ESSAYS IN IMMIGRATION ECONOMICS AND EDUCATION

Gloria Allione

A DISSERTATION

in

Economics

Presented to the Faculties of the University of Pennsylvania

in

Partial Fulfillment of the Requirements for the

Degree of Doctor of Philosophy

2019

Supervisor of Dissertation

Dirk Krueger,
Walter H. and Leonore C. Annenberg
Professor in the Social Sciences and
Professor of Economics

Co-Supervisor of Dissertation

Petra E. Todd,
Edmund J. and Louise W. Kahn Term
Professor of Economics

Graduate Group Chairperson

Jesús Fernández-Villaverde, Professor of Economics

Dissertation Committee:

Iourii Manovskii, Professor of Economics

Dedicated to Francesco

Acknowledgements

I am deeply indebted to my main advisor Dirk Krueger for his patience, support and encouragement. I always looked up to him for guidance, first as Graduate Chair and then as my thesis supervisor, and he never failed to give me very valuable advice. I would like to express my gratitude to my co-advisor Petra Todd for her precious insights and helpful suggestions. I also want to thank Yourii Manovskii for his kindness and willingness to take part in my dissertation committee.

I was lucky to meet Rebecca Stein, with whom I worked as teaching assistant for three years. She gave me the opportunity to help her developing the first Mass Open Online Course in Microeconomics Principles at the University of Pennsylvania, from which we derived our joint paper. Her enthusiasm for my ideas and appreciation for my work helped me gain more confidence in my skills. I will always be thankful to her.

I would like to thank Kelly Quinn for always coming to the rescue and Gina Conway for helping me before even meeting me.

I owe a special thank to Andrea Brandolini, the Head of Statistical Analysis Directorate where I work at the Bank of Italy. Having my best interests at heart, he strongly encouraged me to finish my PhD, giving me the last push needed to conclude my dissertation.

My years in Philadelphia would not have been the same without my friends. Yumi Koh and Mauricio Calani shared with me all the good and bad moments of my PhD since the very beginning: we made a good team. Dove McCrary was my flatmate during my most difficult year at Penn: she lifted my spirits so many times with her smile and quirkiness. Chiara Margaria, Alessandro Arlotto, Alberto Ciancio, Fabrizio Lecci and Stefano Tracà were my Italian family here in the United States: every Thanksgiving we spent together made me miss home a little bit less.

I am extremely grateful to my parents and family for always making me feel their unconditional love. The last and foremost words of affection are for my husband Francesco. I would have never come to the end of this journey if it weren't for you. Thanks for all the sacrifices you patiently endured, for taking such a good care of me and always believing I could make it.

ABSTRACT

ESSAYS IN IMMIGRATION ECONOMICS AND EDUCATION

Gloria Allione

Dirk Krueger

Petra E. Todd

Historically, the United States have been the meeting point for people, ideas and knowledge from all over the world. The first two chapters of my dissertation focus on the composition and impact of the U.S. immigrants; while the last chapter describes a new way for U.S. universities to export education worldwide, analyzing the case of a Mass Open Online Course.

In Chapter 1, I study how visa policies may affect the immigrant distribution in terms of both observable skills, as education, and unobservable traits, as ability or motivation. I develop a model of migration choice, estimated via simulated method of moments using the New Immigrant Survey dataset. I find that wage differentials, combined with skill transferability, act as the main force driving self-selection. Positive self-selection on unobservables is stronger among non-college graduates. The dynamic framework gives the opportunity to simulate the immigrant distribution resulting from alternative visa policies. In this work, I apply the visa-point system proposed by the RAISE Act to the model. The results suggest that the scheme is successful in attracting migrants with the desired observable characteristics but it reduces the positive selection on unobservable skills.

In Chapter 2, I use a firm-level database on high-skill temporary visa (H-1B) applications to study the effects of H-1B workers on patents. To address endogeneity, I construct an instrumental variable by interacting H-1B quotas with the number of

Asian incumbent inventors within the firm. An increase in H-1B applications has a positive impact in terms of firm patent activity, future new inventors and patent activity by incumbents. It also leads to a higher diversity in citations by country and class, as evidence of positive spillovers.

Chapter 3 (joint with Rebecca Stein) describes our Mass Open Online Course in microeconomics principles. Using the Cox proportional hazard model, we relate the high attrition rate to demographic data, finding that younger students, U.S. participants and females are less likely to complete the course. The results provide useful insights for future MOOCs: more flexibility may benefit time-constrained students but it is important to keep the structure and sense of community, both instrumental to completion.

Contents

Acknowledgements iv

Abstract vi

Contents viii

List of Tables x

List of Figures xii

1 Immigrant Self-Selection and Visa Policies 1

1.1 Introduction 1

1.2 Related Literature 4

1.3 A Brief Overview of the U.S. Visa System 6

1.4 Data 8

1.5 Model 12

1.5.1 Preliminaries 12

1.5.2 Model Specifications 14

1.6 Estimation Strategy 19

1.7 Results 23

1.8 Policy Experiment 36

1.8.1 Experiment Design 36

1.8.2 Experiment Results 40

1.9 Conclusions 42

2 The Impact of H-1B Visa on U.S. Innovation **45**

2.1 Introduction 45

2.2 Data 48

2.3 Empirical Framework 51

2.3.1 Analysis by Assignee 51

2.3.2 Analysis by Incumbent Inventor 55

2.3.3 Citation Analysis 57

2.4 Results 58

2.5 Conclusions 62

3 Mass Attrition: An Analysis of Drop Out from Principles of Microeconomics MOOC **64**

3.1 Introduction 64

3.2 Model and Results 73

3.3 Conclusions 84

A Appendix to Chapter 1 **88**

A.1 On multiple migration opportunities 88

Bibliography **90**

List of Tables

1.1 Distribution of employment-based immigrants over visa categories . . 7
1.2 Summary Statistics 11
1.3 Parameters set exogenously 24
1.4 Probability of winning the diversity lottery 24
1.5 Parameters of the home-country wage function 26
1.6 Parameters of the U.S. wage function 27
1.7 Parameters of the probability of a U.S. job offer 28
1.8 Parameters of the migration cost function 30
1.9 Model fit - Migrants' distribution by origin and education 34
1.10 Model fit - Share of EB migrants by origin 34
1.11 Share of high-type migrants by origin and education 35
1.12 Share of high-type migrants by origin and visa category 36
1.13 Visa point system under the RAISE Act 38
1.14 Policy experiment - Migrants' distribution by origin and education . . 40
1.15 Policy experiment - Share of migrants with a job offer by origin . . . 42
1.16 Policy experiment - Share of high-type migrants by origin and education 43

2.1 Impact of H-1B workers on patents by assignee 59
2.2 Impact of H-1B workers on the number of new inventors 60
2.3 Impact of H-1B workers on incumbent inventors (extensive margin) . 61
2.4 Impact of H-1B workers on incumbent inventors (intensive margin) . 62
2.5 Impact of H-1B workers on the diversity index of citations 63

3.1 Statistics on course activity 67

3.2 Statistics on first-week activity 69
3.3 Demographic data 71
3.4 Video and quiz retention by first-week activity 75
3.5 Video and quiz retention by survey participation 77
3.6 Video retention by demographics (focus on Reasons for taking the course) 79
3.7 Video retention by demographics (focus on Age and Occupation) . . . 80
3.8 Quiz retention by demographics (focus on Age and Occupation) . . . 85

List of Figures

1.1 Model fit - Migrants' age distribution by origin 31

1.2 Model fit - Migrants' wage distribution in their home country 32

1.3 Model fit - Migrants' wage distribution in the U.S. 33

1.4 Policy experiment - Migrants' age distribution by origin 41

Chapter 1

Immigrant Self-Selection and Visa Policies

1.1 Introduction

Every year, around 1 million foreigners are lawfully admitted for permanent residence in the United States, obtaining what is commonly known as a 'green card'. According to the 2010 Census, the share of foreign born over the total U.S. population is 12.9%. Given the relevance of the phenomenon, many studies have investigated the economic impact of immigration (Kerr and Kerr, 2011). This has motivated U.S. policy-makers to design visa policies more and more selective. In particular, the idea of introducing a visa-point scheme similar to the one already applied in other countries as Canada or Australia, has come back several times over the last twenty years. However, the immigrant population is the combined result of two forces, out-selection and self-selection. Hence, any policy must take into account that: a) there are unobservable characteristics, such as ability, motivation or soft skills (Heckman and Kautz, 2012), which are difficult to control for; b) the policy itself may change the incentive to migrate, differently for immigrants with different characteristics.

To understand the effects of a visa-point system on the composition of U.S. green card holders, I develop a life cycle model of migration from the rest of the world to the United States. I focus on migration for work reasons so I include in the model two of the current U.S. visa categories: the employment preference and the diversity lottery. In case of the employment-based visa, the foreign worker receives a job offer from the United States, which he can decide whether to accept or reject. For the diversity visa, the immigrant can choose whether to apply to the lottery or not. In the model, the acceptance of a U.S. job offer or the winning of a diversity visa implies migrating to the United States, without the possibility of returning to their home country. Everything being equal, the ex-ante probability of accepting a U.S. job offer is higher than the probability of applying to the lottery because the diversity migrant enters the United States without the certainty of being employed nor the knowledge of its prospective wage. Foreign workers are different in terms of observable characteristics, such as age, education or country of origin, and unobservable skills, low versus high types.

I use the New Immigrant Survey data to estimate the model parameters through the simulated method of moments. The survey is a micro-level dataset, collecting present and past information among the cohort of new legal immigrants admitted between May and November 2003. In particular, it includes retrospective questions about wages and occupation status before leaving the home country or immediately upon migration, in addition to a rich set of demographic, economic and social variables about the current status. Since the migration cost depends also on variables that do not enter the wage function, such as distance to the United States, I am able to identify the value and distribution of unobservable skills by matching simulated moments about home-country and U.S. wages and the probability of migrating to their data equivalent.

Migrants coming from the OECD high-income countries, Eastern Europe and Asia are positively selected in terms of unobservables. In particular, the proportion of high types is larger among uneducated workers. Since they do not benefit from the gain in the college premium, positive self-selection is stronger for them to compensate their lower incentive to migrate. Skills, both observable and unobservable, are not perfectly transferable for natives of Latin America, Africa and Middle East, so wage differentials encourage the uneducated and low-type workers to migrate more than the others. In the meanwhile, though, the high unemployment rates in some source countries of Sub-Saharan Africa give to the educated high-type individuals a greater incentive to move to the United States, counterbalancing some of the negative selection.

In order to assess the effect of a change in the visa system on the distribution of U.S. immigrants, I run a counterfactual experiment and apply the point allocation scheme proposed by the bill S. 1720, introduced in the Senate in 2017. The scheme is successful in attracting mostly educated, young workers from the OECD high-income or the East and South Asia countries. However, the overall proportion of high-type immigrants decreases. The effect is driven by the increased share of Canadian workers among migrants. Being contiguous to the United States, their migration cost is small and so the proportion of low types that find beneficial to migrate is larger than in the countries sharing the same wage differentials. The proposed visa system gives additional points to highly compensated job offers but removes the need of employer sponsorship. This particular feature, included in the counterfactual experiment, results in the vast majority of migrants entering the United States without employment. However, in East and South Asia, where the percentage of workers moving consequently to a U.S. job offer is the largest, the points awarded to highly compensated job positions are effective in selecting the high types. For the remaining regions – Eastern Europe and Central Asia, Latin America, Africa and Middle East – the probability of being admitted under this point allocation scheme is so small that, when given the possibility of migrating, both low and high types take it, without waiting for another opportunity. Hence, for them the positive or negative selection on unobservables, detected in the baseline model, lessens.

The paper is structured as follows. Section 1 provides a review of the related literature. In Section 2, I draw a brief picture of the American visa system, with a particular focus on the employment-based and the diversity visa categories. Section 3 describes the New Immigrant Survey dataset. The model is presented in Section 4. Section 5 explains the estimation strategy and Section 6 contains the estimation results. Section 7 provides a description and analysis of the policy experiment. Section 8 draws some concluding remarks.

