45.9%	44.6%	41.7%
with Brazilian population shares	53.5%	52.7%	52.2%
with Brazilian occupational shares	34.7%	31.1%	28.6%

(a) Frictions in the labor market

	1993	1999	2004
Baseline	39.4%	40.0%	36.1%
with Brazilian wage gaps	39.4%	40.1%	36.3%
with Brazilian population shares	44.0%	45.8%	43.0%
with Brazilian occupational shares	27.0%	25.7%	15.8%

(b) Frictions due to human capital

Table 1.14: Counterfactual output growth in India

1.6.3 Robustness analysis

In this section, I test the previous results for robustness. I compute the output gain with different values of θ, η, and σ. The exercise is done separately by allowing frictions in the labor market and in the acquisition of human capital. The results in Table 1.15 and Table 1.16 show the gain in output in 2010 for Brazil and in 2004 for India when all frictions are removed.

The first row of Table 1.15 shows the output gain in Brazil due to removing frictions with changing η, holding other parameters constant. As can be seen, the results with changing η are robust. The gain does not change much with changing σ, except for $\sigma = 15$. The change in gain from the baseline case when $\sigma = 15$ is 4–6%. With changing θ, the difference from the baseline case is the highest when $\theta = 8.4$.

Frictions due to labor market				
	$\eta = 0.25$	$\eta = 0.15$	$\eta = 0.5$	$\eta = 0.1$
changing η	27.5%	27.5%	27.5%	27.5%
	$\sigma = 3$	$\sigma = 4.5$	$\sigma = 15$	$\sigma = 2.75$
changing σ	27.5%	29.8%	32.8%	26.9%
	$\theta = 3.44$	$\theta = 4.16$	$\theta = 5.6$	$\theta = 8.4$
changing θ	27.5%	27.5%	27.5%	27.5%
Frictions due to human capital				
	$\eta = 0.25$	$\eta = 0.15$	$\eta = 0.5$	$\eta = 0.1$
changing η	21.4%	19.3%	27.4%	18.3%
	$\sigma = 3$	$\sigma = 4.5$	$\sigma = 15$	$\sigma = 2.75$
changing σ	21.4%	23.0%	25.1%	20.9%
	$\theta = 3.44$	$\theta = 4.16$	$\theta = 5.6$	$\theta = 8.4$
changing θ	21.4%	20.1%	18.4%	16.4%

Table 1.15: Output gain in Brazil due to removed frictions

The results in Table 1.16 display the gain in output in 2004 with changing parameters in India. For the case with friction in the labor market, changing η does not change the gain in output relative to the baseline case. With changing σ, the output varies from the baseline by 20% when $\sigma = 15$. Varying θ shows no difference from the baseline gain.

The pattern of output gain in the case of frictions in human capital acquisition is different from frictions in the labor market. The gain differs from the baseline by 4–5% when $\eta = 0.5$, by 10–11% when $\sigma = 15$ and by 1–3% with different values of θ.

Frictions due to labor market				
	$\eta = 0.25$	$\eta = 0.15$	$\eta = 0.5$	$\eta = 0.1$
changing η	41.2%	41.2%	41.2%	41.2%
	$\sigma = 3$	$\sigma = 4.5$	$\sigma = 15$	$\sigma = 2.75$
changing σ	41.2%	43.2%	21.3%	40.5%
	$\theta = 3.44$	$\theta = 4.16$	$\theta = 5.6$	$\theta = 8.4$
changing θ	41.2%	41.2%	41.2%	41.2%
Frictions due to human capital				
	$\eta = 0.25$	$\eta = 0.15$	$\eta = 0.5$	$\eta = 0.1$
changing η	36.1%	34.6%	40.6%	33.8%
	$\sigma = 3$	$\sigma = 4.5$	$\sigma = 15$	$\sigma = 2.75$
changing σ	36.1%	37.2%	22.9%	35.8%
	$\theta = 3.44$	$\theta = 4.16$	$\theta = 5.6$	$\theta = 8.4$
changing θ	36.1%	35.6%	34.7%	33.3%

Table 1.16: Output gain in India due to removed frictions

1.7 Conclusion

The purpose of the paper was to investigate the labor market outcomes in Brazil and India and document their effects on aggregate productivity. I showed that there are disadvantaged groups in Brazil and India. The share of groups other than white men in Brazil and other men in India in highly skilled occupations is low. Only 1–2 % of brown men and women in Brazil, and 1% of women in India are occupied in highly skilled occupations. The wage gap between groups and the privileged group in these countries is also significant. The earnings of brown women in Brazil and scheduled caste women in India are 48–50% lower and than the earnings of privileged men with similar characteristics.

The effect of the resulting occupational choice from frictions in the labor market and in the acquisition of human capital is significantly negative. The augmented Roy model used by Hsieh et al. (2013) allowed me to estimate the potential gains to output from reducing frictions in human capital accumulation and the labor market. Reducing frictions faced by various groups in Brazil and India increases the aggregate productivity of the countries. In particular, the results suggest that reducing frictions to half may increase output by 9–20% in Brazil and by 14–22% in India. Removing frictions increases the aggregate productivity of the countries by 21–42% in Brazil and 36–45% in India.

REFERENCES

Acemoglu D., Reward structures and the allocation of talent, European Economic Review, 39(1):17–33, 1995.

Altonji J. and Blank R., Race and gender in the labor market, in O. Ashenfelter and D. Card, Handbook of labor economics, Amsterdam, 1999.

Antecol H. and Bedard K., The racial wage gap: The importance of labor force attachment differences across brown, Mexican, and White Men, Journal of Human Resources, 39(2), 564–583, 2001.

Arabsheibani R., Carneiro F., and Henley A., Gender wage differentials in Brazil: Trends over a turbulent era, World Bank working paper, 2003.

Baumol W., Entrepreneurship: Productive, unproductive, and destructive, Journal of Political Economy, 1990.

Deshpande A., The grammar of caste: Economic discrimination in contemporary India, New Delhi: Oxford University Press, 2011.

Deshpande A., Goel D. and Khanna S., Bad karma or discrimination? Male–women wage gaps among salaried workers in India, Centre for Development Economics Working Paper 243, Delhi School of Economics, 2018.

Eaton J. and Kortum S., Technology, geography, and trade, Econometrica, 70(5): 1741–1779, 2002.

Heckman J., Sample Selection Bias as a Specification Error, Econometrica, 1979.

Hsieh C., Hurst E., Jones C. and Klenow P., The allocation of talent and U.S. economic growth. mimeo, University of Chicago-Booth, 2013.

Hsieh C. and Klenow P., Misallocation and Manufacturing TFP in China and India, The Quarterly Journal of Economics, 124(4):1403-1448, 2009.

Klasen S., Low schooling for girls, slower growth for all? Cross-country evidence on the effect of gender inequality in education on economic development, World Bank Economic Review, 16:345-373, 2002.

Lagerlof N., Gender Equality and Long-Run Growth, Journal of Economic Growth, Vol. 8, 403-426, 2003.

Lovell P., The geography of economic development and racial discrimination in Brazil, Development and Change, 24(1):83–102, 1993.

Loureiro P., Carneiro F., and Sachsida A., Race and gender discrimination in the labor market: an urban and rural sector analysis for Brazil, Journal of Economic Studies, Vol. 31 Issue: 2, pp.129-143, 2004.

Macinko J., Mullachery P., Proietti F., and Lima-Costa L., Who experiences discrimination in Brazil? Evidence from a large metropolitan region, International Journal for Equity Health, 11(80):1–11, 2012.

Madalozzo R., Occupational segregation and the gender wage gap in Brazil: an empirical analysis, Economia Aplicada, 14(2), 147-168, 2010.

Madeshwaran S. and Attewell P., Caste Discrimination in the Indian Urban Labour Market: Evidence from the National Sample Survey, Economic and Political Weekly Vol. 42(41), 2007.

McFadden D., Conditional Logit Analysis of Qualitative Choice Behavior, in P. Zarembka, ed., Frontiers of Econometrics, New York, NY: Academic Press, pp. 105–142, 1974.

Minnesota Population Center, Integrated Public Use Mocrodata Series, International: Version 6.3. Technical report, Minneapolis: University of Minnesota, 2014.

Mulligan C. and Rubinstein Y., Selection, Investment, and Women's Relative Wages over Time, The Quarterly Journal of Economics, 2008.

Munshi K. and Rosenzweig M., Traditional Institutions Meet the Modern World: Caste, Gender and Schooling Choice in a Globalizing Economy, American Economic Review, 96(4): 1225-52, 2006.

Roy A., Some thoughts on the distribution of earnings, Oxford economic papers, 3(2):135–146, 1951.

Telles E., Race in another America: the significance of skin color in Brazil. Princeton University Press, 2014.

Thorat S. and Attewell P., The Legacy of Social Exclusion: A Correspondence Study of Job Discrimination in India, Economic and Political Weekly, 2007.

Thorat S. and Newman K., Blocked by caste: Economic discrimination in modern India, New Delhi: Oxford University Press, 2010.

Chapter 2

IMMIGRANT CHARACTERISTICS IN LOW-INCOME COUNTRIES

2.1 Introduction

The analysis of immigrants in developed countries has shown that labor market outcomes of immigrants and natives differ. Immigrants experience occupational downgrading on arrival, work in lower-paid occupations and earn less than natives with similar experience and schooling. Over time, the earnings gap between immigrants and natives closes. Immigrants experience occupational upgrading as they adjust to the host country labor market. The initial disadvantage in the labor market outcomes between immigrants and natives diminishes.

Studies have explained the differences in earnings between immigrants and natives by the differences in human capital endowments that immigrants from different countries possess (Hendricks (2002), Schoellman (2012), Lagacos et al. (2017), Hendricks and Schoellman (2018)). Others have proposed that the difference in earnings is due to the skill loss experienced by immigrants when moving from the source to the host country (Chiswick (1978, 1979, 1980), Duleep and Regets (1997, 1999, 2002)). To distinguish between these two causes of earnings difference, I investigate immigrants in low-income countries for whom skill transfer is likely to be less of a problem. Since lower earnings of immigrants are explained by the lack of transferable skills, I assume that immigrants in low-income countries do not face skill loss since they are occupied in higher-paid occupations and their earnings are higher than those of natives. The additional evidence that immigrants in low-income countries do not experience a skill loss in low-income countries is that the share of college-educated migrants in high-skill occupations is higher than that of non-migrant workers in the source countries.