1.2 Related Literature

Since the seminal paper of Borjas (1987), there has been great interest in the topic of migrants' self-selection. Most of the literature builds around the Roy model, that predicts positive selection from source countries where the rate of return to skills is lower than in the destination country and negative selection when it is higher. However, the evidence suggests that highly educated workers are over-represented also in the emigrant population from less developed countries contradicting some of the model predictions. Grogger and Hanson (2011) propose a linear utility model where absolute wage differentials can explain positive selection from poorer countries. Wealth constraints are another possible explanation for low migration rates among less skilled workers, as argued by Chiquiar and Hanson (2005) who find evidence of intermediate selection among Mexican emigrants to the United States, consistent with a negative selected distribution truncated from below. I construct a modified Roy model in a dynamic framework, which allows me to take into account also the option value of the migration choice (Burda, 1995; Locher, 2001). The possibility of moving to the U.S. becomes also an insurance against future bad shocks, so workers may choose to postpone their migration.

Most recent studies (Borjas *et al.*, 2018; Bertoli *et al.*, 2016) analyse self-selection in terms of unobservables characteristics. Borjas *et al.* (2018) use administrative data from Denmark to decompose the self-selection in total earnings between observables and unobservables. They conclude that unobservables play the dominant role and, under certain conditions, may reverse the self-selection based only on observables. My model can capture the same mechanism by including unobservables discrete types and see whether it is supported by the U.S. data.

A parallel strand of the literature has studied how immigration policies affects the self-selection of migrants. Bianchi (2013) identifies two different components: a size effect and a composition effect, working in opposite directions. At lower levels of restrictions, the composition effect may dominate the size effect so that increasing the migration cost leads to higher skill distribution among migrants. Bianchi argues that, in light of these conclusions, there may be no need for skill-dependent policies as visa-point systems, which are difficult to tailor, but migration costs can be as much effective in screening the migrant pool. Stark *et al.* (2017) compare the impact of an entry fee versus a quota when the receiving country is interested in both the quality of migrants and their occupation. They conclude that a differentiated entry fee induces positive self-selection among migrants and allows to control their share by occupation. Bertoli *et al.* (2016) investigate how immigration policies targeting observable characteristics may influence the immigrant composition in terms of unobservables. In case of positive self-selection on the unobservables, increasing the selectivity of migrants on observables may lead to a decrease in their overall productivity to the point that its better to admit an uneducated high-type migrant rather than an educated low-type one. Through my estimated model, I can run a counterfactual experiment with a visa-point system to study this same question empirically, not only from a theoretical point of view.

Ortega and Peri (2013) use a dataset on annual bilateral migration flows covering

15 OECD destination countries and 120 sending countries for the period 1980-2006 to establish, among other results, that migration policies do play a role in determining the size of the flows. Jasso and Rosenzweig (2009) compare employment immigrants in Australia, which applies a visa-point system, and the United States, finding that the selection mechanism has little impact on the characteristics of skill migration. However, the two countries attract a different pool of migrants, given their geographic location, so my counterfactual experiment may lead to further insights. A very recent study by Chassamboulli and Peri (2018) uses a search model in general equilibrium, calibrated to the migrant flows between the U.S. and the rest of the world, to analyze the economic effects of changing immigration policies. The paper studies the impact of reducing the size of particular visa categories, shifting to others, for example from family reunification to employment-based. Differently from my work, it does not get into the details of the selection mechanism of work immigrants so the two studies may be considered complementary to each other.

1.3 A Brief Overview of the U.S. Visa System

The architecture of the current U.S. immigration system is built on the Immigration Act of 1990. The act established two preference categories for permanent immigrants: family-sponsored and employment-based. Each category is subject to a worldwide limit: between 226,000 and 480,000 for family preference visas and 140,000 for employment-based. The exact value is determined annually, depending on the number of preference visas in the previous fiscal year. There is also a per-country annual limit, equal to 7% of the total. Immediate relatives of U.S. citizens are not subject to any numerical limitation but they contribute to the family-sponsored limit for the following year.[1]

On average, employment-based preferences count for 12% of the total number of visas. They include 5 different categories: priority workers; professionals with advanced degrees or of exceptional ability; skilled workers, professionals (without advanced degrees), and needed unskilled workers; special immigrants (e.g., ministers, religious workers, and employees of the U.S. government abroad); and investors. Table 1.1 shows the distribution of employment-based (EB) preferences across categories in the fiscal year 2003, including accompanying spouses and children who participate to the numerical limit. Two thirds of the EB visas are allocated to the second and third category. In the vast majority of cases, applicants must be sponsored by their U.S. employer, hence possessing a job offer prior to petition. Generally, a labor certification issued by the Secretary of Labor is also needed to attest that there are not sufficient U.S. workers available for the job position in question and that wages and working conditions of similarly employed U.S. workers will not be negatively affected by the employment of the foreign worker.

Table 1.1: Distribution of employment-based immigrants over visa categories

EB1 - Priority workers	15.89%
EB2 - Professionals with advanced degree or of exceptional ability	19.91%
EB3 - Skilled workers, professionals, unskilled workers	56.36%
EB4 - Special immigrants	7.79%
EB5 - Investors	0.05%

The Act of 1990 also established the Diversity Immigrant Visa Program (DV), known as 'green card lottery', to promote immigration from underrepresented countries. To enter the lottery, applicants must have been born in an eligible country, i.e. one with fewer than 50,000 admissions over the preceding five years, and have

[1]The family preference limit is the minimum between 226,000 and 480,000 minus the previous year's total of immediate relatives of U.S. citizens.

at least a high school degree or two years of qualifying work experience. Originally, the number of visas allocated through the Diversity Program was 55,000 but it was reduced to 50,000 in 1999. The visas are distributed among six geographic regions according to a formula based on immigrant admissions during the preceding 5 years and the total population of the region. Each country can receive up to 7% percent of the available DVs in one year. DV winners are then randomly selected by region, with no difference between countries belonging to the same region, except for the aforementioned limit. Table 1.4 presents the list of countries declared ineligible for the 2003 Lottery. The remaining visas are allocated to refugees, asylees and other special categories which account for few admissions. In general, applicants may be already living in the United States as temporary workers, foreign students, refugees, or sometimes even undocumented. If so, we talk about adjustments of status to legal permanent residence; while, for applicants living abroad we talk about new arrivals.

1.4 Data

The New Immigrant Survey (NIS) is a longitudinal dataset targeting all the immigrants admitted to the United States through a permanent visa between March and November 2003. It includes both new arrivals and adjustments of status. For the Adult Sample[2], 12,500 immigrants were drawn from the U.S. administrative records, representing 4.3% of the entire population. The response rate was 68.6% for a total of 8,573 cases. Acknowledging the research interest in studying and comparing immigrants with different visas, the sample was stratified to allow for representativeness across the main visa categories. Four strata were defined: spouses of U.S. citizens,

[2]The NIS includes also a Child Sample, covering immigrants younger than 18 who entered as child-of-U.S.-citizen or adopted orphans under five years of age. The sample is not relevant for my research purpose.

employment-preference immigrants, diversity lottery winners and other immigrants. Employment-based immigrants and diversity lottery winners were over-sampled, at respectively two and three times their frequency, making up 16.5% and 13.5% of the sample. Conversely, spouses of U.S. citizens were under-sampled at half their natural rate, being 16.5% of the sample. The survey questions the new legal permanent residents about a variety of topics (family, schooling, health, earnings, etc.) but what makes the dataset incredibly valuable in light of my research question is the information about their careers at home and in the United States. Having retrospective data on their home country wages allows me to identify the unobserved types among migrants and thus estimate self-selection. They were asked about:

- the first job they had in their home country (occupation, industry, hours worked, earnings when they *first* started the job,...);

- the last job they had in their home country (occupation, industry, hours worked, earnings just *before* they left job,...): the information was recorded only if they had another job other than the first one;

- the first job they had in the U.S. (occupation, industry, hours worked, earnings when they first started the job,...);

- the current job at the time of the interview (2003 and follow-up 2007).

The project comprised a follow-up interview in 2007-2009 but the response rate was particularly low, 46.1% of the previous respondents. Nevertheless, Massey *et al.* (2017) provides weights effective in correcting for the nonresponse bias, so I use the second round data to fill the missing values for some variables of the first round (for example, the wage of their last job at home in case it was not recorded in the first round) and to capture the wage progression in the U.S.

Given that the model makes some simplifying assumptions, I need to select a sub-sample of the available cases. First, I consider only employment-based preference (EB) and diversity (DV) principal[3] migrants, since I do not study family migration.[4] I also leave out female migrants, not modeling the labor choice. The dataset includes also the adjustments of status, so half of the sample consists of migrants who arrived in the U.S. prior to 2003. I set the year of migration at the time of their first job in the United States.[5] As for the age span, I consider only those who migrated between 18 and 64 years. I define educated immigrants as those with at least a bachelor degree, earned in their home country. Since I take education as exogenous, I need to exclude all the migrants who received their bachelor degree in the United States:[6] those represent 3.78% of my remaining sample.

The resulting sample counts 1,551 observations, with diversity visa slightly outnumbered by employment-based visas. Table 1.2 shows some summary statistics. The sub-samples by visa category are very different in terms of country of origin and education. Most DV migrants come from Eastern Europe, Central Asia and Sub-Saharan Africa, while the majority of EB migrants are from East and South Asia and OECD high-income countries. The difference in provenience stems from two factors: the ineligibility of some countries for the diversity lottery, as India or Canada, and the different probability of getting a job offer in the United States. Latin America

[3] I exclude accompanying spouses and children.

[4] For family migration, see Gemici (2011), who estimates a dynamic model of internal migration with intra-household bargaining using U.S. data, and Lessem (2018), who estimates a model where the migration decision depends on the location of the spouse using survey data from the Mexican Migration Project.

[5] If sooner, I use the year of last entry where last entry stands for the last time they entered the United States without leaving for more than 60 days.

[6] The question in the survey is about the highest degree received so in case of degrees higher than bachelor and earned in the United States, I cannot directly identify those who graduate from college in their home country. However, I observe the number of years of schooling spent in their home country: if higher than 16, I consider them to have received their bachelor before migrating and I keep them in the sample.

Table 1.2: Summary Statistics

	Visa category		
	Diversity	**Employment-based**	**Total**
	%	%	%
Country of origin			
OECD high-income countries	4.1	23.8	14.8
Eastern Europe & Central Asia	29.2	5.0	16.1
East Asia & South Asia	11.1	51.7	33.1
Latin America	3.4	12.9	8.6
Africa Sub-Saharan	38.6	3.2	19.4
Middle East & North Africa	13.5	3.4	8.1
Total	100.0	100.0	100.0
Highest degree received			
None or lower than HSD	11.3	10.7	11.0
High School Degree	20.0	4.0	11.3
Associate Degree	5.8	1.8	3.6
Bachelor Degree	29.3	35.6	32.8
Master Degree	12.3	30.3	22.1
PhD	2.1	9.0	5.9
JD/MD	1.7	2.3	2.0
Unspecified	17.5	6.3	11.4
Total	100.0	100.0	100.0
Age at migration			
18-21	6.6	5.2	5.9
22-25	19.5	20.4	20.0
26-29	22.7	26.2	24.6
30-33	17.2	15.7	16.4
34-40	18.9	19.0	19.0
41-50	11.4	10.6	11.0
51-64	3.7	2.9	3.2
Total	100.0	100.0	100.0
Arrival year			
1965-1980	0.0	0.6	0.3
1981-1990	1.1	8.9	5.4
1991-1995	1.3	19.5	11.2
1996-2000	7.3	59.5	35.7
2001-2004	90.3	11.5	47.5
Total	100.0	100.0	100.0
Adjustment of status			
New arrival	92.9	21.9	54.4
Adjusted status	7.1	78.1	45.6
Total	100.0	100.0	100.0
N	709	842	1,551

accounts only for 8.6% of the entire sample, in contrast with the overall statistics where Mexico alone makes 20% of the total admissions. This comes from leaving out family preference visas. Not surprisingly, EB migrants are on average more educated than DV migrants: most of EB visa categories require a bachelor degree. In terms of age distribution, there are no striking difference between the two sub-samples: half of migrants come to the United States before thirty while arrivals after the age of 50 are quite rare, at around 3%. As mentioned before, the sample includes both new arrivals and adjustments of status: if the former are the vast majority of DV visas, the latter represent almost 80% of the EB migrants, with 411 new legal permanent residents adjusting their status from a temporary work visa. As for the arrival year, most migrants entered, permanently, the United States after 1995. For the EB sample, this is coherent with the maximum length of stay of a temporary work visa, which is three or six years, depending on the occupation.[7]

1.5 Model

1.5.1 Preliminaries

The economy is divided in two regions: the United States and the Rest of the World. Workers from the Rest of the World can migrate to the United States through two alternative channels: being sponsored by a U.S. employer for an employment-based visa or applying to the diversity visa lottery. Once they migrate, they cannot go back to their homeland.

From the model perspective, the two visa processes differ under two aspects: the

[7] I observe some new-arrival immigrants whose last year of entry is before 2003. They were probably still registered as resident abroad when they applied for a permanent visa. This explains the inconsistency between the higher share of new arrivals with respect to the percentage of arrivals in the period 2001-2004.

labor status of the worker in the first period upon migration and the informational structure of the decision problem. In case of the employment-based visa, the worker must receive a U.S. job offer so (a) he is surely employed in the first period after migrating and (b) he observes his prospective U.S. wage before making his decision. Instead, under a diversity visa, the worker needs to spend some time searching for a job after arriving in the United States so (a) he may be unemployed in the first period and (b) he obviously does not know his future U.S. wage when choosing whether to migrate.[8] The assumption is consistent with the statistics in my sample where only 8% of the DV migrants working at the time of the interview were offered their current job before moving to the United States, against 81% of the EB migrants.[9]

The timeline is as follows: in each period, after observing their home-country labor status and income shock, workers receive a U.S. job offer with some probability p, which they can accept, moving to the United States through an EB visa, or reject, staying in their home country; then, only if they did not get a U.S. job offer, they can apply to the diversity lottery and, in case of winning with probability q, enter the United States with a DV visa. In both scenarios, the individuals who migrate incur a one-time migration cost, which they observe at the time of their decision.

[8]The model by Borjas (1987) assumes perfect knowledge of both domestic and foreign wages, as in the EB case; while Bertoli (2010) introduces uncertainty in the migration decision through the same mechanism used in the DV case: workers observe the shock realization of their domestic wage but not the foreign one. He argues that the different informational structure negatively affects the theoretical predictions on the migrants' self-selection in unobservables.

[9]Since migrants were only asked about their current job in this question, and it may not be the first one they got upon migration, I considered only those whose arrival year matches the year they started their current job.

1.5.2 Model Specifications

State Space and Initial Conditions

The two regions are indexed by $m \in \{0 : \text{Rest of the World}, 1 : \text{United States}\}$. Workers are heterogeneous in their level of education $e \in \{U, E\}$ and their country of birth b, each country belonging to a specific set B_i, where $i = 1, \ldots, 6$. In addition, they are characterized by two possible unobservable types $\theta \in \{L, H\}$, who permanently differ in terms of their skills (Heckman and Singer, 1984; Keane and Wolpin, 1997). The initial type probability, $p(\theta = H)$, is equal to 0.5 and it is independent on the education level or country of origin.[10] The state space at time t is:

$$\Omega_t = \{S, h_{0,t}, h_{1,t}, l_t, y_t, M_t\}$$

where $S = \{b, e, \theta\}$ denotes the set of time-invariant state variables and $M_t \in \{0, 1\}$ is the worker's location at age a, so it is equal to one if the worker has already migrated to the United States and zero otherwise. Age a and time t are perfectly exchangeable with $a = t + 18$ and $t = 0, \cdots, 46$ so age is omitted from the state variables.

The individual's labor market experience $h_{m,t+1}$ at time $t + 1$ is determined by his work experience $h_{m,t}$, labor status l_t and location M_t in the previous period. I distinguish between experience acquired in the home country, $m = 0$, or in the U.S. $m = 1$.

$$h_{m,t+1} = \begin{cases} h_{m,t} + 1 & \text{if } l_t = 1 \text{ and } M_t = m \\ h_{m,t} & \text{otherwise} \end{cases}$$

with $h_{m,0} = 0$ and $m \in \{0, 1\}$.

[10]The assumption is very strong but it is necessary since I do not observe wage data for non-migrants.

At time $t+1$, the worker is unemployed with probability $\delta_{M_{t+1}}(b)$.

$$l_{t+1} = \begin{cases} 0 & \text{w.p. } \delta_{M_{t+1}}(b) \\ 1 & \text{w.p. } (1 - \delta_{M_{t+1}}(b)) \end{cases}$$

Since the level of schooling is taken as exogenous, the labor status of educated workers is set equal to zero for their college years $l_t(e = E) = 0$ with $t = 0, \cdots, 3$.[11] In case of unemployment, income y_{t+1} is equal to a constant $\underline{y}$; otherwise, it follows the home-country or U.S. wage function $w_{M_{t+1}}(\cdot)$, depending on the worker's location.

$$y_{t+1} = \begin{cases} \underline{y} & \text{if } l_{t+1} = 0 \\ w_{M_{t+1}}(S, h_{0,t+1}, h_{1,t+1}) & \text{if } l_{t+1} = 1 \end{cases}$$

Wage Functions

At this stage, both home-country and U.S. wages are taken as given: I implicitly assume the flow of migrants is not large enough to impact prices.

In the home country, the wage function depends on education, work experience and type. The variable GDP per worker is added to take into account the different country productivities:

$$\log\, w_0(b, e, \theta, h_0) = \mu_0^0 + \mu_1^0 \log(\text{GDP per worker}) + \mu_{2,B_i}^0 \mathbb{1}\{e = E\} + \\ + \mu_{3,B_i}^0 h_0 + \mu_{4,B_i}^0 h_0^2 + \mu_{5,B_i}^0 \mathbb{1}\{\theta = H\} + \varepsilon_0$$

where $b \in B_i$ and $\varepsilon_0 \sim \mathcal{N}(0, \sigma_0^2)$. The coefficients on education, experience and type

[11] Usually, in case of exogenous education, the initial age is set at around 22 years old so to avoid these adjustments. However, the structure of my problem does not allow me to do the same: first, 6% of the migrants in my dataset moved to the United States before they were 22 so I would need to exclude them from the sample; second, since most non-college graduates start working before 22, so I would not be able to use their home-country first wage observations in my estimation.

differ by country group B_i.

The U.S. wage function shares a similar structure but separates home-country and U.S. work experience to allow for the possible depreciation of the former. Also, it does not need to include the variable GDP per worker, being all wages from the same country production function.

$$\log\, w_1(b,e,\theta,h_0,h_1) = \mu_0^1 + \mu_{2,B_i}^1 \mathbb{1}\{e=E\} + \mu_{3,B_i}^1 h_1 + \mu_{4,B_i}^1 h_1^2 + \\ + \mu_{5,B_i}^1 h_0 + \mu_{6,B_i}^1 h_0^2 + \mu_{7,B_i}^1 \mathbb{1}\{\theta=H\} + \varepsilon_1$$

where $b \in B_i$ and $\varepsilon_1 \sim \mathcal{N}(0,\sigma_1^2)$.

The different coefficients on education, experience and type, at country group level, allow for different wage dynamics, consistently with the findings in the literature. Several empirical studies state that skill transferability is imperfect and heterogeneous across migrants (Jasso *et al.*, 2002; Chiswick *et al.*, 2005; Chiswick and Miller, 2012). Using the NIS dataset, Lessem and Sanders (2013) shows that returns to US experience are higher for countries with lower GDP.

The home-country and U.S. income shocks are independent from each other.

Migration Opportunity

Since the model does not include the education choice, I prevent college graduates from migrating before the end of their studies, by setting the probability of receiving a U.S. job offer or winning the diversity lottery both equal to zero for the first four periods of their life ($e = E$ and $t < 4$).

In general, the probability of receiving a job offer from a U.S. employer depends on the level of education, the home-country experience and the country of origin:

$$p_t(b, e, h_0 \,|\, e = U \text{ or } t \geq 4)$$
$$= \Big(1 + \exp\big(\alpha_{0,B_i} + \alpha_1\, \mathbb{1}\{e = E\} + \alpha_2\, h_0 + \alpha_3\, h_0^2 + \alpha_4\, \mathbb{1}\{b \in B_{eng}\}\big)\Big)^{-1}$$

where $b \in B_i$. The constant is equal across countries belonging to the same group B_i, but it differentiates according to whether English is the official language ($b \in B_{eng}$).

The probability of winning the diversity lottery is equal to:

$$q_t(b \,|\, e = U \text{ or } t \geq 4) = \begin{cases} 0 & b \text{ is not an eligible country} \\ q_{D_i} & b \in D_i \text{ is an eligible country} \end{cases}$$

where $\{D_i\}_{i=1,\ldots,6}$ are the six geographical regions defined by the Immigration and Nationality Act for the purposes of the Diversity Immigrant Visa Program.

Migration Cost

The migration cost is measured in utility units as follows:

$$\Delta_t(b, e) = \kappa_{0,b} + \kappa_1 e + \kappa_2 t + \kappa_3 t^2 + \xi$$

where $\xi \sim \mathcal{N}(0, \sigma_\xi^2)$.

The country-specific component $\kappa_{0,b}$ depends on several variables: distance to the United States, contiguity, existence of colonial ties, GDP per capita, immigrants from the same source country already leaving in the U.S. and whether English is the official language. This set of variables is commonly used in the gravity models analyzing international trade flows (Ortega and Peri, 2013). The geographical variables are from the GeoDist Database (Mayer and Zignago, 2011).

Value Functions

Utility is expressed in logarithmic terms. Once the worker has migrated to the United States, there is no decision left to make because the model does not consider return migration. Hence, his value function is:

$$V_t^1(\Omega_t) = \log(y_t) \,+\, \beta \mathbb{E} V_{t+1}^1(\Omega_{t+1})$$

On the other hand, workers still living in their home country are called, in every period, to choose whether or not to move to the United States, after incurring a migration cost $\Delta_t(b,e)$. In case of a U.S. job offer, $j = 1$, the migration decision implies accepting the offer; otherwise, $j = 0$, it translates into applying to the diversity lottery.

$$W_t(\Omega_t|\, j, w_1) = \max\big\{ V_t^0(\Omega_t)\,;\, \widetilde{V}_t^1(\Omega_t|\, j, w_1) - \Delta_t(b,e) \big\}$$

When making his location choice, the employment-based migrant is aware he will be employed at the prospective wage w_1 in his first period in the United States, so:

$$\widetilde{V}_t^1(\Omega_t|\, j = 1, w_1) = V_t^1(S, h_{0,t}\,, h_{1,t}\,, l_t = 1, y_t = w_1)$$

Conversely, the diversity migrant may be unemployed with probability δ_{DV} and does not observe his labor status nor his wage if employed. Hence,

$$\begin{aligned}\widetilde{V}_t^1(\Omega_t|\, j = 0) = (1 - \delta_{DV})\, \mathbb{E} V_t^1\big(S, h_{0,t}\,, h_{1,t}\,, l_t = 1, y_t = w_1(\Omega_t)\big) \\ + \delta_{DV}\, V_t^1\big(S, a_t, h_{0,t}, h_{1,t}, l_t = 0, y_t = \underline{y}\big)\end{aligned}$$

The agent choosing to stay in his home country will face three possible scenarios in the next period: receiving a U.S. job offer with probability p; not receiving a U.S.

job offer but applying and winning a diversity visa with probability $(1-p)\cdot q$; not receiving a US job offer nor winning a diversity visa with probability $(1-p)\cdot(1-q)$. Therefore, his value function is:

$$
\begin{aligned}
V_t^0(\Omega_t) = \log(y_t) \; + \; \beta\Big(& p_{t+1}(\,.\,)\,\mathbb{E}W_{t+1}(\Omega_{t+1}|\, j=1, w_1) + \\
& + \big(1-p_{t+1}(\,.\,)\big)\, q_{t+1}(\,.\,)\,\mathbb{E}W_{t+1}(\Omega_{t+1}|\, j=0) + \\
& + \big(1-p_{t+1}(\,.\,)\big)\big(1-q_{t+1}(\,.\,)\big)\,\mathbb{E}V_{t+1}^0(\Omega_{t+1})\Big)
\end{aligned}
$$

The worker's problem can be solved recursively, starting from the last period T when the continuation values are equal to zero ($V_{T+1}^m = 0$). The distributional assumptions on the wage and migration cost functions lead to closed-form solutions of the expected value of the shocks, conditional on the optimal migration choice.