To test if the difference in immigrants' earnings is due to the human capital endowments of immigrants from different countries, I follow Hendricks (2002). His findings suggest that the unobserved human capital of immigrants does not vary as much as the income per capita of the source countries. A possible explanation of this is that immigrants from poor countries are more positively selected than immigrants from rich countries. In particular, the studies have shown that immigrants are positively selected on education and that the selection is higher for immigrants from low-income countries (e.g. Schoellman (2012) and Hendricks and Schoellman (2018)). A more positive selection of poor country immigrants results in higher unobserved skills than those of workers in the source country. Thus, the self-selection of immigrants in developed countries drives a wedge between the human-capital endowments of immigrants and source country workers. The analysis in low-income countries shows that immigrants are selected much less on schooling and that the differential selection of immigrants is less than in developed countries, which suggests that the variation in immigrant earnings should be higher in low-income countries than in developed countries.

The disadvantage in the labor market outcomes between immigrants and natives closes over time. The labor market adjustment of immigrants to the host country labor market in developed countries is shown by the closing of the gap in earnings and the gap in occupational distribution between immigrants and natives. Chiswick (1978, 1979, 1980) documents that immigrants adjust to the environment by investing in host-country-specific skills, which leads to faster growth of their earnings relative to natives. Duleep and Regets (1997, 1999, 2002) explain the adjustment by the accumulation of human capital. The opportunity cost of investing in human capital is lower for immigrants than for natives in developed countries. Thus, immigrants accumulate more human capital than natives. In low-income countries, immigrants are paid more and are in better-paid occupations than natives. Then, due to the higher opportunity cost of investing in human capital, immigrants should accumulate less human capital than natives with similar characteristics.

The purpose of this paper is to investigate empirically whether the difference in immigrant earnings in low-income countries is due to the difference in human capital endowments and to investigate if low-earning natives accumulate more human capital than immigrants. I investigate immigrants in low-income countries by using census data from Brazil, Mexico, and Venezuela. The analysis shows that, although immigrants are positively selected on education, they are closer to the source country workers in terms of attained education levels than immigrants in developed countries. The relationship between the average schooling years of migrants and non-migrants shows that the selection of immigrants from poor countries is not as high as the studies in developed countries show. This result in low-income countries translates to larger differences in the unobserved human capital of immigrants. By using immigrant earnings, I estimate the unobserved human capital of immigrants. The analysis of immigrant earnings in low-income countries indicates that the variation of unobserved human capital of immigrants is larger than findings in developed countries show. In particular, the slope that describes the relationship between immigrant unobserved human capital and the source country income is 0.31–0.37 in low-income countries, whereas in the US it is 0.14. This result supports the findings of Hendricks and Schoellman (2018) that differences in human capital across immigrants are large.

I also show that immigrants in low-income countries earn more and are in better-paid occupations than natives. The analysis of immigrants in Brazil indicates that the earnings gap between immigrants and natives closes over time and that natives upgrade their occupations relative to immigrants. The initial earnings of the 1980–1990 and 1990–2000 arrival cohorts were 30% and 40% higher than natives with similar characteristics, respectively. Over time, the earnings gap between 1980–1990 and 1990–2000 arrival cohorts and natives has decreased to 26% and 31%, respectively. The wage gain from changing occupational distribution relative to natives for immigrants with high school degrees or less is -0.74% and -6.6%, and for college-educated immigrants is -1.13% and -6.09% after 10 and 20 years of stay in Brazil, respectively. This shows that natives in Brazil accumulate more human capital than immigrants, which supports the findings in developed countries that lower-earning groups accumulate more human capital.

This paper is organized as follows. Section 2.2 reviews the literature, Section 2.3 describes the census data obtained from the Integrated Public Use Microdata Series, Section 2.4 examines if the difference in earnings is due to human capital or skill transferability, Section 2.5 discusses the selection of immigrants and compares the characteristics of migrants and non-migrants, Section 2.6 provides empirical evidence on unobserved skill differences of immigrants, Section 2.7 describes the findings related to assimilation of immigrants, and Section 2.8 concludes.

2.2 Literature review

Studies in developed countries have shown that immigrant earnings are lower than those of natives with comparable characteristics and that immigrants are occupied in lower-paid occupations. Duleep and Dowhan (2002), by analyzing immigrants in the US, found that 1965–1969 arrival cohorts earned 17% and cohorts who immigrated after 1969 earned 28–46% less than natives with similar observable characteristics. Antecol et al. (2003) document that recent immigrants in Australia, Canada, and the USA earn 5.3%, 43.8% and 52.9% lower than natives, respectively. Winkelmann (2005) shows that immigrant earnings in New Zealand are 20–25% lower than those of natives. Immigrants work in lower-paid occupations relative to natives with similar characteristics. Dustman et al. (2014) point out that recent immigrants in the UK work in lower-paid occupations although they are better educated than the overall population. Zorlu (2013), by investigating immigrants in the Netherlands, documents that Turkish and Moroccan immigrants are in jobs at the lower levels of skill distribution.

2.2.1 Immigrant earnings and human capital

Hendricks (2002) has proposed that immigrant earnings reflect the human capital endowments of the source countries. This comes from the fact that immigrants migrate with the source country human capital and face similar skill prices in the host country labor market. Given observable characteristics, immigrant unobserved skills relative to natives are derived from their earnings. His findings suggest that the unobserved human capital of immigrants does not vary as much as the income per capita of the source countries, meaning that there is not much difference between the unobserved human capital of the immigrants from poor countries and that of the immigrants from rich countries. A possible explanation of this is that immigrants from poor countries are more positively selected than immigrants from rich countries.

Human capital across immigrants also differs by the quality of the source country schooling. Schoellman (2012) estimates the quality of schooling for countries by analyzing the return to schooling of immigrants in the US. He shows that immigrants from developed countries have higher return to schooling than do immigrants from developing countries. He documents the importance of the measure of education quality estimated from immigrant earnings in accounting for cross-country income differences. Taking into account country differences in education quality increases the contribution of schooling in cross-country income differences by 10% to 20%.

Another measure of unobserved human capital differences across countries is proposed by Lagacos et al. (2017). They investigate the differences in returns to experience for immigrants from different countries. The paper finds that returns to experience accumulated in the source countries are higher for immigrants from developed countries than for immigrants from developing countries. By building the model of life-cycle human capital accumulation, Lagacos et al. point out that immigrants from poor countries accumulate less human capital in the source countries than do immigrants from rich countries.

By using the New Immigrant Survey (NIS), Hendricks and Schoellman (2018) investigate wage gains from migration for immigrants in the US. The NIS data allow control of the selection and skill transferability of immigrants by observing the wages of workers in the source as well as in the host country. Wage gains experienced by immigrants after migration are assumed to come from host-country-specific factors such as physical capital and Total Factor Productivity (TFP), and the remaining part from the human capital of the workers. By estimating the human capital of immigrants from different countries, the paper documents that human capital accounts for 60% of cross-country income differences.

2.2.2 Skill transferability

An alternative explanation of the earnings disadvantage of immigrants relative to natives is that immigrants cannot transfer their skills to the host country labor market and they lack host-country-specific skills. The studies assume that immigrants cannot fully utilize the source-country human capital due to lack of specific skills. Immigrants lack knowledge about the host country job opportunities, have less occupation-specific training, and they also lack host-country-specific credentials, such as diplomas or licenses specific to the host country labor market. To increase the value of their source country human capital, immigrants learn languages and attend trainings.

Chiswick (1978, 1979, 1980) pointed out that earnings of immigrants are depressed initially because immigrants experience a skill loss in the host country labor market. Immigrants have different levels of skill transferability because of the differences in the source and host countries' cultural and economic environments, languages, returns to schooling and experience. He theorized that low-skill-transferability immigrants will experience lower initial earnings than high-skill-transferability immigrants and natives.

According to Duleep and Regets (1997, 1999, 2002), immigrants from developed countries have higher earnings than other immigrants because they have higher transferable skills due to similarity in economic opportunities in the source and the host countries.

Only migrants with high transferable skills in developed countries migrate because workers with low transferable skills do not have the incentive to migrate due to the higher opportunity cost of investing in source-country-specific skills. Workers from developing countries have higher incentives to migrate and invest in host-country-specific skills due to having lower opportunity cost of investing in human capital.

Other studies assign a secondary role to skill transferability in explaining the labor market outcome of immigrants. It is important to take into account immigrants' skill transferability, because imperfect skill transfer leads to the understatement of the source country human capital endowments. By observing the same worker in two labor markets, Hendricks and Schoellman (2018) analyzed if immigrants were able to transfer their skills to the host country. Their findings suggest that immigrants experience imperfect skill transferability due to the occupational downgrading when they move from the source country to the host country. Lagacos et al. (2017) also show that immigrants in the US face imperfect skill transferability by analyzing the share of college-educated workers in high-skill occupations. They find that the share of college-educated immigrants in the US in high-skill occupations is lower than the share of college-educated non-migrants in those occupations in the source countries.

2.2.3 Assimilation

After spending some time in the host country, immigrants catch up with natives in terms of earnings and occupations. Chiswick (1978, 1979, 1980) proposed that immigrants lack host-country-specific skills when they initially arrive. Then immigrants acquire required host country skills and become more adapted to the host country labor market, which leads to faster growth of their earnings relative to natives. He measured immigrant earnings growth by using a single cross-sectional data set. He compared the earnings of recently arrived immigrants to the earnings of immigrants with similar observable characteristics who had been in the country longer. Duleep and Regets (1999, 2002) modified Chiswick's model by adding the opportunity cost of investing in human capital. Low-skill-transferability immigrants experience faster earnings growth than natives and high-skill-transferability immigrants due to their lower opportunity cost of investing in human capital. This leads to higher growth of earnings for immigrants relative to natives and the moving to better-paid occupations, thus closing the earnings gap relative to natives.

Many other studies have investigated immigrant assimilation by adopting varying methodologies. Cohort differences in observable and unobservable characteristics have been found to be important in investigating immigrant assimilation. Estimation of earnings growth by taking into account cohort-specific characteristics (Borjas (1985, 1987, 1992)) has shown that entry earnings of immigrants differ by the years of immigration. The other important question in studying immigrant assimilation is the role of the source country human capital in immigrants' adjustment to the host country labor market. The studies have pointed out the importance of distinguishing between the education and the experience obtained in source and in host countries (e.g. Friedberg (2000), Schoeni (1997), Bratsberg and Ragan (2002), Akresh (2006, 2007), Cohen-Goldner and Eckstein (2008) and Lagakos et al. (2017)). The findings suggest that education and labor market experience obtained in different countries are not perfect substitutes, the return to immigrants' schooling and experience is generally less than that of natives, or human capital obtained in developed countries is more valuable than human capital obtained in low-income countries.