1.6 Estimation Strategy

Estimation is carried by simulated method of moments (SMM). Specifically, the SMM parameter estimates are chosen to minimize the weighted average distance between a set of sample moments and their simulated analogs. The weights are given by the inverse of the estimated variances of the sample statistics.

My wage observations are limited to U.S. immigrants who became legal permanent residents in 2003. I do not observe workers who never left their origin country ($M_T = 0$), but only those who moved to the United States at some point in their life ($M_T = 1$). Then, the conditional expected value of their wages is equal to:

$$
\mathbb{E}[\log w_{m,t}|X, M_T = 1] = X\mu^m + \mu^m_{\text{type},b}\cdot p(\theta = H|M_T = 1) + \mathbb{E}[\varepsilon_{m,t}|M_T = 1]
$$

where X collects all the worker's observable characteristics and m characterizes his location, i.e. home country, $m = 0$, or Unites States, $m = 1$. Since the income shock is independent over time, his conditional expectation depends only on the contemporaneous migration choice. Self-selection on unobservable skills can be detected in the conditional probability of being a high-type worker $p(\theta = H|M_T = 1)$. In absence of self-selection, it is equal to the unconditional probability $p(\theta = H)$. Otherwise, it may be greater, if self-selection is positive, or smaller, if it is negative. The conditional probability depends on the policy function, in so far as incentives to migrate are different between low and high types. Since wages enter the policy function, identification requires some exogenous variable affecting the migration choice but not the salary: in this case, distance to the United States, included in the migration cost, serves as instrument.

In my simulation, I use Barro and Lee (2013) dataset on educational attainment as initial distribution. The data include 146 countries from 1950 to 2010. They are disaggregated by sex, 5-year age intervals and four level of schooling: no formal education, primary, secondary and tertiary education. College graduates are drawn from the tertiary education pool while uneducated workers from the residual male population.

The estimation process requires to draw some individuals from each cohort, simulate the arrival of a job offer from the United States or the winning of the diversity lottery, and construct the model aggregate statistics by selecting only those who choose to migrate in 2003. However, this would entail the unnecessary simulation of workers who would never even have the chance of migrating. Given the very low probability of winning the diversity lottery or receiving a job offer, it would be computationally challenging to draw a number of individuals sufficient to end up with a satisfying sample for the simulation moments. Hence, I adopt a different approach, by restricting the initial distribution to "potential immigrants", i.e. individuals who get at least one opportunity of migrating through an employment-based or diversity visa. I also consider the 2003 cohort to be representative of the entire U.S. permanent immigrant population, since my model does not separate the age and time dimension.

My estimation procedure develops as follows:

1. Draw potential immigrants from the sampling distribution $g(S, a, v)$, where g is chosen to assure sub-samples by region and education large enough to compute reliable aggregate statistics.[12]

2. Simulate potential immigrants' life, assuming each of them gets one chance to migrate through a visa $v \in \{\text{EB,DV}\}$ at age a.[13] In terms of the model, this translates into the arrival of a job offer, for the EB sub-sample, or the unanticipated winning of the diversity lottery, in case of application, for the DV sub-sample.

3. Discard all the individuals who decide to stay in their home country.

4. Construct the sampling weights, as the likelihood ratio between the true probability distribution $f(.)$ and the sampling distribution $g(.)$:

$$\omega(S, a, v) = \frac{f(S, a, v)}{g(S, a, v)}$$

where the probability of being a potential EB immigrant from country b, with

[12] For instance, I sample potential DV migrants twice as much as potential EB migrants, given that they have a lower propensity to migrate.

[13] In the appendix, I discuss the correction needed to account for multiple migration opportunities.

education e and age a, is equal to:[14]

$$f(S, a, EB) = p(S, a) \cdot f_{BL}(S, a)$$

while the probability of being a potential DV immigrant is:

$$f(S, a, DV) = q(S, a) \cdot (1 - p(S, a)) \cdot f_{BL}(S, a)$$

The term $f_{BL}(S, a)$ is the worlwide population distribution by country, education and age in 2000 according to the Barro and Lee (2013) dataset.

5. Compute the simulated moments:

$$\tilde{m}(\vartheta) = \frac{\sum_i (\tilde{m}_i(\vartheta) \cdot \omega_i)}{\sum_i \omega_i}$$

6. Choose the parameters ϑ such that:

$$\min_{\vartheta} \, [m - \tilde{m}(\vartheta)]' W [m - \tilde{m}(\vartheta)]$$

where W is a diagonal matrix with the inverse of the estimated variances of the sample moments m.

I draw four million potential migrants and I end up with around 35,000 effective migrants at the estimated parameters. I use 339 moments to estimate 65 parameters. The aggregate statistics employed in the estimation process consists of three

[14]The probability of receiving a job offer depends also on work experience, so a transformation is needed:

$$p(b, e, a) = \sum_{h_0} p(b, e, a, h_0) Pr(h_0 | b, e, a)$$

Home-country work experience, conditional on age, country and education, follows a binomial distribution $B\big(n(e, a), p_h(b)\big)$ with parameters $n(e, a) = \max\big(a - 18 - 4 \cdot \mathbb{1}\{e = E\}\big)$ and $p_h(b) = \delta_0(b)$.

groups. Home-country wage moments include the first and last wage before migration by region and education, work experience categories and set of countries grouped by geographical location, so to exploit the differences in distance to the United States. U.S. wage moments consists of the first and current wage (2003 and follow-up 2007) by region and education, U.S. work experience categories, home-country work experience categories and set of countries grouped by geographical location. Finally, the migrant distribution moments include the proportion of workers for every of the above statistics, the ratio of DV winners by region and education, their age at arrival and the share of EB and DV migrants who unsuccessfully applied for a diversity visa before. The overall probability of migration is used to identify the size of the migration cost.

1.7 Results

Some of the parameters are set exogenously (Table 1.3). The discount factor β is equal to $1/1.04 = 0.9615$. The home-country unemployment rates, $\delta_{0,b}$, come from the ILOSTAT database and they refer to male labor force. I set δ_{DV}, the probability of being unemployed in the first year upon migration under the diversity visa, to be equal to 0.2263, which is the share of unemployed DV migrants among new arrivals in my sample. The foreign-born unemployment rate in the U.S., δ_{US}, is 0.063 (OECD data). Given the structure of my dataset, I do not observe many migrants with long work experience, neither in the U.S. nor in their home country. In the United States, even with the adjustments of status, I have very few cases that have been working for more than fifteen years. In the origin country, I face a similar issue since most migrants left before they were forty. Hence, I impose a threshold to the number of work experience years h counted in the home-country and U.S. wage function and I replace h_m, for $m = \{0, 1\}$, with $\tilde{h}_m = \min(h_m, H)$ and $H = 22$.

Table 1.3: Parameters set exogenously

β	0.9615
$\log(\underline{y})$	0
$\delta_{0,b}$	country unemployment rate
δ_1	0.063
δ_{DV}	0.2283
$q(b)$	see Table 1.4
$H = \max(h)$	22
σ_ξ	1

Given that this is a partial equilibrium model, I also need to determine exogenously the probability of winning the diversity visa lottery, otherwise I would not be able to identify the migration cost. For my purpose, I use the ratio between winners and on-line applicants by diversity region in the DV-2007 Lottery, which is the earliest year for which data by the U.S. Department of State are available.

Table 1.4: Probability of winning the diversity lottery

Region	**Probability**
Africa	0.0113
Asia	0.0044
Europe	0.0105
South America	0.0063
North America	0.0061
Oceania	0.0401
Ineligible countries	0

Countries ineligible in 2003 and 2004: Canada, China (mainland born), Colombia, Dominican Republic, El Salvador, Haiti, India, Jamaica, Mexico, Pakistan, the Philippines, South Korea, the United Kingdom and dependent territories, and Vietnam. Persons born in Hong Kong SAR, Macau SAR, Taiwan, and Northern Ireland were eligible to apply for the DV-2003 and DV-2004 lottery.

I define six group of countries B_i: OECD high-income countries, Eastern Europe[15] and Central Asia, East and South Asia, Latin America, Sub-Saharan Africa, North Africa and Middle East. Since I do not have enough wage observations to separately identify the last three groups, I assume that Latin America, Africa and Middle East share the same parameters in the home-country and U.S. wage function.

Tables 1.5–1.8 report the estimated parameters and their asymptotic standard errors. In the home country wage-function (Table 1.5), the college premium is between 0.40 and 0.60 for all four groups, except for Easter Europe and Central Asia where it is considerably lower. The same is true also for the type premium. On the other hand, the coefficient on experience is much higher than in the other areas. The standard deviation of the income shock is quite large, which is not surprising given that observations from very different countries are pooled together in the same wage function, so high variability is to be expected.

In the United States (Table 1.6), the college premium is higher than in the home country for all four groups, except Latin America, Africa and Middle East. Similarly, high-type migrants from these regions see their wage premium decrease upon migration, while the opposite is true for high-type workers from OECD high-income countries, Eastern Europe and Asia. Thus, the results show a different skill transferability across regions, consistently with other studies (Jasso and Rosenzweig, 2009). The coefficient on U.S. experience is very close to the one in their home-country for migrants belonging to the OECD high-income group while it is much higher for the others. In case of skilled workers from Latin America, Africa and Middle East, it means that over time they might offset some of their initial disadvantage, thanks to a better job match or the acquisition of new skills (Lessem and Sanders, 2013). The

[15] The exact title should be non-OECD European countries, however for sake of clarity I use Eastern Europe since the two mainly correspond to each other.

Table 1.5: Parameters of the home-country wage function

		OECD high-income	Eastern Europe & Central Asia	East & South Asia	Latin America, Africa & Middle East
constant (μ_0^0)	6.3324				
	(0.0116)				
log(GDP per worker) (μ_1^0)	0.1725				
	(0.0004)				
education (μ_{2,B_i}^0)		0.5391	0.1461	0.4188	0.5686
		(0.0021)	(0.0007)	(0.0012)	(0.0740)
experience (μ_{3,B_i}^0)		0.0966	0.1920	0.0362	0.0748
		(0.0003)	(0.0005)	(0.0001)	(0.0002)
experience sqrd (μ_{4,B_i}^0)		-0.0025	-0.0049	-0.0004	-0.0003
		(0.0000)	(0.0000)	(0.0000)	(0.0003)
type (μ_{5,B_i}^0)		1.8582	0.2709	1.5197	1.3408
		(0.0046)	(0.0021)	(0.0044)	(0.0049)
sd. income shock (σ_0)	2.1190				
	(0.0072)				

Asymptotic standard errors in parentheses

high return to U.S. experience could also be related to further education pursued by the migrants once in the United States, for example a Master: since my model only discriminates between college and non-college graduates. I exclude from the sample those who got their bachelor degree after migrating, but I do not account for other specialization degrees. Interestingly, the coefficient on home-country experience is relevant only for the OECD high-income group. The standard deviation of the income shock is much lower than in the home-country wage function since in this case there is no cross-country variability.

Table 1.7 shows the parameters governing the probability of receiving a job offer from a U.S. employer. The constant on Eastern Europe and Central Asia is very low, compared to the other countries. Education has a positive coefficient, as expected given that most of the EB visa categories require a bachelor degree to apply.

The estimated migration cost is remarkably high: the average for those who mi-

Table 1.6: Parameters of the U.S. wage function

		OECD high-income	Eastern Europe & Central Asia	East & South Asia	Latin America, Africa & Middle East
constant (μ_0^1)	7.9973				
	(0.0093)				
education (μ_{1,B_i}^1)		0.6838	0.4270	0.7552	0.4871
		(0.0023)	(0.0025)	(0.0018)	(0.0634)
US exp. (μ_{2,B_i}^1)		0.0983	0.3844	0.2054	0.2931
		(0.0003)	(0.0006)	(0.0003)	(0.0004)
US exp. sqrd (μ_{3,B_i}^1)		-0.0016	-0.0190	-0.0072	-0.0128
		(0.0000)	(0.0000)	(0.0000)	(0.0000)
HC exp. (μ_{4,B_i}^1)		0.0358	0.0146	0.0095	0.0123
		(0.0002)	(0.0001)	(0.0000)	(0.0129)
HC exp. sqrd (μ_{4,B_i}^1)		-0.0006	-0.0006	-0.0006	0.0000
		(0.0000)	(0.0000)	(0.0000)	(0.0007)
type (μ_{6,B_i}^1)		2.2791	1.0858	2.0603	0.7999
		(0.0045)	(0.0018)	(0.0071)	(0.0037)
sd. income shock (σ_1)	1.2263				
	(0.0048)				

Asymptotic standard errors in parentheses

Table 1.7: Parameters of the probability of a U.S. job offer

constant (α_{0,B_i})	
OECD high-income	-8.2638
	(0.0762)
Eastern Europe & Central Asia	-30.4239
	(175.06)
East & South Asia	-8.1509
	(0.0720)
Latin America	-6.5409
	(0.0894)
Sub-Saharan Africa	-10.3173
	(1.0888)
North Africa & Middle East	-8.8568
	(0.1766)
education (α_1)	1.6349
	(0.0638)
experience (α_2)	0.0301
	(0.0117)
experience sq. (α_3)	-0.0027
	(0.0006)
English off. Lang. (α_4)	-0.0001
	(0.0700)

Asymptotic standard errors in parentheses grated is equal to 14.76 in utility terms. It is equivalent to receiving in each period a U.S. salary twice the one in the home country, assuming a remaining work life of forty years. It is consistent with the estimates by Alker and Kennan (2011) who find the U.S. inter-state moving cost to be on the order of $300.000 for high-school graduates. The migration cost is indeed increasing with distance to the U.S. so workers from farer countries will move only if their gains from migration are higher. The cost function decreases with contiguity, colonial ties[16] and English being the official language. The coefficient on the percentage of compatriots already living in the U.S. is negative, consistently with the presence of positive network effects (Bertoli and Rapoport, 2015). College graduates have lower migration costs, in line with the evidence that educated workers migrate more even when they have less incentive in terms of wage differentials. The negative coefficient on GDP per capita suggests the presence of budget constraints for migrants from poorer countries.

Figures 1.1–1.3 and Tables 1.9–1.10 show the model fit. The model catches quite well the migrants' age profile by origin. It slightly overestimates the share of workers by Latin America, Africa and Middle East leaving their home country after 45 years old (Figure 1.1). Figure 1.2 reports the distribution of the first wage in the migrants' home country, first column, and the last one before migrating, second column. The simulation returns wage distributions more disperse than in the data, especially in the OECD high-income countries and in East & South Asia. This is due to the large value of the estimated income shock standard deviation, as a result of the wage function trying to fit observations from very different countries. Figure 1.2 reports the distribution of the migrants' first wage in the United States, first column, and the one they declared in the 2003 interview, second column. The bimodal distribution of

[16]The United Kingdom and the Philippines are the only two countries with colonial ties to the United States.

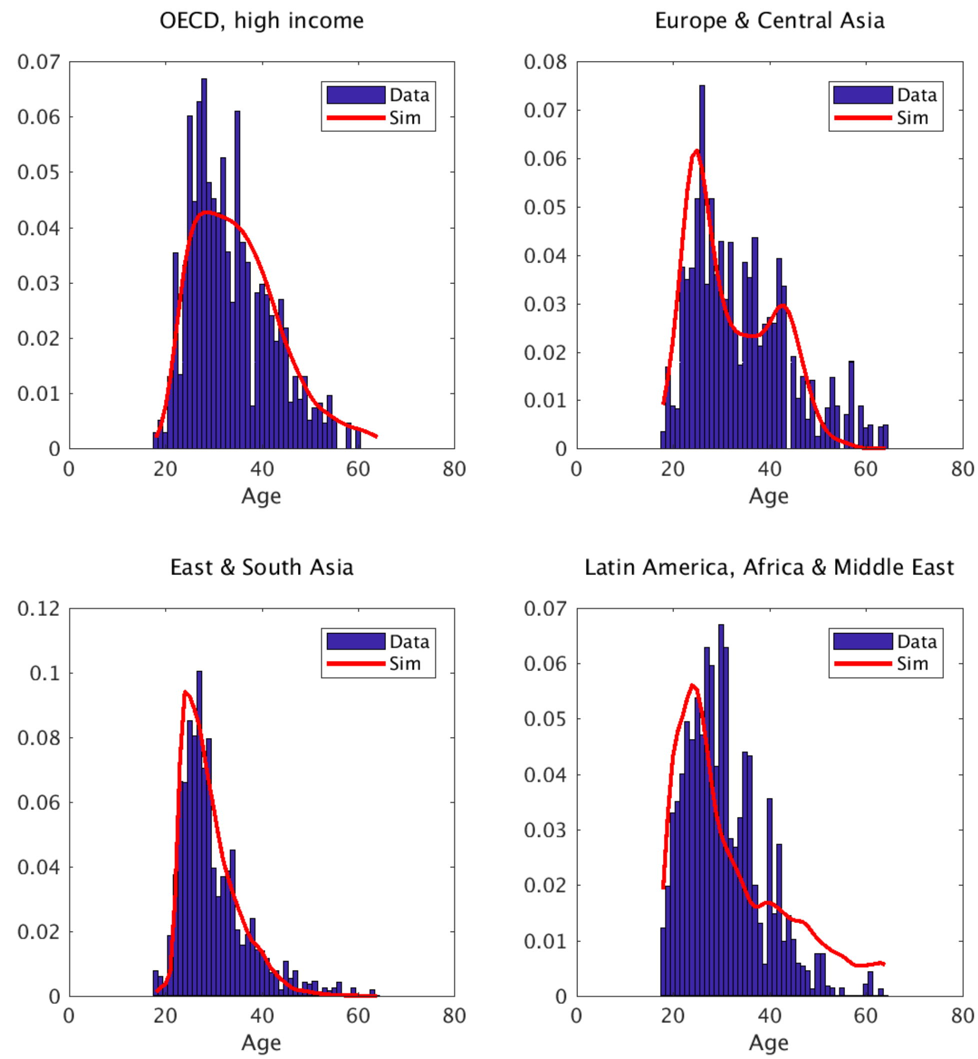
Table 1.8: Parameters of the migration cost function

constant (κ_0)	50.4573
	(0.1104)
log(Distance)	1.0676
	(0.0024)
contiguity	-13.0574
	(0.0584)
colonial ties	-5.4405
	(0.0511)
English off. Lang.	-5.2015
	(0.0173)
% pop. in US in 2000	-1.9620
	(0.0054)
log(GDP per capita)	-2.9713
	(0.0074)
education (α_1)	-10.1175
	(0.0246)
age over 18 (α_2)	-0.0884
	(0.0004)
age over 18 sq. (α_3)	0.0011
	(0.0000)

Standard errors in parentheses

the 2003 simulated wage comes from the process I use to construct the statistic[17].

Figure 1.1: Model fit - Migrants' age distribution by origin



[17]In the data, each individual was interviewed at a different point of their U.S. tenure, depending on the year they migrated. The time dimension does not appear in my model, so I need to mimic the interview occurence to compare the simulated wage to the one in the first-round of the survey. I define the variable $\varrho = \frac{\text{n. years since arrival}}{\text{total years will live in US}}$ and I use the empirical distribution $f(\varrho)$ by country of origin and education to assign different probabilities to different observation times. The distance between the data and the simulated statistic is therefore the combination of two components: the difference in wages by U.S. experience and the difference in U.S. experience at the time of the interview.

Figure 1.2: Model fit - Migrants' wage distribution in their home country

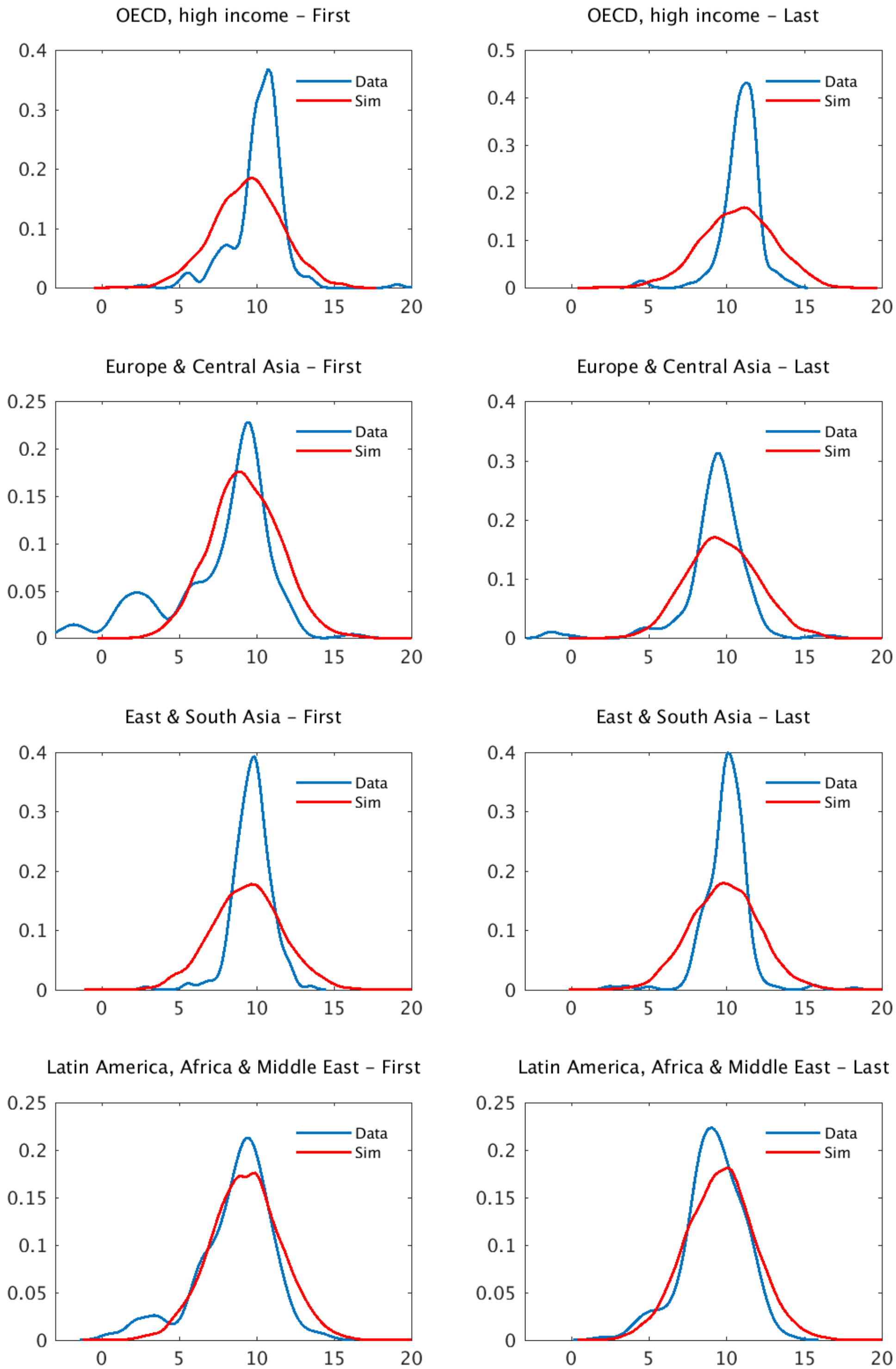


Figure 1.3: Model fit - Migrants' wage distribution in the U.S.