Assimilation of immigrants to the host country labor market in terms of occupational mobility was also investigated by many researchers. Studies found evidence that immigrants initially downgrade then, after accumulating human capital, upgrade their occupations. The studies use two types of data, longitudinal and cross-sectional. The studies using longitudinal data (Chiswick et al. (2005), Akresh (2006), Hendricks and Schoellman (2018)) have found that most of the immigrants experience occupational downgrading and move to higher-paid occupations after spending time in the host country. The studies using cross-sectional data (Green (1999), Barrett and Duffy (2007), Mattoo et al. (2008), Chiswick and Miller (2008, 2009), Dustman et al. (2014), Zorlu (2013, 2016)) have found similar evidence about the occupational mobility of immigrants. The occupational distribution of recently arrived immigrants resembles the distribution of relatively uneducated natives: immigrants start working at lower-paid jobs, then, with the duration in the host country, their occupational distribution improves and they move to higher-paid jobs. Thus, immigrants are more occupationally mobile than natives.

2.3 Data

The analysis uses census data from the Integrated Public Use Microdata Series (IPUMS) for the following countries and periods: USA (1970, 1980, 1990, 2000), Brazil (1980, 1991, 2000, 2010), Venezuela (1981, 1990, 2001), and Mexico (1970, 1990, 2000, 2010). The variable names, coding schemes, and documentation are consistent for most samples, which makes the analysis more comparable across periods and countries. This section explains the data used in the paper, and issues related to the consistency of the certain variables across sample periods and countries.

The data are analyzed for the sample periods that have key variables. For example, the Brazilian census has key variables available for the following years: 1991, 2000, and 2010. The following restrictions are made to the data: 1) the analysis is restricted to individuals whose ages are between 20 and 65, 2) individuals who are on active military duty and unemployed individuals are excluded, 3) observations with missing income are excluded, and 4) immigrant source countries with fewer than 100 observations for the US and 50 observations for other countries are dropped.

Table 2.1 provides information on the number of observations after imposing the restrictions in Brazil, Mexico, and Venezuela. Brazilian data show that the number of immigrants in the samples dropped from 12,369 in 1991 to 7,612 in 2010. The analysis of the information on the internet does not indicate that the Brazilian government imposed restrictions on immigrants, which shows that the decline is not associated with any change in immigration policy. In fact, Villen (2017) provides data that show a large increase in working visas during this period. According to his study, the number of immigrant workers increased from 5,376 in 1993 to 14,741 in 2000, and to 55,471 in 2010.

The number of immigrants in the sample in Mexico increased from 4,333 in 1990 to 5,405 in 2010. The data also show that the number of immigrants in Venezuela constitutes a substantial part of the overall population of the country: 51,037 immigrants versus 275,984 and 49,021 immigrants versus 515,826 of the total working population for the samples in 1981 and 2001, respectively. Most of the immigrants in Venezuela are from Colombia and constitute a half of the total immigrant population in the sample. The large-scale Colombian immigration can be explained by the long border and the Colombian conflict since 1980.

Brazil			
Survey year	1991	2000	2010
Number of observations	2,045,386	2,448,835	3,349,398
Number of immigrants	12,369	10,304	7,612
Mexico			
Survey year	1990	2000	2010
Number of observations	1,290,567	1,536,279	1,713,270
Number of immigrants	4,333	4,910	5,405
Venezuela			
Survey year	1981	1990	2001
Number of observations	275,984	331,398	515,826
Number of immigrants	51,037	42,752	49,021
Immigrants from Colombia	23,368	26,614	27,348

Table 2.1: Sample data

Some variables that IPUMS provides are harmonized across countries and some are not. Harmonized variables that I use are edattaind, age, incearn, sex, empstat, hrswork, bplcountry and yrimm. Emptat indicates a person's employment status, which I use to identify employed individuals. Incearn reports the individual's total income in the previous month or year, expressed in the currency of the respective country. Hrswork shows a person's hours worked per week, which is used to compute hourly wages. Hrswork is not available for some surveys. Bplcountry shows the respondent's country of birth. It allows me to identify immigrants in the sample and their source countries. Year of immigration is captured by yrimm, which is available only for Brazil and the US. A person's educational attainment, identified by variable "edattaind," shows the person's educational attainment in terms of the level of schooling completed. Table 2.2 gives the detailed coding of the variable, from which I construct a variable that indicates a person's number of schooling years completed, the third column of Table 2.2. There is a limitation in constructing years of schooling from edattaind because it will show only the approximate number of years of schooling. For example, there is a discontinuity between 8 and 12 years of schooling, and it also will not allow me to identify individuals with more than 16 years of schooling.

Code	Label	Schooling years
	Less than primary completed	
110	No schooling	0
120	Some primary	2
130	Primary (4 years)	4
	Primary completed, less than secondary	
211	Primary (5 years)	5
212	Primary (6 years)	6
	Lower secondary completed	
221	General and unspecified track	7
222	Technical track	8
	Secondary completed	
311	General track completed	12
312	Some college/university	13
320	Technical track	13
321	Secondary technical degree	13
322	Post-secondary technical education	13
400	University completed	16

Table 2.2: IPUMS Educational attainment

From the available data I construct hourly wages, experience, and a variable that identifies immigrant cohort. I construct hourly wages from monthly earnings and hours worked per week. Cohorts of immigrants in Brazil are identified by migration years. Experience is constructed from the individual's age and years of schooling completed as age minus schooling minus 6. As I discussed previously, there is a problem with recording years of schooling correctly for some observations, which leads to a difficulty in recording the potential experience for some observations. It may overstate the actual potential experience if actual years of schooling is higher and understate if the actual years of schooling is lower.

Table 2.3 summarizes the key variables in the sample by providing their means, standard deviations, and minimum and maximum values. Mean age and mean years of schooling have increased over time for all countries. The difference in mean earnings between 1990 and 2000 in Mexico and Brazil is related to the currency revaluations conducted by the corresponding countries, Mexico in 1993 and Brazil in 1994.

	mean	sd	min	max
1991				
age	35.8	11.2	20.0	65.0
educ	5.37	4.63	0.0	16.0
lwage	11.0	1.1	1.1	16.8
2000				
age	36.6	11.2	20.0	65.0
educ	6.36	4.65	0.0	16.0
lwage	5.8	1.0	0.0	13.5
2010				
age	37.9	11.6	20.0	65.0
educ	7.7	4.88	0.0	16.0
lwage	6.6	0.9	0.0	14.4

(a) Brazil

	mean	sd	min	max
1990				
age	35.9	12.2	20.0	65.0
educ	5.59	4.5	0.0	16.0
lwage	13.0	1.3	0.0	18.3
2000				
age	36.8	12.3	20.0	65.0
educ	5.9	4.54	0.0	16.0
lwage	7.6	0.9	0.7	13.8
2010				
age	38.0	12.5	20.0	65.0
educ	6.42	4.42	0.0	16.0
lwage	8.2	0.9	0.0	13.8

(b) Mexico

	mean	sd	min	max
1981				
age	34.9	11.2	20.0	65.0
educ	5.95	3.9	0.0	16.0
lwage	7.6	0.8	0.7	14.4
1990				
age	35.7	11.2	20.0	65.0
educ	5.55	4.3	0.0	16.0
lwage	8.7	0.8	0.7	11.5
2001				
age	36.8	10.8	20.0	65.0
educ	7.89	4.0	0.0	16.0
lwage	6.0	1.1	0.0	12.9

(c) Venezuela

Table 2.3: Sample data

The analysis also uses the data on occupations, which are available from IPUMS. These indicate an individual's primary occupation, classified according to the system used by the respective census of each country. Country surveys have different classification systems for occupations. Moreover, the Brazilian survey has varying classifications for different periods. To make data comparable across years and countries, I harmonize the occupational coding to the 1990 census occupational classification system used by Hsieh et al. (2013) and aggregate to 19 occupations.[1] The aggregation of the occupations to the broader category is done by merging related occupational categories. For instance, executive, administrative, and managerial occupations are merged into one broader category as management-related occupation.

Section 2.5 about the selection of immigrants compares the education levels of migrants and non-migrants. The data on the education of non-migrant populations are taken from Barro and Lee (2013). The data set provides educational attainment information for 146 countries from 1950 to 2010. The data are disaggregated by sex and by five-year age intervals. The advantage of the data set is that it has information on average years of schooling for countries which are analyzed in the paper. I compare years of schooling of the population aged 25 and over in 2000 for Venezuela[2] and in 2010 for other countries.

I also use the Penn world table database (PWT version 8.1). The database contains information on countries' relative income levels and populations, covering most of the countries analyzed in this paper. It is used to compare source country GDP per capita to the human capital of immigrants. The GDP per capita of a country is computed by dividing real GDP at chained PPPs to the population of the country.

2.4 Human capital vs skill transferability

Studies have explained the differences in earnings across immigrants and natives by the differences in human capital endowments that immigrants from different countries possess. Immigrants from developed countries have higher unobserved human capital, higher returns to schooling, and higher returns to experience than those coming from developing countries. Alternative interpretations of the difference in earnings between immigrants and natives are that immigrants cannot fully transfer their skills or that they lack host-country-specific skills. Immigrants from developing countries have lower

[1] The detailed occupational coding is provided in the Appendix.

[2] Venezuela has immigrant data only up to 2001.

transferable skills than immigrants from developed countries. Only workers with high transferable skills migrate from developed countries since workers with low transferable skills do not have an incentive to migrate due to having a higher opportunity cost of investing in host-country-specific skills.

To distinguish between two interpretations of earnings difference across immigrants, I investigate immigrants in low-income countries for whom skill transfer is less likely to be a problem. Since studies explain lower initial earnings of immigrants as due to their initially lacking transferable skills, I assume that immigrants in low-income countries do not face skills loss because they work in higher-paid occupations and their earnings are higher than those of natives. The share of immigrants in high-skilled occupations also shows that immigrants in low-income countries experience less skills loss than immigrants in high-income countries. By following Lagacos et al. (2017), I compute the shares of college-educated migrant and non-migrant workers in high skilled occupations.[3] Figure 2.1 compares the frequencies of working in high-skilled jobs by migrants and non-migrants. Thus, if the share of college-educated migrants working in high-skill occupations is lower (higher) than those of non-migrants with similar characteristics in the source countries, then migrants experience (do not experience) a skills loss. As can be seen from the figure, most of the immigrants in the US experience a skills loss, as shown by the lower share of college-educated migrants in high-skill occupations than non-migrants. In Brazil, most of the immigrants do not experience a skills loss.

Another way to see it is to look at the average share of immigrants working in high-skilled occupations. In Brazil, the average share is 0.84, meaning that college-educated immigrants in Brazil tend to work at high-skill occupations at a higher frequency than those in the US. It is clear that immigrants in the US are less likely to work at high-skill occupations.

[3]High-skill occupations are defined as professionals, technicians, associate professionals, legislators, senior officials and managers.

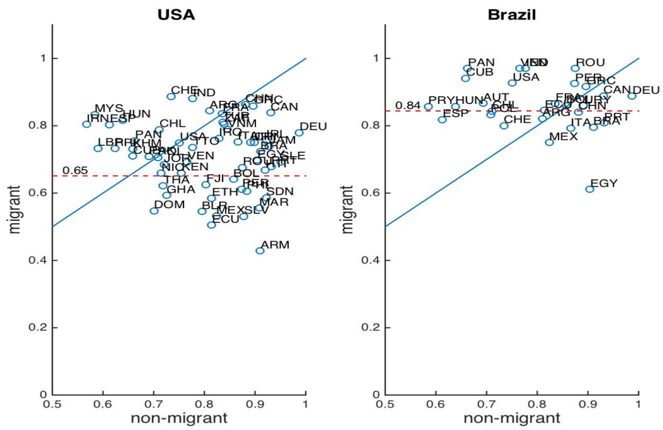


Figure 2.1: The share of college-educated workers in high-skilled occupations

2.5 Selection

Immigrants are a selected group of the source country populations. Theoretical studies suggest that immigrants are negatively selected if they are from poor and unequal countries (Borjas (1987)). Others document positive selection due to higher migration costs faced by lower-educated immigrants (Chiquiar and Hanson (2005)). Host country networks also play a role in self-selection by reducing costs and increasing returns to migration ((Orrenius and Zavodny (2005), McKenzie and Rapoport, (2010)).

Empirical studies in the US have documented that immigrants are positively selected based on education (e.g. Schoellman (2012), Hendricks and Schoellman (2018), Lagacos et al. (2017)). Schoellman (2012) hypothesized that if immigrants are positively selected on education then they must be positively selected on cognitive ability, which suggests that they are more productive than non-migrants. A recent study by Hendricks and Schoellman (2018) using longitudinal data of immigrants in the US provides evidence that immigrants are highly selected on characteristics such as education and wages, and that immigrants from poor countries are selected much more on these characteristics.

To investigate if immigrants in low-income countries represent a selected group of the source country population, I construct a measure of selection that compares education levels of migrants and non-migrants of the various source countries. This measure of selection will show if there is a difference in characteristics between migrants and non-migrants. Then I compare immigrants in various host countries to find out if the measure of selection varies across different host countries.

Migrants are considered to be selected on education if the average education level of migrants differs from the average education level of non-migrants. In particular, if the level of education of immigrants is higher than that of non-migrants then there is a positive selection on education; if it is lower then there is a negative selection on education. To construct average years of schooling of immigrants, I use an educational attainment variable which is available in the censuses of the respective countries from IPUMS. The data provide an information on respondents' education level and country of birth. By using information on years of schooling and country of birth, I construct weighted average years of schooling for the source countries, which is then defined as the average years of schooling of migrants. Average years of schooling of non-migrants are taken from the Barro-Lee data set. The data set contains information on the educational attainment of the populations of many countries by five-year age groups. For each country I construct average years of schooling for the population aged between 25 and 65 by weighting the schooling of each age group by the corresponding share of the group. Next I compare the resulting average years of schooling of immigrants in the host countries and average years of schooling of non-migrants in the source countries.

Figure 2.2 compares the average years of schooling of migrants and non-migrants in 2000. The vertical and the horizontal axes represent average years of schooling of migrants and non-migrants, respectively. From the plot for the US, we see that data points lie above the 45 degree line, meaning that immigrants in the US are more educated than non-migrants. We also see that the level of selection increases for poor countries. The selection on education is the highest for immigrants from India, as shown by the large distance from the 45 degree line. This supports findings in other studies documenting that immigrants in the US are positively selected on education and that the degree of positive selection increases for poor countries. Schoellman (2011) documents that immigrants in the US are positively selected on years of schooling and that the selection is highest for immigrants from Afghanistan, Nepal, Sierra Leone, and Sudan, who have 10 to 12 years of schooling difference from non-migrants.

Figure 2.2 also compares the education levels of migrants in Brazil, Mexico, and Venezuela to those of the non-migrants. The analysis shows that immigrants are positively selected but the degree of selection is lower than in the US. This is clearly seen for Brazil and Venezuela, as the data points lie around the 45 degree line. We still see that immigrants from poor countries are positively selected, but the degree of positive selection is not as strong as in the US. The main takeaways from the analysis are that immigrants from poor countries in the US are selected by a factor of 3-4, while immigrants in other countries are selected by a factor of 2-3. All immigrants from rich countries in the US are positively selected. This is not true for all immigrants from rich countries in other host countries. Some immigrants are positively selected and some are negatively selected on education.

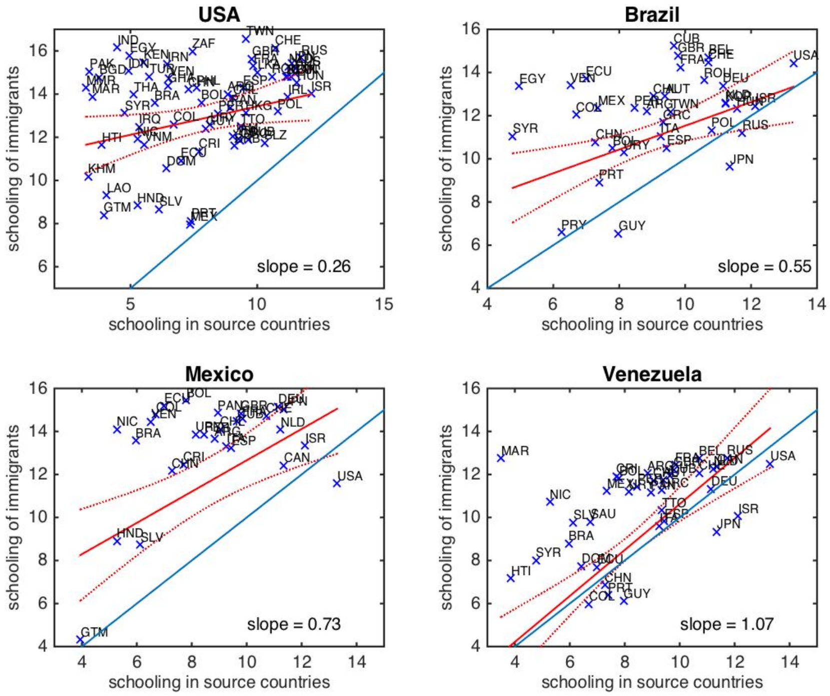


Figure 2.2: Education levels of immigrants vs non-migrants

The relationship between the educational attainment of migrants and that of non-migrants is also shown by the regression line and the slope in Figure 2.2. The regression was adjusted by the source country weights. The slopes of the lines describe the relationship between immigrant and non-migrant schooling and show how steep it is. The higher slope indicates that selection on education is closer between source countries. For example, the slope of 0.26 in the US is the lowest among four countries, meaning that "low school" countries are more selected on education than "high school" countries. The highest slope belongs to Venezuela, which indicates that immigrants from "high school" and "low school" countries are equally selected.

This section explored the question of immigrant selection in different host countries. The main takeaway from this section is that selection on education is higher for immigrants in the US than in other host countries.

2.6 Unobserved skill differences

Immigrant earnings are positively related to source country incomes. Immigrants from developed countries earn more than immigrants from developing countries given observable characteristics. Thus, immigrant earnings convey some information about the development of source countries. This observation has encouraged many studies to investigate labor market outcomes of immigrants in host countries. Studies that investigate immigrant earnings and their occupations have documented that immigrants differ in unobserved skills, return to schooling and experience accumulated in source countries.

To estimate the unobservable skills differences of immigrants, I follow Hendricks (2002). The idea is to compare the estimated unobserved skills of immigrants with the source country income. First, wage regression is estimated on the native sample by including controls on observable characteristics. The earnings for given characteristics of immigrants are predicted using the estimates of the regression, then the residual earnings are computed by subtracting the predicted earnings from the observed earnings. The resulting value is the residual earnings of immigrants relative to natives. Given observable characteristics, the residual earnings show the difference in the unobserved skills of immigrants.

The procedure of estimating residual immigrant earnings relative to natives uses the following regression equation:

$$y_{it}^n = \alpha^n + \beta^s s_{it}^n + \beta^a D_{it}^n + \mu_t^n + \epsilon_{it} \tag{2.1}$$

where:

y_{it}^n - wage per hour;

s_{it}^n - schooling years;

D_{it}^n - a dummy variable that represents the following age groups (15–20, 21–25, ...

66+);

μ_t^n - time effect.

A superscript n on the variables means that the equation is estimated using only natives. Assuming that the return to education, age and time are the same for immigrants and natives, I compute predicted earnings from the above equation for all individuals in a sample. Then, by subtracting the predicted earnings from the actual earnings, I get residual earnings. Given individual residual earnings, I compute mean residual earnings by the source country from which the immigrants originate.

The resulting country-specific residual earnings are plotted in Figure 2.3. The figure plots residual immigrant earnings relative to natives on the y-axis and source country GDP per capita on the x-axis. The points on the figure represent the source country relative to the US. The residual earnings are plotted against relative per capita GDP of countries in 2000.[4] For example, the residual earnings of immigrants from Japan are on average 20% higher than the residual earnings of natives in the US, and GDP per capita is 30% lower than GDP per capita of the US. There are two takeaways from the graph. First, the skills differences of immigrants in the US have a positive relationship with earnings per capita. Second, residual earnings for most immigrants are below 0, meaning that the unobserved skills of most of the immigrants are below those of natives, which is not surprising since immigrants in the US come from countries where income is lower than in the US.

I compute residual earnings for immigrants in Brazil, Mexico, and Venezuela by using census data[5] for the countries. Figure 2.3 plots the earnings gap of immigrants in these countries relative to per capita income of the source countries. We see that earnings vary positively with GDP per capita in all host countries. The other takeaway from these plots is that immigrant earnings are higher than the earnings of natives. For example,

[4]If I measure source country GDP in the year before migration, I get the same results. The reason is that the slope between the weighted GDP per capita and GDP per capita in 1990 and 2000 is 0.96 and 0.90, respectively.