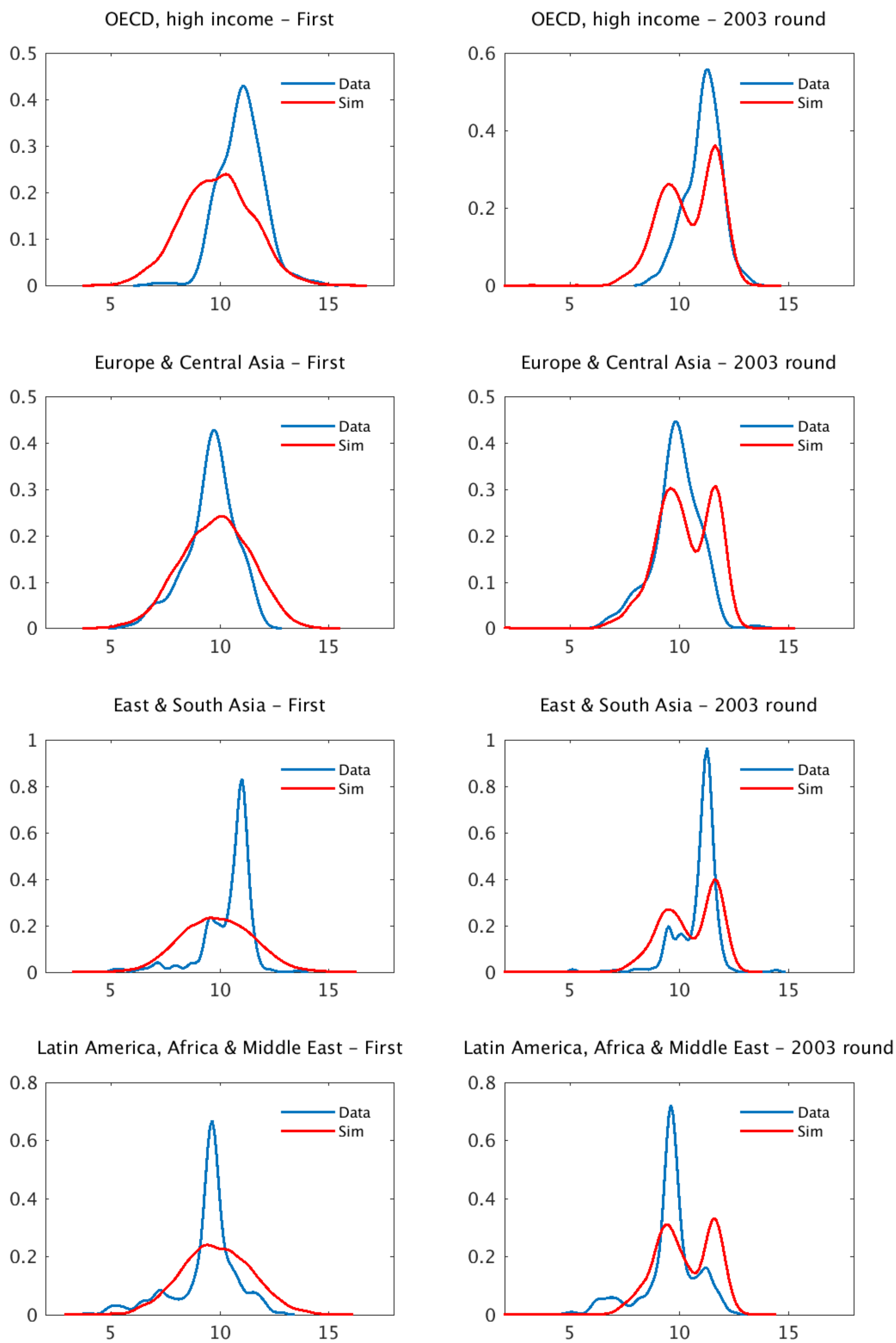


Table 1.9 reports the distribution of migrants by origin and education. The total allocation per group of countries is quite close to the data. The model overestimates the share of college graduates for all groups except Latin America, Africa and Middle East where it is lower. On average, the share of employment-based migrants resulting from the simulation falls short of 15% when compared to the data (Table 1.10).

Table 1.9: Model fit - Migrants' distribution by origin and education

	Uneducated		Educated	
	Actual	Predicted	Actual	Predicted
OECD high-income	0.0277	0.0123	0.1230	0.1422
Eastern Europe & Central Asia	0.0684	0.0359	0.0815	0.1390
East & South Asia	0.0511	0.0181	0.2890	0.3604
Latin America, Africa & Middle East	0.2077	0.1697	0.1517	0.1223

Table 1.10: Model fit - Share of EB migrants by origin

	Actual	Predicted
OECD high-income	0.8420	0.6986
Eastern Europe & Central Asia	0.1646	0.0000
East & South Asia	0.8026	0.6656
Latin America	0.7798	0.4954
Sub-Saharan Africa	0.0866	0.0119
North Africa & Middle East	0.2253	0.1968

Overall, the percentage of high-type migrants is equal to 66.37%. Since it is higher than the initial probability, 50%, it implies positive self-selection among migrants. This is particularly true for those coming from the OECD high-income countries, Eastern Europe and Central Asia, East and South Asia. This is not surprising since in these areas the wage differential is larger for the high-type workers so they have more incentive to move to the United States. Interestingly, the positive self-selection is stronger for uneducated workers. Since they are less likely to migrate because they do not benefit from the higher college premium and they bear a greater migration cost, the high-type incentive plays an even bigger role for them. This is case of selection on observable and unobservable skills going in opposite directions, one of the possible scenarios predicted by the model in Bertoli (2010). On the other hand, migrants from Latin America, Africa and Middle East, who see their high-type premium decrease once in the United States, are negatively selected. In Latin America, the result is more evident for educated workers, who already have a lower incentive to migrate due to the reduced college premium. The situation is reversed in Sub-Saharan Africa, despite the wage function parameters being the same as in Latin America. At a closer look, most of the migrants from Sub-Saharan Africa come from countries with very high unemployment rates, on the order of 22%. The unemployment spell is more painful to educated workers because their wage, in times of employment, is larger. In order to escape the huge unemployment rates, college graduates from Sub-Saharan Africa are more willing to move to the Unites States than their Latin America counterparts. Their incentive is so large to reverse the prediction on unobservables' self-selection, based uniquely on wage differentials, between educated and uneducated workers, which is fulfilled in case of Latin America. The result shares the same logic as Rendon and Cuecuecha (2010), who argue that labor market mobility is as important as wage differentials.

Table 1.11: Share of high-type migrants by origin and education

	Uneducated	Educated
OECD high-income	0.7846	0.7848
Eastern Europe & Central Asia	1.0000	0.9149
East & South Asia	0.9884	0.8576
Latin America	0.3079	0.1939
Sub-Saharan Africa	0.0195	0.2772
North Africa & Middle East	-	0.0116

Table 1.12: Share of high-type migrants by origin and visa category

	DV	EB
OECD high-income	0.9795	0.7007
Eastern Europe & Central Asia	0.9324	0.9489
East & South Asia	0.8045	0.8937
Latin America	0.3135	0.2115
Sub-Saharan Africa	0.1290	0.2746
North Africa & Middle East	0.0000	0.0591

1.8 Policy Experiment

1.8.1 Experiment Design

Since the Immigration Act of 1990, U.S. policy-makers made several attempts to radically change the structure of the American visa system. So far, none of them has succeeded to pass the vote of both houses of the Congress. The latest one is the S. 1720 bill, also known as the RAISE (Reforming American Immigration for Strong Employment) Act, introduced in the Senate in August 2017. On top of changes to the family preference path, the bill would end the Diversity Visa Program and introduce a visa-point system[18]. Table 1.13 displays the details of the proposed point allocation scheme. The new system would favor immigrants between 26 and 30 years old, English proficient, with an advanced degree in scientific and technological fields, especially if earned in the United States. Employer sponsorship would not be necessary any more, but a job offer with a high enough salary would give the applicant additional points. Investors and individuals who achieved extraordinary results in their life, as Nobel laureates or Olympic medal winners, would be awarded extra points as well. Only immigrants accruing more than 30 points would be eligible to submit an

[18]A previous bill, S. 744, also proposed a visa-point system. The bill passed the vote in the Senate in June 2013, but was never discussed in the House of the Representatives.

application. The worldwide annual limitation would remain 140,000 visas as it is now for the employment preference category. No limit would be imposed at the country level. Point-based visas would be issued to the highest ranked applicants up to the achievement of the annual limit.

The bill has been at the center of debate with opposite factions arguing about the possible effect on the U.S. economy. A step back is necessary to ask: how would the migrants' distribution, in terms of observable and unobservable characteristics, change if the new visa-point system was in place? I embed a simplified version of the proposed point allocation scheme in my estimated model to answer the question. Each potential immigrant gets a number of points equal to:

$$v(S,a,j) = \sum_{i=1}^{8} \nu_{j,i}\, \mathbb{1}\{a \in A_i\} + \nu_e\, \mathbb{1}\{e = E\} + \sum_{i=i}^{5} \nu_{\mathrm{ep},i}\, \mathbb{1}\{\mathrm{toefl}_{b,e} \in \mathrm{EP}_i\} +$$
$$+ \nu_\theta\, \mathbb{1}\{b \in \{\text{OECD high-income, East \& South Asia}\}, \theta = \mathrm{H}, j = 1\}$$

where $S = \{b, e, \theta\}$, $\nu_{a,i}$ are the points assigned by age category A_i, $\nu_e = 4$, i.e. the difference in points between high-school and non-U.S. college graduates, $\nu_{\mathrm{ep},i}$ are the points assigned by English proficiency category EP_i and $\nu_\theta = 8$, i.e. the number of points allocated to applicants with a highly compensated job offer[19]. English proficiency is by country of origin and level of schooling: the distribution comes from the score data for the TOEFL exam[20]. Not all the immigrants getting a U.S. job offer, $j = 1$, are awarded additional points: first, they must be high-type workers, $\theta = H$, and second, they must come from the OECD high-income or the East & South Asia countries. The two conditions were chosen to reflect the intent behind the high

[19] Annual salary offered is at least 200% but less than 300% of the median household income in the state of employment.

[20] Test and Score Data Summary for TOEFL iBT Tests (January 2017 – December 2017 Test Data). TOEFL, which stands for Test of English as a Foreign Language, is a standardized test to measure the English language ability of non-native speakers.

Table 1.13: Visa point system under the RAISE Act

	Points
Age (10 points maximum) Between 18 and 21	6 points
Between 22 and 25	8 points
Between 26 and 30	10 points
Between 31 and 35	8 points
Between 36 and 40	6 points
Between 41 and 45	4 points
Between 46 and 50	2 points
51 or older	0 points
Formal education (13 points maximum)	
U.S. or Foreign High School Degree	1 point
Foreign bachelor's degree	5 points
U.S. Bachelors Degree	6 points
Foreign master's degree in Science, Technology, Engineering or Mathematics (STEM)	7 points
U.S. STEM Masters Degree	8 points
Foreign Professional Degree or Doctoral STEM	10 points
U.S. Professional Degree or Doctoral STEM	13 points
English language proficiency (12 points maximum)	
1st – 5th deciles	0 points
6th – 7th deciles	6 points
8th decile	10 points
9th decile	11 points
10th decile	12 points
Extraordinary achievement (40 points maximum)	
Nobel Laureate or comparable recognition	25 points
Individual Olympic medal or first place in a comparable international sporting event	15 points
Job offer/highly compensated employment (13 points maximum)	
Annual salary offered is at least 150% but less than 200% of the median household income in the state of employment	5 points
Annual salary offered is at least 200% but less than 300% of the median household income in the state of employment	8 points
Annual salary offered is at least 300% of the median household income in the state of employment	13 points
Investment (12 points maximum)	
Investment of at least \$1.35 million but less than \$1.8 million in a U.S. New Commercial Enterprise (NCE); maintain the investment for three years and play active role in managing the NCE as primary occupation	6 points
Investment of at least \$1.8 million in a U.S. NCE	12 points
Valid (pre-existing) offer of admission under family preference category	2 points

compensation criterion. The high types earn, on average, a greater wage than the low types and the difference is particularly striking in case of workers coming from the OECD high-income or the East & South Asia countries, where the type premium in the U.S. wage function is above two.

An applicant gets a visa if

$$v(S,a,j) + \epsilon \geq \underline{v} \qquad \text{where} \qquad \epsilon \sim \mathcal{N}(0,1)$$

Some of the attributes considered by the RAISE point allocation scheme are not included in my model (for example, degrees higher than bachelor or received in the United States) so it would not be fair to use the threshold of $\underline{v} = 30$, established by the bill. Moreover, I am interested only in the composition effect of the point-based system so I would like to keep the size of the migrant population unchanged. Hence, the threshold is set at $\underline{v} = 27.318$, such that the number of foreign workers migrating to the United States is the same as in the baseline model.

Then, the probability of migrating upon visa application is, in absence of a U.S. job offer:

$$\tilde{q}(S,j) = 1 - \Phi\big(\underline{v} - v(S,j,m=0)\big)$$

while, in presence of a U.S. job offer:

$$\tilde{p}(S,j,h_0) = p(b,e,h_0) \cdot \Big(1 - \Phi\big(\underline{v} - v(S,j,m=1)\big)\Big)$$

The assumptions about the different informational structure and labor status of migrants with or without job offer stay valid.