[5]Data from the following censuses are used: Brazil (1991, 2000, 2010), Mexico (1990, 2000, 2010) and Venezuela (1981, 1990, 2001).

the residual earnings of immigrants from Korea in Brazil are 80% higher and the residual earnings of immigrants from Argentina in Mexico are 60% higher than those of natives, given observable characteristics. Only the earnings of immigrants from Guatemala in Mexico and immigrants from Colombia and Grenada in Venezuela are equal to or lower than the earnings of natives in the respective host countries.

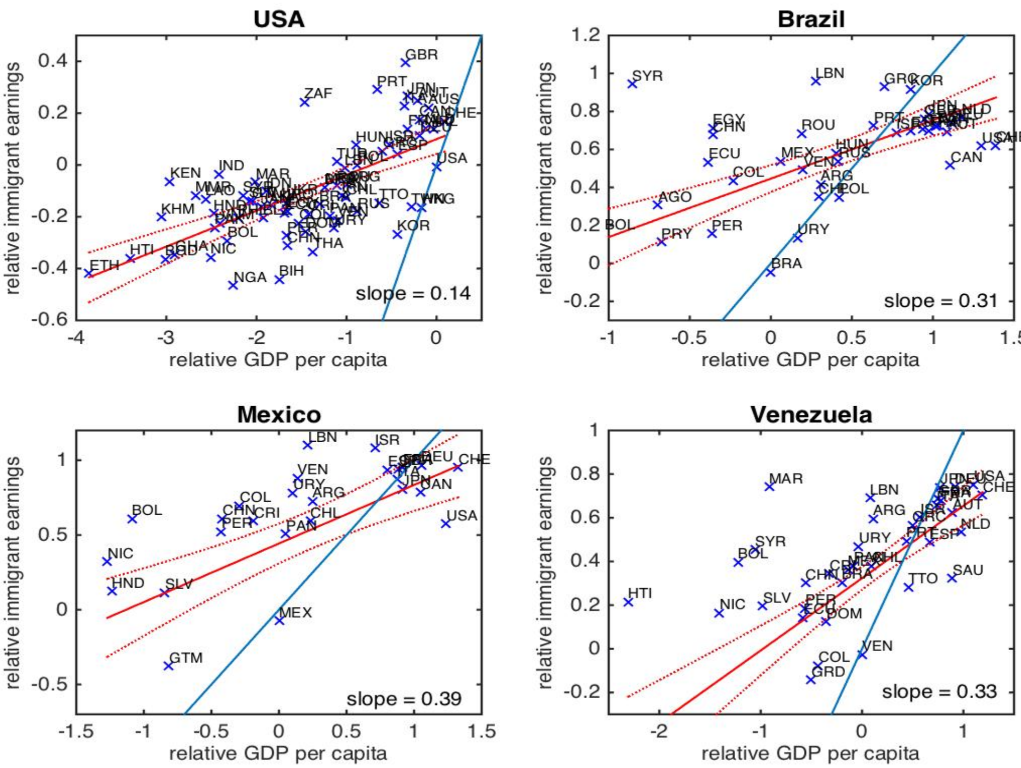


Figure 2.3: Earnings gap of immigrants relative to natives

If immigrants represent a random group of the source country population, then the estimated unobserved human capital of the immigrants measure the unobserved human capital of the source country worker. A positive selection in characteristics that increase the productivity of immigrants will result in higher unobserved skills than those of workers in the source country. As shown in Figure 2.2, immigrants in the US are positively selected on education, and the selection is higher for immigrants from low-income countries. A more positive selection of low-income country immigrants results in unobserved skills higher than those of workers in the source country. Figure 2.3 suggests that the unobserved human capital of immigrants estimated from immigrant earnings does not vary as much as the income per capita of the source countries. The slope of 0.14 in the US indicates that a 1% increase in relative income is associated with a 0.14% increase in relative residual skills.

In low-income countries, immigrants are much less selected on schooling and the differential selection of immigrants is less than in the US, which should result in larger differences in the unobserved human capital of immigrants. The slopes in the figure indicate that the gap in unobserved human capital across immigrants in low-income countries is larger than in the US. A 1% change in source country income is associated with a 0.31%, 0.39%, and 0.33% change in residual earnings of immigrants in Brazil, Mexico, and Venezuela, respectively. The high variation in immigrant earnings in low-income countries permits the assumption that the difference in human capital endowments across countries is much larger. This result supports the findings of Hendricks and Schoellman (2018).

2.7 Assimilation

Earnings between immigrants and natives converge over time. Duleep and Regets (1999 and 2002) have explained the convergence of earnings by different rates of earnings growth due to different rates of human capital accumulation. Immigrants from a developing country arriving at a developed country accumulate human capital more due to the lower opportunity cost of investing in human capital.

In this section I will investigate immigrant earnings growth relative to natives in low-income countries. In sub-section 2.7.1 I provide some theoretical background on the human capital investment model. In sub-section 2.7.2 I proceed with empirical analysis that shows the relative growth of immigrants' earnings in low-income countries. In sub-section 2.7.3 I analyze the occupational distribution of immigrants and investigate if immigrant earnings growth is accompanied by change in occupational distribution.

2.7.1 Human capital investment model

This section briefly describes the two-period human capital investment model developed by Ben-Porath.

Individual maximizes the following discounted two-period model:

$$\max_{i} w_t(1-i_t)h_t + e^{-rt}w_{t+1}[h_t(1-\delta)+(h_t i_t)^{\alpha}] \tag{2.2}$$

where i_t is the investment at time t, h_t is the stock of human capital at time t, δ is the depreciation rate of human capital and α is the rate of return to investment. The solution of the problem returns the following identity that relates investment to the stock of human capital:

$$i_t = \left(\frac{w_{t+1}}{w_t e^{-rt}\alpha}\right)^{\frac{1}{\alpha-1}} \frac{1}{h_t} \tag{2.3}$$

The model predicts that investment decreases with h_t. Thus immigrants in developed countries due to having lower skills relative to natives invest more in human capital than do natives. In studying immigrant earnings growth in the US, Duleep and Regets documented that immigrants have lower initial earnings, but higher earnings growth than natives. Immigrants coming to developed countries with low initial earnings have lower opportunity costs of investing in human capital than natives. Then immigrants accumulate more human capital than natives, which makes their earnings grow faster than those of natives.

2.7.2 Empirical analysis

Here I check if the results of the empirical analysis match the predictions of the human capital investment model. To investigate immigrant earnings in Brazil, I follow

Duleep et al. (2014). I estimate the following wage equation on data that pool immigrants and natives:

$$\begin{aligned} y_{it} &= \beta X_{it} + \theta s_{it} + \beta^{Exp} Exp_{it} + \beta^{Exp^2} Exp_{it}^2 \\ &+ \beta^{im} im + (\theta^{im} s_{it} + \beta^{imExp} Exp_{it} + \beta^{imExp^2} Exp_{it}^2) \times im + \epsilon_{it} \end{aligned} \tag{2.4}$$

where:

im - immigrant indicator;

X_{it} - observable characteristics other than education and experience;

s_{it} - years of schooling;

Exp_{it} - total potential experience;

The interaction of experience and schooling with immigrant dummy allows the estimation of a country-specific return to schooling ($\theta + \theta^{im}$) and a country-specific return to experience ($\beta^{Exp} + \beta^{Exp^2} + \beta^{imExp} + \beta^{imExp^2}$). I estimate the above equations separately for each census and for each year-of-entry cohort that can be followed from the immigrants' initial years in the host country. I define $a_{it}^{s,exp} = \beta^{im} + \theta^{im} s_i$ $+ \beta^{imExp} Exp_{it} + \beta^{imExp^2} Exp_{it}^2$ as the mean log earnings gap of the immigrant i in year t with schooling s_{it} and experience Exp_{it} relative to the native with similar observable characteristics. The changing of this earnings gap over time will show the earnings growth of immigrants relative to natives.

The wage equation 2.4 was first estimated for the cohort of immigrants aged 25–45 who entered the host country during the ten[6] years prior to a census. Then, using the censuses that were conducted after 10 and 20 years, I estimate the same equation for the same cohort but aged 35–55 and 45–65, respectively. I estimate the wage equation on samples from the US and Brazil. Figures 2.4 and 2.5 display entry earnings and growth of earnings for immigrants with average years of schooling in the US and Brazil, respectively. As shown in Figure 2.4, immigrants in the US initially earn lower than natives. The initial

[6]For the US I use cohort of immigrants that entered the US during the five years prior to census.

earnings of immigrants who arrived in 1975–1980 are 27% lower than those of comparable natives, and the gap shrinks over time to 5%. A similar pattern of assimilation is observed for immigrants who arrived in 1985–1990. Immigrants on arrival earned 25% less than natives with similar characteristics and the gap had decreased to 14%. This result is in line with the studies in developed countries that document higher earnings growth for immigrants than natives.

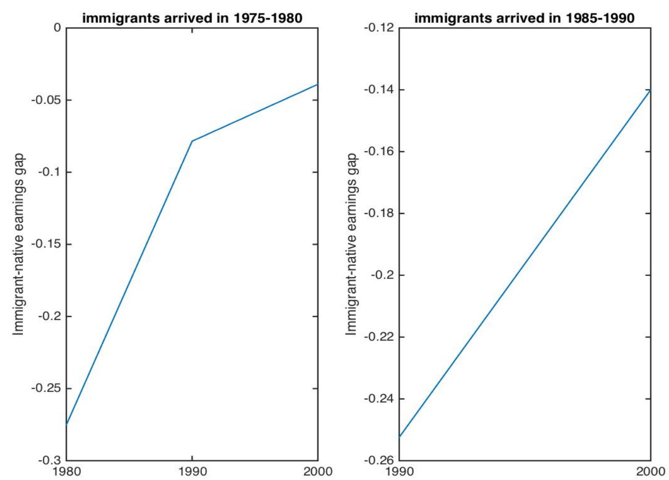


Figure 2.4: Assimilation of immigrants in the US

Using the 1990, 2000, and 2010 Brazilian census and equation 2.4 I estimate cross-sectional regression to investigate the earnings growth for immigrants and natives in Brazil. Figure 2.5 shows immigrant earnings relative to those of natives. Initial earnings of both immigrant cohorts are higher than those of the natives. The 1980–1991 arrival cohort with average years of schooling earn 30% more than natives with similar characteristics. After twenty years the earnings, gap between immigrants and natives has shrunk by 4% to 26%, meaning that the earnings growth of natives was higher than the earnings growth of immigrants. The relative earnings path of the 1990–2000 arrival cohort shows a similar trend. Immigrants initially earned 40% more than comparable natives and the gap had decreased to 31% over time. Thus, immigrants in Brazil earn more than natives, but the gap closes over time. This observation in Brazil supports the accumulation of human capital hypothesis: natives experience higher earnings growth than immigrants due to having a lower opportunity cost of investing in human capital.

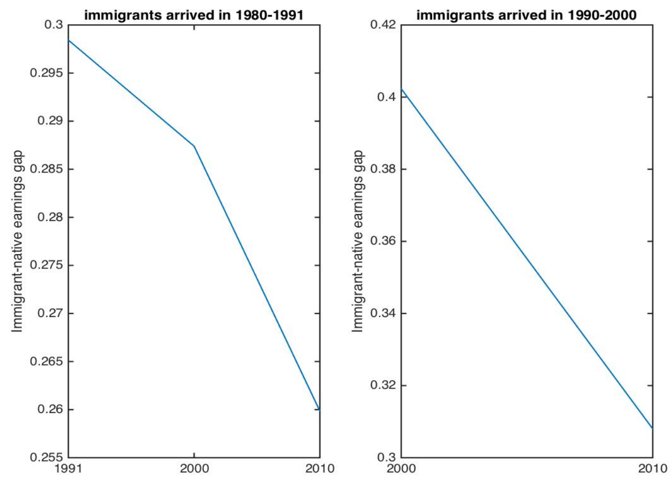


Figure 2.5: Assimilation of immigrants in Brazil

Selective out-migration can bias the results. The studies in developed countries document that the least skilled out-migrate at higher rates (e.g. Lubotsky (2007)). I check if there is a selective out-migration of immigrants in observables in Brazil by analyzing the change in average schooling and the fraction of highly educated immigrants as a given cohort stays in the country. Dividing the immigrant cohorts in the population by schooling, I compute the average years of schooling and the fraction of highly educated workers over time. Highly educated workers are defined as individuals with more than 12 years of schooling. The measures of education level of immigrant cohorts were adjusted by the source country weights.

In Appendix, Figure A1 compares the education of immigrant cohorts in Brazil. Average years of schooling of 1980–1990 arrival cohorts didn't change much with years of stay in Brazil, 12.12, 11.85, 12, respectively in 1990, 2000, and 2010. The fraction of highly educated 1980–1990 arrival cohorts slightly increased over time, 47%, 48%, and 50% in 1990, 2000, and 2010, respectively. Average years of schooling of 1990–2000 arrival cohorts slightly increased with years since migration, but not by much, 12.23 and

12.39 years of schooling in 2000 and 2010, respectively. The fraction of highly educated 1990–2000 arrival cohorts also increased over time, 55% in 2000 and 56% in 2010. A small increase in the average years of schooling for 1990-2000 cohorts and increase in the fraction of highly educated immigrants for both cohorts suggests the out-migration of lower educated immigrants. This would likely bias results in favor of higher wage growth for 1990-2000 arrival cohorts relative to natives in Brazil.

2.7.3 Occupational distribution and mobility of immigrants

In the previous section I showed that the gap in earnings between immigrants and natives closes over time. In the US, immigrants initially earn less than natives and over time immigrant earnings grow faster than those of natives. In Brazil, immigrants earn more than natives and the gap in earnings closes due to the higher growth of earnings of natives. Does the occupational distribution of immigrants also converge to that of natives? The studies in developed countries document that immigrants initially experience occupational downgrading, then over time they upgrade their occupations, meaning that they move to higher-paid occupations. This observation in developed countries supports the human capital accumulation model that immigrants with low opportunity cost of investing in human capital accumulate more human capital than natives. In this section I investigate if immigrants in low-income countries experience occupational mobility.

This section analyzes if the occupational distributions of immigrants and natives differ. Does the occupational distribution of immigrants change over time? I analyze the occupational distribution of immigrants relative to natives in the US and Brazil, Mexico and Venezuela. Then I conduct some counterfactual experiments that help me compare the occupational distribution of immigrants over time.

Occupational ranking

To compare the occupational distribution of immigrants and natives, I rank occupations. Following Hendricks and Schoellman (2018), I rank occupations by mean occupational earnings.[7] Mean occupational earnings are estimated from the following specification:

$$y_i^n = \alpha^n + \beta^n X_i^n + \sum_j \omega_j D_{ji}^n + \epsilon_i \tag{2.5}$$

where X^n is a vector of observable characteristics such as education, experience, sex, regional dummies, and marital status and D_{ji}^n is a dummy representing occupation j for individual i. From the above regression I estimate ω_j - occupational dummies, which is used to sort occupations.

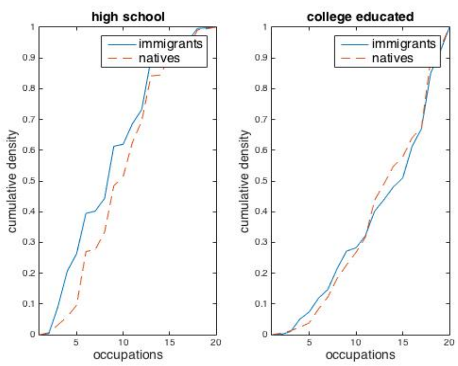
Occupational distribution

Now that I have a ranking of occupations, I can compare the occupational distributions of immigrants and natives. First I divide the sample into four age groups: 20–29, 30–39, 40–49 and 50+, and two education groups: individuals with high school degrees or less and individuals with college degrees. For each education and age cell I calculate occupational distributions across occupations for immigrants and natives p_i^{im} and p_i^n. I sort the occupations based on the occupational dummies estimated above and construct the cumulative densities for different groups of immigrants and natives.

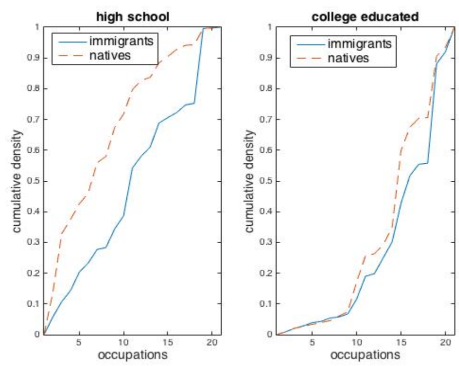
Figure 2.6 plots the densities for immigrants and natives aged 30–40 with high school and college degrees. The horizontal axis represents sorted occupations and the vertical axis represents cumulative density based on these occupations. The analysis of the occupational distributions of workers in the US shows that immigrants with high

[7] Chiswick et al. (2005), Chiswick and Miller (2008) and Akresh (2006) used the International Socio-Economic Index of Occupational Status (ISEI), which allows determination of the status of occupations. It is derived from the International Standard Classification of Occupations (ISCO) by using data on education, occupation, and income.

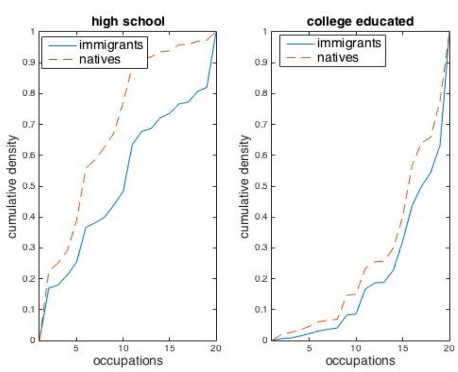
school degrees or less are concentrated in lower-paid occupations than natives with similar characteristics. Immigrants and natives with college degrees have close occupational distributions, which are concentrated in higher-paid occupations. In Brazil, Mexico, and Venezuela we see a different picture. In these countries, the distribution of immigrants in both groups first-order stochastically dominates the distribution of natives, meaning that immigrants work in higher-paid occupations than natives. There is a larger difference in occupational distribution between immigrants and natives with high school degrees or less than between college-educated immigrants and natives. Natives with college degrees are closer in occupational distribution to immigrants with comparable characteristics.



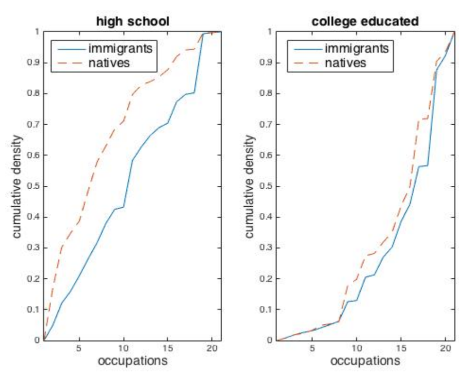
(a) USA



(b) Brazil



(c) Mexico



(d) Venezuela

Figure 2.6: Occupational distribution of immigrants

Occupational mobility

I showed that occupational distributions of immigrants and natives differ. Here I investigate if occupational distributions of immigrants relative to natives change over time. The studies in developed countries have documented that immigrants gradually upgrade their occupations relative to natives after accumulating human capital and investing in the host-country-specific skills. Do immigrants in low-income countries upgrade their occupations after arrival? To answer this question I conduct an experiment where I quantify the overall gain from the occupational distribution.

I compare occupational distributions of immigrants on arrival with the occupational distributions of the same cohort after some period in the host country. I estimate the change in mean occupational earnings between the two groups. The change in occupational mean earnings has two components: a change in composition and a change in distribution. I am interested in a change in distribution. To separate these effects I use the Oaxaca decomposition method:

$$\sum_c w_{c,t+1} y_{c,t+1} - \sum_c w_{c,t} y_{c,t} = \sum_c (w_{c,t+1} - w_{c,t}) * y_{c,t+1} + \sum_c (y_{c,t+1} - y_{c,t}) * w_{c,t} \qquad (2.6)$$

where $y_{c,t}$ is the mean log earnings relative to natives of the immigrant from country c in time t, and $w_{c,t}$ is the share of immigrants from country c in time t, which is available from the data. Mean log relative immigrant earnings $y_{c,t}$ is estimated from the occupational earnings relative to natives in the following way:

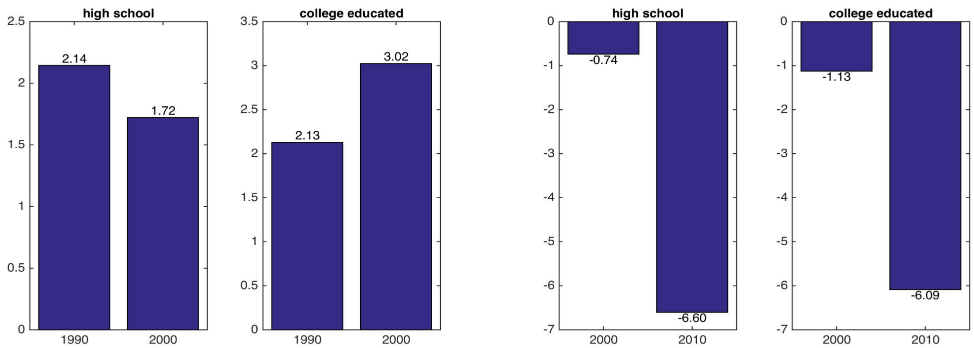
$$y_{c,t} = \sum_j (w_{c,j,t} - w_{n,j,t}) * y_{j,t} \qquad (2.7)$$

where $w_{c,j,t}$ is the share of immigrants from country c in occupation j in time t, and $y_{j,t}$ is the earnings in occupation j estimated from equation 2.5.