1.8.2 Experiment Results

Figure 1.4 and Tables 1.14–1.16 show the results from the policy experiment. The quasi totality of migrants are college graduates. 95% of them come from the OECD high-income (64%) or the East & South Asia (31%) countries. The geographical distribution is not driven by the points allocated to highly compensated job positions, as the proportion of workers moving to the United States with a job offer already in place decreases dramatically. That is because in the baseline model, some major countries, as India or Canada, were completely excluded from the Diversity Visa Program so employer sponsorship was their only way into the United States. Age at migration is highly concentrated around 26-30 years old. So, the proposed visa-point system works in terms of observable characteristics, channeling the migrants' pool towards the desired attributes, maybe even too so given the lack of differentiation in terms of schooling, age and origin.

Table 1.14: Policy experiment - Migrants' distribution by origin and education

	Uneducated		Educated	
	Baseline	Experiment	Baseline	Experiment
OECD high-income	0.0111	0.0000	0.1458	0.6419
Eastern Europe & Central Asia	0.0377	0.0000	0.1362	0.0101
East & South Asia	0.0174	0.0000	0.3522	0.3072
Latin America, Africa & Middle East	0.1799	0.0000	0.1195	0.0408

The next step is looking at the unobservable characteristics, i.e. the type distribution. The point allocation scheme tries to capture some of the unobservables by rewarding highly compensated job offers. However, the direct effect is small given the low proportion of migrants leaving their home country with secure employment prospects. Overall, there is no improvement in the proportion of high-type workers, actually the percentage slightly decreases from 66.37% to 63.54%. The reduction is

Figure 1.4: Policy experiment - Migrants' age distribution by origin

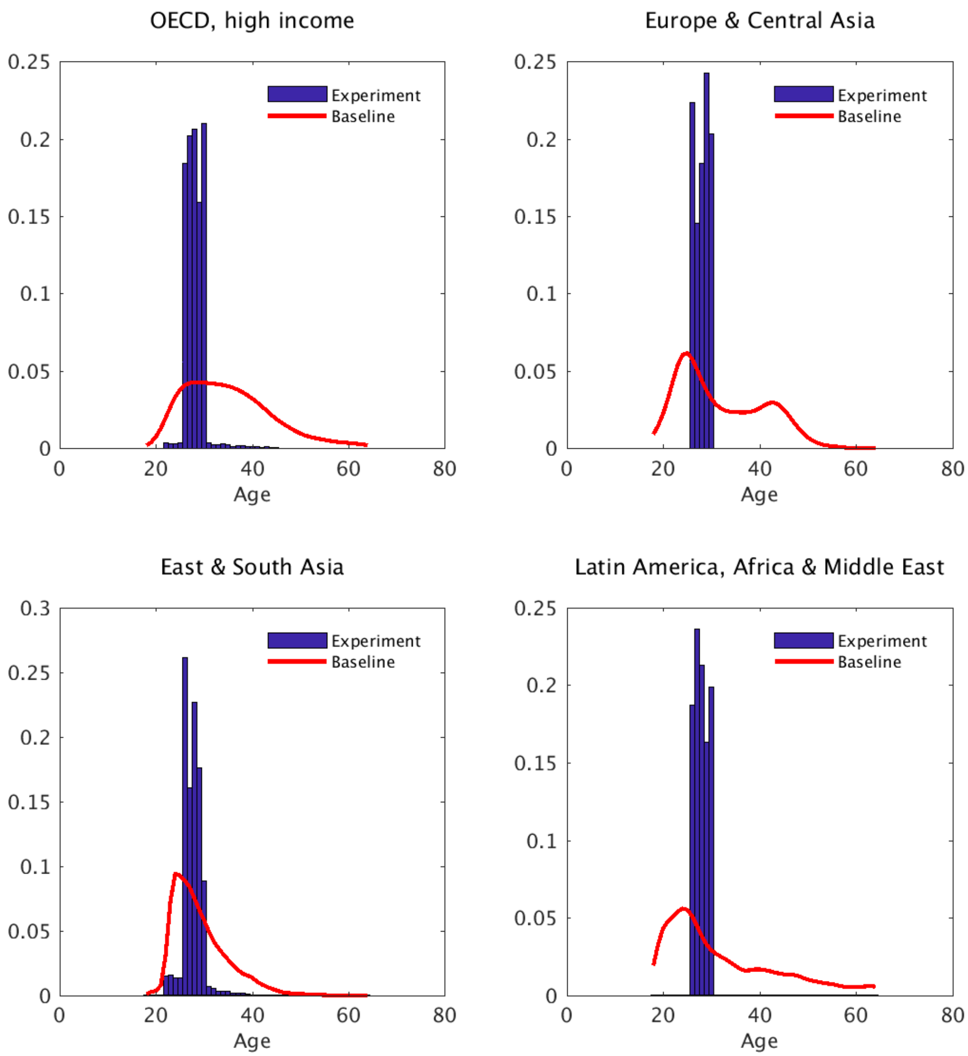


Table 1.15: Policy experiment - Share of migrants with a job offer by origin

	Baseline	Experiment
OECD high-income	0.6812	0.0391
Eastern Europe & Central Asia	0.0000	0.0000
East & South Asia	0.6740	0.1468
Latin America	0.4752	0.0078
Sub-Saharan Africa	0.0114	0.0002
North Africa & Middle East	0.2215	0.0015

due to migrants from the OECD high-income countries, in particular to Canadian workers, whose weight is more than doubled in the group, now that they can get a visa without a job offer. Their share of low types is higher compared to other OECD countries, because they have a smaller migration cost, due to contiguity to the United States. The proportion of high types has increased among East & South Asia workers: since they are the one migrating the most with a job offer, it may be the direct consequence of the additional points allocated to highly compensated employment. As for the remaining areas, the change in the type distribution is consistent with workers having more incentive to migrate, all things being equal. Actually, their probability of getting a visa has decreased and so is the continuation value in their home country. Hence, individuals with lower incentive to migrate, namely low types in Eastern Europe & Central Asia and high-type in Latin America, Africa & Middle East, now move more often to the United States. The effect is captured thanks to the dynamic structure of the model.

1.9 Conclusions

This paper develops and estimates a life cycle model of migration choice from the rest of the world to the United States through two possible paths: the employment preference visa or the diversity program. The wage function includes both observable

Table 1.16: Policy experiment - Share of high-type migrants by origin and education

	Baseline	Experiment
OECD high-income	0.7888	0.5139
Eastern Europe & Central Asia	0.9213	0.8448
East & South Asia	0.8747	0.9190
Latin America	0.2675	0.4452
Sub-Saharan Africa	0.1260	0.4321
North Africa & Middle East	0.0098	0.2004

skills, in terms of college degree, and unobservable skills, in terms of high versus low types. The estimated parameters predict positive self-selection on both type of skills for migrants coming from Europe, Asia and other OECD high-income countries: the degree of selection on unobservables is actually higher for uneducated rather than educated workers, because their incentive to move to the United States is lower. Then, I conduct a policy experiment by applying a simplified version of the visa-point system proposed by the RAISE Act to the model. I find that the new system would bring to the U.S. only young and educated workers, mostly coming from OECD high-income countries or East and South Asia. However, without per-country limitations, the policy attracts an excessive number of migrants from the closest countries, and since their migration cost is lower they are less favorably selected in terms of unobservable skills. In this case, less diversity translates into less uniqueness. To fix the distortion, the point allocation scheme should increase the weight to highly compensated job offers, which allow to identify high types.

There are several prospects for future research. First, the model takes education as exogenous, while it may play an important role in the migration decision, given that the U.S. are the country with the largest number of international students. Therefore, it is necessary to endogenize the schooling choice in two directions: whether to pursue a college degree and where to pursue it, in its own country of origin or in the United

States. Second, visa policies are now built in a partial equilibrium framework where the probability of being admitted does not depend on others' behavior. Bringing the model to the general equilibrium framework would allow to take into account the visa numerical limitations and better evaluate the consequences of a policy change.

Chapter 2

The Impact of H-1B Visa on U.S. Innovation

2.1 Introduction

The United States are the country of immigrants by definition. Many have pointed out to the key contribution of foreign talents to the U.S. innovation leadership in the world. It is hardly surprising if we look at some raw data: in the 2000 Census, immigrants represent 24% of the STEM (Science, TEchnology and Math) workforce with a bachelor and 47% of the STEM workforce with a doctorate (Kerr and Lincoln, 2010). The numbers are impressive after considering that their share over the the total labor force reaches just 12%. Wadhwa *et al.* (2007) finds that 24.2% of international patent applications[21] filed from the U.S. in 2006 have at least one foreign inventor. Another work by Wadhwa *et al.* (2012) reports that 24.3% of the engineering and technology companies founded between 2006 and 2012 list at least one key founder as foreign-born and the percentage rises to 43.9% if we restrict our attention to the Sylicon Valley.

Given the above figures and their policy implications, the issue has drawn the attention of many economic researchers seeking to shed light on the link between innovation and immigration. My paper finds its collocation in this current of the literature. I use data about patents and H-1B applications to test different hypothesis.

[21] International patent applications are filed under the Patent Cooperation Treaty which allows the inventor to seek patent protection for an invention simultaneously in each of the 147 countries supporting the Treaty.

H-1B visa were introduced by the Immigration Act of 1990 as non-immigrant visas for temporary workers in specialty occupations. The term "specialty occupation" does not lead to a formal requirement but the usual interpretation implies the worker has at least a bachelor. The H-1B visa lasts 3 years after which it can be renewed for the same period of time. The U.S. government regulates the entry of temporary workers through quotas.

Workers entering the United States under H-1B visa represent the perfect measure for skilled immigration. Unfortunately, there are no micro data on H-1B issuances but there are firm-level records on the Labor Condition Application (LCA) the employer needs to submit before petitioning for the visa. The database has been used by Ruiz *et al.* (2012) to study the demand for H-1B workers in U.S. metropolitan areas and marginally by Kerr and Lincoln (2010) to infer the link between H-1B visa and U.S. inventions by ethnicity. However, this is the first time, at least at my knowledge, which the database is used in such an extensive way. The final product is a unified firm-level dataset of patents and LCA from 2001 to 2007, offering many insights on the mechanism through which skilled immigration affects innovation.

The first step of my analysis aims at finding the overall effect of H-1B visa on the firm patent activity. Chellaraj *et al.* (2008) conduct a time series analysis on patent grants revealing a positive impact of foreign graduate students over time. Maskus *et al.* (2010) look more closely to the science and engineering departments from 1973 to 1988 but they don't find any significant difference between native and foreign doctoral students in terms of contribution to publications and citations. They argue that this is a sign of the optimal behavior of universities in choosing PhD students. The state level analysis by Hunt and Gauthier-Loiselle (2010) reveals that a 1% increase in the share of high-skilled immigrants leads to 15% more patents. Hunt (2011) focus on the distinction between different categories of visa finding that immigrants who entered through a work or student visa outperforms natives in terms of wage, patents and start-up founded. Kerr and Lincoln (2010) find that higher H-1B admissions increase the number of patents by inventors with Chinese and Indian names in cities and firms which are more dependent on H-1B relatively to others. However, they have limited effect on native patenting.

My results suggest that LCA have a small but positive effect on innovation: a 10% increase in applications leads to 0.6% more patents at firm level. The estimates improve when I use an instrumental variable to address the endogeneity of LCA, signaling the presence of measurement error in the first place. The instrument is constructed by the interaction of two variables: H1-B quotas over time and the number of Asian inventors who patented within the firm in the 90s. It is similar to the one already used in the immigration literature (DiNardo and Card (2000), Hunt and Gauthier-Loiselle (2010), Kerr and Lincoln (2010)).

Another channel through which skilled immigration affects innovation is the market size effect (Acemoglu, 2002): the profitability of patents increases pushing firms to invest more in R&D and to hire more inventors. In my data, the hypothesis is confirmed by the positive impact of LCA on the number of future new inventors.

Immigrants can benefit the firm but harm the native workers in terms of wages (Borjas, 2005) or displacement (Borjas, 2004). Pekkala Kerr and Kerr (2013) observe that workers in STEM occupations leaving the firm in high immigration periods have harder time to find a new job and suffer larger wage declines compared to non-STEM workers. The key point is the degree of substitutability between foreign and native workers, as noticed by Ottaviano *et al.* (2005). By allowing imperfect substitutability, as supported by their estimates, they find that a 10% increase in total employment due to immigration can lead to an overall increase in wages of 3-4%. I don't observe wages for native workers so I can't test any hypothesis in that direction. However, I can exclude any crowd-out effect by means of H-1B workers: the probability for incumbent inventors to patent again in the future in the same firm increases as the number of LCA grows.