$\sum_c w_{c,t+1} y_{c,t+1} - \sum_c w_{c,t} y_{c,t}$ is the overall gain of an immigrant from staying in a host country for one more period, $\sum_c (w_{c,t+1} - w_{c,t}) * y_{c,t}$ is the component of the gain due to a change in country compositions and $\sum_c (y_{c,t+1} - y_{c,t}) * w_{c,t}$ is the component of the gain due to a change in the occupational distribution of immigrants. I am interested in the second component of the gain.

The computed gain in earnings which summarizes the difference in occupational distributions of immigrants is plotted in Figure 2.7. The wage gain for immigrants in the US that comes from changing the occupational distribution shows that immigrants in both education groups experienced occupational upgrading after arrival. The analysis of 1975–1980 arrivals shows that immigrants with a high school degree or less have an increase in wages: 2.14% and 1.72% increase after 10 and 20 years of stay in a host country, respectively. The same cohort of college-educated immigrants also experienced an occupational upgrading: 2.13% and 3.02% increase in average wages after 10 and 20 years of stay in a host country, respectively. This shows that after some time the occupational distribution of immigrants shifts towards higher-paid occupations.

The analysis of the occupational mobility of immigrants in Brazil shows a different picture. Figure 2.7 shows the gain in wages from changing the occupational distribution of immigrants in Brazil. Immigrants with high school degrees and college-educated immigrants experienced an occupational downgrading relative to natives, meaning that natives with similar characteristics move to higher-paid occupations. The counterfactual wage gain from changing the occupational distribution for immigrants with a high school degree or less is -0.74% and -6.6% after 10 and 20 years of stay in a host country, respectively. The counterfactual wage gain from changing occupational distributions for college-educated immigrants is -1.13% and -6.09% after 10 and 20 years of stay in a host country, respectively.



(a) USA (b) Brazil

Figure 2.7: Relative wage gain from changing the occupational distribution of immigrants

2.8 Conclusion

I estimate residual immigrant earnings for Brazil, Mexico, and Venezuela. I show that earnings of immigrants are steeper in low-income countries than in the US. One of the explanations of this is that immigrants in these countries are less selected on education than immigrants in the US. Investigation of selection on education in Brazil and Venezuela shows that the level of selection does not change with source countries.

Given the earnings gap between immigrants and natives, I investigate if the gap closes over time. Analysis of immigrants in Brazil shows that immigrants start with higher earnings relative to natives and that the gap between natives closes over the period. The occupational distribution of immigrants in Brazil shows that immigrants work in higher-paid occupations relative to natives. After spending some time in the host country immigrants in Brazil downgrade their occupations relative to natives. This result in low-income countries is consistent with the predictions of the human capital accumulation model: low-earning groups accumulate more human capital than high-earning groups.

REFERENCES

Akresh I., Occupational mobility among legal immigrants to the United States, The International Migration Review, 40 (4), 854–884., 2006.

Akresh I., U.S. immigrants' labor market adjustment: Additional human capital investment and earnings growth, Demography, 44 (4), 865–881., 2007.

Akresh I., Occupational trajectories of legal U.S. immigrants: Downgrading and recovery, Population and Development Review, 34:3 435–456, 2008.

Antecol H., Cobb-Clark D., and Trejo S., Immigration policy and the skills of immigrants to Australia, Canada, and the United States, Journal of Human Resources, 38 (1), 192–218., 2003.

Barrett A. and Duffy D., Are Ireland's Immigrants Integrating into its Labour Market?, Papers from Economic and Social Research Institute (ESRI), 2007.

Barro R. and Lee J., A new data set of educational attainment in the world, 1950-2010, Journal of Development Economics, 104, 184-198, 2013.

Blinder A., Wage Discrimination: Reduced Form and Structural Estimates, Journal of Human Resources. 8 (4): 436–455., 1973.

Borjas G., Assimilation, changes in cohort quality, and the earnings of immigrants. Journal of Labor Economics, 3, 463–489, 1985.

Borjas G., Self-Selection and the Earnings of Immigrants, American Economic Review, 77 (4), 531-553, 1987.

Borjas G., National origin and the skills of immigrants. In: Borjas, G.J., Freeman, R.B. (Eds.), Immigration and the Work Force. University of Chicago Press, Chicago, 1992a.

Borjas G., The slowdown in the economic assimilation of immigrants: Aging and cohort effects revisited again, Journal of Human Capital 9:4: 483–517, 2015.

Bratsberg B. and Ragan J., The impact of host country schooling on earnings: A study of male immigrants in the United States, Journal Human Resources, 37 (1), 63–105., 2002.

Chiswick B., The effect of Americanization on the earnings of foreign-born men, Journal of Political Economy, 897–922, 1978a.

Chiswick B., The economic progress of immigrants: Some apparently universal patterns, Contemporary Economic Problems, American Enterprise Institute, Washington, DC, 1979.

Chiswick B., An Analysis of the Economic Progress and Impact of Immigrants, Department of Labor monograph, N.T.I.S. No. PB80-200454, National Technical Information Service, Washington, DC., 1980.

Chiswick B., Lee Y. and Miller P., Family Matters: The Role of the Family in Immigrant's Destination Language Skills, Journal of Population Economics, 18(4), pp.631-647, 2005a.

Chiswick B., Lee Y. and Miller P., Parents and Children Talk: English Language Proficiency within Immigrant Families, Review of Economics of the Household, 3(3), pp.243-268, 2005b.

Chiswick B., Lee Y. and Miller P., Immigrant Earnings: A Longitudinal Analysis, Review of Income and Wealth, 51(4), pp.485-503, 2005c.

Chiswick B., Lee Y. and Miller P., A longitudinal analysis of immigrant occupational mobility: A test of the Immigrant Assimilation Hypothesis, International Migration Review 39:2: 332–353, 2005.

Chiquiar D. and Hanson G., International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States, Journal of Political Economy, pp. 239-281, 2005.

Cohen-Goldner S. and Eckstein Z., Labor Mobility of Immigrants: Training, Experience, Language, and Opportunities, International Economic Review, 2008.

Duleep H. and Dowhan D., Insights from longitudinal data on the earnings growth of U.S. foreign born men, Demography 39 (3), 485–506., 2002a.

Duleep H. and Dowhan D., Revisiting the Family Investment Model with Longitudinal Data: The Earnings Growth of Immigrant and U.S.-Born Women, IZA Discussion Paper No. 568, Institute for the Study of Labor, Bonn, Germany, 2002b.

Duleep H. and Regets M., Immigrant entry earnings and human capital growth, Research in Labor Economics, pp. 297-317, 1997.

Duleep H. and Regets M., Immigrants and human-capital investment, American Economic Review, 89:2: 186–191, 1999.

Duleep H. and Regets M., The Elusive Concept of Immigrant Quality: Evidence from 1970–1990, IZA Discussion Paper No. 631, Institute for the Study of Labor, Bonn, Germany, 2002.

Duleep H., Liu X., and Regets M., Country of Origin and Immigrant Earnings, 1960-2000: A Human Capital Investment Perspective, IZA Discussion Paper, 2014.

Dustmann C., Frattini T. and Preston I., The Effect of Immigration along the Distribution of Wages, Review of Economic Studies, 2013.

Friedberg R., The labor market assimilation of immigrations in the United States: The role of age at arrival, Unpublished Manuscript, Brown University, 1992.

Friedberg M., You can't take it with you? Immigrant assimilation and the portability of human capital, Journal of Labor Economics 18, 221–251, 2000.

Green D., Immigrant occupational attainment: assimilation and mobility over time, Journal of Labor Economics 17, 49–79., 1999.

Hendricks L., How Important Is Human Capital for Development? Evidence from Immigrant Earnings, American Economic Review, 92 (1), 198-219., 2002.

Hendricks, L. and Schoellman T., Human capital and development accounting: New evidence from wage gains at migration, The Quarterly Journal of Economics, Volume 133, Issue 2, 2018.

Hsieh C., Hurst E., Jones C. and Klenow P., The allocation of talent and U.S. economic growth. mimeo, University of Chicago-Booth, 2013.

Jain A. and Peter K., A Joint Hazard-Longitudinal Model of the Timing of Migration, Immigrant Quality, and Labor Market Assimilation, IZA Discussion Paper, 2017.

Lagakos D., Moll B., Porzio T., Qian N., and Schoellman T., Life-cycle wage growth across countries, Unpublished Manuscript, Princeton University, 2016.

Lubotsky D., Chutes or Ladders? A Longitudinal Analysis of Immigrant Earnings, Journal of Political Economy, 115 (5), 820-867, 2007.

Mattoo A., Neagu I., and Özden Ç., Brain waste? Educated immigrants in the U.S. labor market, World Bank Policy Research Working Paper, vol. 3581. Washington, D.C., 2005.

Manuelli R. and Seshadri A., Human Capital and the Wealth of Nations, American Economic Review, 104 (9), 2736-2762., 2014.

McKenzie D. and Rapoport H., Self-Selection Patterns in Mexico US Migration: The Role of Migration Networks, Review of Economics and Statistics 92 (4), 811-821., 2010.

Orrenius P. and Zavodny M., Self-selection among undocumented immigrants from Mexico, Journal of Development Economics, 78 (1), pp. 215-240., 2005.

Oaxaca R., Male-Female Wage Differentials in Urban Labor Markets, International Economic Review, 14 (3): 693–709., 1973.

Schoellman T., Education Quality and Development Accounting, Review of Economic Studies, 79 (1), 388-417. 2012.

Schoellman T., Refugees and Early Childhood Human Capital, American Economic Journal: Macroeconomics, forthcoming.

Schoeni R., New evidence on the economic progress of foreign-born men in the 1970s and 1980s, Journal Human Resources, 32, 683–740, 1997.

Zorlu A., Occupational adjustment of immigrants in the Netherlands, Journal of International Migration and Integration 14:4: 711–731., 2013.

Villen P., Immigrant labor in Brazil (2007—2015): The legal and the undocumented circuits , RUDN Journal of Economics, 2017.

Winkelmann R., Immigration: The New Zealand experience. In: Zimmermann, K.F., European Migration-What Do We Know? Oxford University Press, 2005.

APPENDIX

Code	Occupations names	Code	Occupations names	Code	Occupations names
1	Executives, Administrative and Managerial	23	Information clerks	45	Electronic repairer
2	Management related	24	Records processing, non-financial	46	Misc. Repairer
3	Architects	25	Records processing, financial	47	Construction
4	Engineers	26	Office machine operator	48	Extractive
5	Math and Computer science	27	Computer and communication equipment operator	49	Precision Production, Supervisor
6	Natural science	28	Mail distribution	50	Precision Metal
7	Health Diagnosing	29	Scheduling and distributing clerks	51	Precision Wood
8	Health Assessment	30	Adjusters and Investigators	52	Precision Textile
9	Therapists	31	Misc. Administrative support	53	Precision other
10	Teachers, Postsecondary	32	Private household occupations	54	Precision, food
11	Teachers, Non-Postsecondary	33	Firefighting	55	Plant and System
12	Librarians and curators	34	Police	56	Metal and Plastic Machine Operator
13	Social scientists and Urban planners	35	Guards	57	Metal and Plastic Processing Operator
14	Social, Recreation and religious workers	36	Food preparation and service	58	Woodworking machine operator
15	Lawyers and Judges	37	Health service	59	Textile machine operator
16	Arts and Athletes	38	Cleaning and building service	60	Printing machine operator
17	Health technicians	39	Personal service	61	Machine machine operator
18	Engineering technicians	40	Farm managers	62	Fabricators
19	Science technicians	41	Farm non-managers	63	Production inspectors
20	Technicians, other	42	Related agriculture	64	Motor Vehicle Operator
21	Sales	43	Forest, fishers and hunters	65	Non-Motor Vehicle Operator
22	Secretaries	44	Vehicle mechanic	66	Freight and material handlers

Table A1: Occupational coding

Table A1 reports information on aggregated categories and their coding. Categories were aggregated on analyzing the IPUMS 1990 Census Bureau occupational classification scheme and Hsieh et al. (2013) data available on authors webpage.

Aggregated code	Broader occupational categories	Disaggregated code
1	Executives, administrative, and managerial	1, 2, 40
2	Architects, engineers, lawyers	3, 4, 15
3	Math, and computer science	5, 6, 13
4	Nurses, therapists, and other health service	7, 8, 9, 37
5	Technicians	17, 18, 19, 20
6	Teachers	10, 11
7	Recreation, religious, arts, athletes	12, 14, 16
8	Administrative support, clerks, record keepers	22, 23, 24, 25, 26, 27, 28, 29, 30, 31
9	Food, cleaning, and personal services and private household	32, 38, 39
10	Fire, police, and guards	33, 34, 35
11	Food prep.	36, 54
12	Farm, related agriculture, logging	41, 42, 43
13	Mechanics	44, 45, 46
14	Construction and extraction	47, 48
15	Precision manufacturing	49, 50, 51, 52, 53
16	Manufacturing operators	55, 56, 57, 58, 59, 60, 61, 62, 63
17	Vehicle operators	64, 65, 66
18	Sales	21
19	Home	67

Table A2: Broader occupational categories

year	variable	N.obs.	mean	sd	min	max
1991	age	964173	37.0	10.0	23.0	60
	educ		5.0	4.5	0.0	16
	hours		43.1	10.8	1.0	100
	exper		26.0	11.8	1.0	54
	lwage*		11.1	1.1	4.1	16.6
2000	age	1204520	38.5	10.3	23.0	60
	educ		5.9	4.6	0.0	16
	hours		43.4	12.8	1.0	100
	exper		26.6	12.0	1.0	54
	lwage		5.9	1.0	0.0	12.6
2010	age	1530715	39.2	10.6	23.0	60
	educ		7.3	5.0	0.0	16
	hours		40.8	12.2	1.0	100
	exper		25.9	12.7	1.0	54
	lwage		6.7	0.9	0.0	13.3

*lwage defines log of earnings

(a) Brazil

year	variable	N.obs.	mean	sd	min	max
1993	age	244197	36.3	11.1	20.0	60
	educ		4.6	5.3	0.0	16
	exper		25.7	12.9	0.0	54
	lwage		5.3	1.1	0.0	10.0
1999	age	256537	36.4	11.0	20.0	60
	educ		5.2	5.5	0.0	16
	exper		25.1	12.9	0.0	54
	lwage		6.0	1.0	2.3	12.6
2004	age	268612	36.7	11.0	20.0	60
	educ		5.5	5.4	0.0	16
	exper		25.2	13.0	0.0	54
	lwage		6.2	1.0	2.3	11.8

*lwage defines log of earnings

(b) India

Table A3: Summary statistics

	1991	2000	2010
All	5.5%	6.3%	10.4%
white men	8.5%	8.7%	12.1%
white women	7.7%	9.4%	17.1%
brown men	1.8%	1.8%	3.8%
brown women	1.9%	2.3%	6.8%

(a) Brazil

	1993	1999	2004
All	6.8%	8.1%	7.8%
Other men	11.5%	13.0%	12.3%
Other women	5.5%	7.0%	6.4%
Scheduled tribe men	3.5%	4.7%	7.0%
Scheduled tribe women	1.2%	1.9%	2.4%
Scheduled caste men	2.5%	3.9%	5.1%
Scheduled caste women	0.5%	1.3%	1.8%

(b) India

Table A4: Share of college-educated

	1991	2000	2010
in labor force:			
employed, not specified	684,648	0	0
at work	0	802,335	1,051,513
have job, not at work	0	35,591	43,666
not in labor force:			
inactive	0	366,792	435,902
housework	285,185	0	0
unable to work, disability	8,569	0	0
in school	8,843	0	0
living on rents	2,891	0	0
retired	33,416	0	0
pensioner	14,662	0	0
Total	1,038,214	1,204,718	1,531,081

(a) Brazil

	1993	1999	2004
in labor force:			
at work	164,079	169,877	183,889
have job, not at work	2,186	2,753	2,123
not in labor force:			
housework	77,793	83,779	81,583
permanent disability	1,184	1,563	2,020
temporary illness	439	464	416
in school	6,588	7,589	7,941
retirees and living on rents	1,154	1,566	1,987
Total	253,423	267,591	279,959

(b) India

Table A5: Activity status

Table A6 and A7 report results of regression of log wage on group dummies and other explanatory variables for Brazil and India. Regression is run separately for each year.

	1991	2000	2010
white women	-0.313	-0.283	-0.255
	0.006	0.005	0.005
brown men	-0.225	-0.252	-0.178
	0.005	0.005	0.005
brown women	-0.519	-0.481	-0.441
	0.006	0.006	0.005
schooling	0.106	0.102	0.08
	0.001	0	0
experience	0.047	0.043	0.034
	0.001	0.001	0.001
experience^2	-0.001	-0.001	0
	0	0	0
(Intercept)	5.515	0.306	1.282
	0.015	0.013	0.013
occup. Dummies	Yes	Yes	Yes
Number of observations	141099	169064	209953
R-squared:	0.51	0.51	0.4
F-test	2025.44	2445.2	1946.39

*Dependent variable is log wages

Table A6: Regression results for Brazil

	1993	1999	2004
other women	-0.336	-0.329	-0.311
	0.003	0.003	0.003
sch. caste men	-0.209	-0.192	-0.209
	0.004	0.003	0.003
sch. caste women	-0.492	-0.489	-0.483
	0.004	0.004	0.003
schooling	0.076	0.078	0.079
	0	0	0
experience	0.021	0.023	0.026
	0	0	0
experience^2	-0.0001	-0.0001	-0.0001
	0	0	0
(Intercept)	6.181	6.901	7.09
	0.01	0.009	
occup. Dummies	Yes	Yes	Yes
Number of observations	216540	226560	242580
R-squared:	0.67	0.75	0.74
F-test	6117.76	9197.53	9758.43

*Dependent variable is log wages

Table A7: Regression results for India

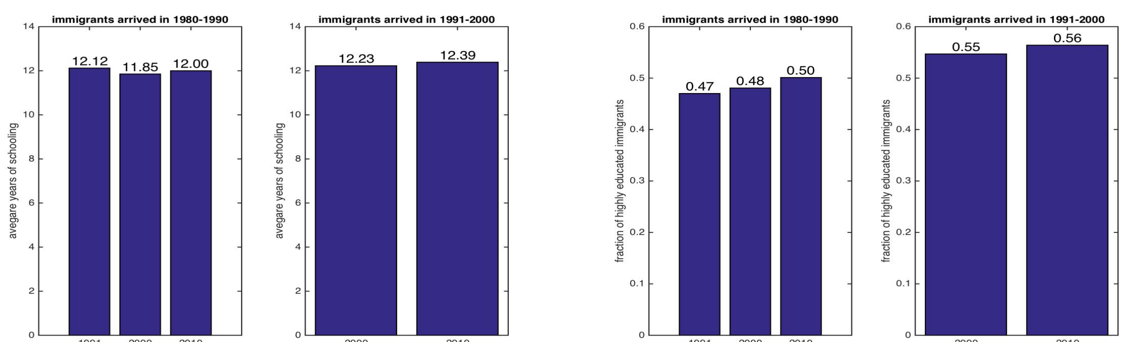
code	**Names of occupational categories**
1	Executives, Administrative, and Managerial
2	Architects, Engineers, Lawyers
3	Math, and Computer Science
4	Nurses, Therapists, and Other Health Service
5	Technicians
6	Teachers
7	Recreation, Religious, Arts, Athletes
8	Administrative Support, Clerks, Record Keepers
9	Food, Cleaning, and Personal Services and Private Household
10	Fire, Police, and Guards
11	Food prep.
12	Farm, Related Agriculture
13	Mechanics
14	Construction and extraction
15	Precision Manufacturing
16	Manufacturing Operators
17	Vehicle Operators
18	Sales
19	Home

Table A8: Occupational categories

	1980-1990	1990-2000
im	0.746	0.22
	0.18	0.136
educ_a	0.139	0.13
	0.002	0.002
educ_im	0.001	0.024
	0.008	0.006
exper	0.09	0.07
	0.005	0.004
exp_2	-0.002	-0.001
	0	0
exp_im	-0.069	-0.019
	0.018	0.015
expim_2	0.002	0.001
	0.001	0
(Intercept)	9.241	3.919
	0.067	0.056
Regional Dummies	Yes	Yes
Number of observations	9529	11164
R-squared:	0.4	0.41
F-test	575.46	707.66

*Dependent variable is log wages

Table A9: Regression results (Brazil)



(a) Average years of schooling

(b) Fraction of highly educated immigrants

Figure A1: Schooling of immigrants over time