The positive impact on the incumbent patent activity is not just at the extensive margin but also at the intensive one. There may be several reasons behind this higher production of ideas. Incumbents may feel threatened by the newcomers and pushed to patent more to escape the competition. As an alternative, the arrival of inventors bringing different culture and expertise may generate profitable collaborations and positive spillovers, as suggested by Alesina *et al.* (2013) and Niebuhr (2006). I distinguish the patents by the ranking of inventors. I argue that patents where the incumbent is the first inventor are likely to be related to the escape competition effect while patents where he appears as one of the subsequent inventors and the newcomer is the first one are likely to be the result of collaboration. Both effects are present in my data but I cannot say which one is prevailing.

Finally, I exploit citations to detect additional spillovers. I construct three indices capturing the diversity of citations by respectively country, state and class. Inventors in firms applying for more H-1B workers cite patents more sparse in terms of country and class. I interpret the results as supporting the presence of positive spillovers.

The paper is structured as follows: Section 2 describes LCA and patents data; Sections 3 presents the empirical strategy applied; Section 4 contains the results of the estimation; Section 5 draws some concluding remarks.

2.2 Data

The LCA database is publicly released by the U.S. Department of Labor (DoL). Every employer petitioning for H-1B visa needs to first file an LCA. The purpose of the

LCA is to ensure that, by hiring foreign workers, the firm does not harm the existing employees. In order to prove that, the employer must declare the work location, the occupation and the proposed wage for the H-1B worker. The compensation must be equal or higher than the prevailing wage for the same occupation and work area. The value of the prevailing wage comes from collective bargaining or it can be required to the State Employment Security Agency. In addition, by signing the LCA, the firm attests that on the application date no strike, lock-out or work stoppage is in course in the same work location and occupation of the H-1B worker.[22] The employer is not required to advertise the position filled by the foreign worker but a copy of the LCA must be published at the workplace or notified to the employee representative.

The LCA is just the first step of the process leading to the issuance of the H-1B visa. Once, it is certified by the DoL, the employer can file a H-1B petition to the US-CIS which makes the final decision. So, even after excluding all the applications which have been withdrawn or rejected by the DoL, the data are inflated with respect to the real number of H-1B issued. In other words, there is no one-to-one correspondence between LCA records and H-1B visa. In addition, the database misses to distinguish between applications for new issuances and renewals. The employer is required to fill an LCA application in both cases but I don't observe the distinction. Ideally, for my analysis, I would restrict the sample only to new issuances.

Despite some measurement error, the LCA database represents a very valuable resource for inferring the number of non-immigrant workers employed in a firm. The data go from 2001 to 2012. The records are divided by fiscal years. However, since the starting date for the certification is reported, I take that as time reference.

Given that my study focus on the impact of the H-1B visa on innovation, I am particularly interested in the STEM occupations. The database uses 3-digit DOT

[22] U.S. Department of Labor

occupation codes from 2001 to 2008, 6-digit 2000 SOC codes from 2009 to 2010 and 6-digit 2010 SOC codes from 2011 on. I identify the STEM occupations through the list provided by the Bureau of Labor Statistics.[23] The document specifies the STEM occupations in terms of 2010 SOC code so I further use the crosswalks between 2010 SOC and 2000 SOC codes[24] and between 2000 SOC and DOT codes[25] to complete the identification. Moreover, I run a word search in the occupation title to refine my classification, for example by including all the observations with the word "engineer" or excluding all the ones with the word "accountant". The passage from DOT to SOC codes may lead to some discrepancies over time on the entire sample. However, after merging the LCA database to the patent data, I restrict my sample to the time interval going from 2001 to 2007 where only DOT codes are used so the classification between STEM and non-STEM occupations is consistent over time.

For patents, I use jointly two different datasets, the dataset by NBER and the dataset by Lai *et al.* (2001). The NBER dataset includes all the USPTO patents granted between 1976 and 2006 and divided by assignee. The advantage of the dataset is the identification of each assignee by a unique code through a name matching algorithm. The code is more accurate than the one assigned by USPTO since it takes care of all the cases where the assignee is the same but the name was recorded differently originating different codes. On the other side, the dataset by Lai *et al.* (2001) covers all the patents granted between 1975 and 2010 divided by inventors. Similarly to the NBER dataset, they use a disambiguation algorithm to identify inventors and assign a unique number to each of them. I merge the two datasets to get full identification of both assignees and inventors.

[23] "Detailed 2010 SOC occupations included in STEM". SOC Policy Committee recommendation to the Office of Management and Business (U.S. government)

[24] Bureau of Labor Statistics

[25] "ETA Form 9035CP General Instructions for the 9035 and 9035E - Appendix I". Department of Labor

I use the application year as time reference since it is the one closer to the moment when the idea has originated. However, since patents are recorded only after they have been granted and the average time between the application and the grant date is about three years, I need to exclude all the patents granted after 2007 to avoid the truncation bias.

In order to merge the LCA and the patent dataset, I first need to standardize the employer names in the LCA records. I use a modified version of the standardization routine applied by Jim Bessen in the NBER Patent Project. By merging the two datasets through a name matching algorithm, I further restrict my sample to 7% of the total number of employers and 29.17% of the total number of LCA applications. The share of identified firms seems small but if we concentrate our attention to the first 100 employers by LCA applications, the percentage grows to 65% both in terms of number of firms and applications. Moreover, most of the excluded employers are consultancy firms which largely use H-1B workers but don't produce any patents so they are irrelevant for my analysis. In sum, I end up with a sample of around 7,000 firms, which I consider being fairly representative of the population of firms hiring H-1B workers and producing patents.

2.3 Empirical Framework

2.3.1 Analysis by Assignee

I start my analysis at the firm level by assessing the relationship between H-1B workers and patents, according to the following regression:

$$\log(P_{i,t}) = \alpha + \beta_1 \log(\mathrm{LCA}_{i,t}) + \beta_2 \log(\mathrm{PS}_{i,t}) + \delta_t + \eta_i + \epsilon_{i,t}$$

where $P_{i,t}$ is the number of utility patents for which firm i applied in year t, $LCA_{i,t}$ is the number of non-immigrant workers in STEM occupations for which firm i filed an LCA in year t and $PS_{i,t}$ is the patent stock of firm i at time t. The panel dimension of my dataset allows me to add firm and year fixed effect to control for unobserved heterogeneity. I use the stock of patents of the assignee as proxy for its size since I don't have any information about the total number of employees and I suspect the size of the firm over time may be correlated to both patents and H-1B applications. In this way, I can solve the omitted variable issue.

In the log-log specification above, all the observations where the number of patents is equal to zero are discarded. So I am using only part of the information I have. The solution suggested by Hausman *et al.* (1984) is to apply a Poisson regression model instead where

$$\log\bigl(\mathbb{E}[P_{i,t}|X_{i,t}]\bigr) = \alpha + \beta_1 \log(\mathrm{LCA}_{i,t}) + \beta_2 \log(\mathrm{PS}_{i,t}) + \delta_t + \eta_i$$

and $P_{i,t}$ is distributed according to a Poisson with parameter $\lambda_{i,t}$ so that $\mathbb{E}[P_{i,t}|X_{i,t}] = \lambda_{i,t}$. The specification is really similar to the log-log case and the interpretation of the result is the same, which makes the comparison easy. In addition, the observations where $P_{i,t}$ is equal to zero are not discharged but they are fully used in the estimation of the model.

If we assume that $P_{i,t}$ has a Poisson distribution, we imply the mean and variance are equal. This is not the case for the patents in the sample where we observe high dispersion. The issue doesn't affect the consistency of the estimated coefficient but its standard error. Hausman *et al.* (1984) propose a Negative Binomial regression model where the mean of the distribution is unchanged but the variance takes into account the overdispersion. For sake of robustness, I will present the estimates for all three models.

All the above regressions could return biased estimates if the error term is correlated to any of the regressor, giving rise to an endogeneity problem. In my case, there may be some concerns: it is easy to imagine a situation where the firm is hit by a positive innovation shock and needs more employees to develop its ideas. So, the increase in H-1B workers would be correlated to the increase in patents, generating a positive coefficient β_1 even without any causal relationship.

In order to solve the endogeneity problem, I need to instrument the H-1B regressor with a variable which is correlated to LCAs but exogenous to the quantity of patents produced. The annual quotas set by the government for the issuance of H-1B visa represent the perfect instrument. They are independent from the error term because they are decided at authority level so they do not depend on the firm decisions.[26] They are also correlated to LCAs because the H-1B petition process is costly[27] so, if the quota is low, we may expect firms to be more cautious in applying for H-1B visas they can hardly get. The H-1B quota does not depend on the country of origin of the worker, differently from the quotas for permanent residency. It was equal to 195,000 visa from 2001 to 2003. It decreased to 65,000 visa in 2004 and then increased again to 85,000 visa in 2006 when the government allowed for an additional 20,000 quota reserved to foreign students graduating from U.S. colleges.

The H-1B quota changes over time but it is constant across firms so we need to find a variable to interact, which could reflect the higher dependency of the firm on H-1B workers. The immigration literature has often used the stock of immigrants from a particular country to instrument the subsequent flow. The first paper to apply this

[26]The lobbying activity by firms which are highly dependent from H-1B visa could invalidate this statement but I am assuming the result of this activity is not immediate so that there is some time displacement between the innovative shock and the increase in H-1B. Time dummies and firm fixed effect should do the rest.

[27]Fees go from a minimum of $1,575 to a maximum of $5,550 per worker.

technique was DiNardo and Card (2000). They study the crowd-out effect of low-skill immigrants on native workers, facing a similar endogeneity problem. However, they argue that immigrants tend to settle in place where communities from the same country are already established. Since most of the low-skill immigration flow in the 80s came from Mexico, they use the relative Mexican population in 1970 as instrument. It doesn't suffer from endogeneity issues and it is correlated with the regressor of interest. I apply a similar approach by using the quantity of Asian inventors who patented in the firm between 1990 and 2000 but not after as my instrumental variable. In 2004, around 65% of the new H-1B visa issuances were from East Asia.[28] It's reasonable to expect non-immigrants workers to prefer firms with a larger number of employees from their same country of origin, for cultural and network reasons. In the meanwhile, by excluding all the Asian inventors who patented also after 2000 I am sure the variable is not correlated to the error term.

Unfortunately, the patent records do not tell the country of birth of the inventor so I identify Asian inventors by matching their surnames to a list of Asian surnames, constructed from the results of Lauderdale and Kestenbaum (2000) and Kerr (2007). I am aware that ethnicity and country of birth may not correspond but I am confident the variable represents a good guess of the firm dependency from H-1B workers. Hence the instrumental variable is:

$$\text{Instrument}_{i,t} = \log(\text{Asian}_i) \times \log(\text{H-1B quota}_t)$$

I am also interested to test the market size effect of H-1B workers. The increase in high-skilled immigration may increase the profitability of patents and push firms to hire even more inventors. I don't observe the time of entry of a new inventor in

[28] "Characteristics of Specialty Occupation Workers (H-1B): Fiscal Year 2004". USCIS

the firm but I can observe if he has ever patented with that assignee before. The regression of interest is:

$$\log(\text{New Inv}_{i,t+1}) = \alpha + \beta_1 \log(\text{LCA}_{i,t}) + \beta_2 \log(\text{PS}_{i,t}) + \delta_t + \eta_i + \epsilon_{i,t}$$

where New $\text{Inv}_{i,t+1}$ is the number of inventors who have never patented in the firm before. Again, in addition to the log-log regression, I use a Poisson and a Negative Binomial regression as robustness check. I suspect there is the same endogeneity issue as before so I also run a 2SLS regression using the instrument I described above.

2.3.2 Analysis by Incumbent Inventor

After examining the overall impact of LCAs on the firm patenting activity, I want to study the effect on incumbent inventors, i.e. inventors already working in the firm when the immigration flow takes place. The issue arises several questions: are foreign and native inventors substitutes or complements in the patent production function? Does the arrival of H-1B workers trigger an escape competition mechanism so that incumbents are pushed into producing more ideas? Are incumbents collaborating with the new inventors on the same innovation?

The analysis starts with a probit model where I explore the effect of LCAs on incumbent patenting activity at the extensive margin:

$$\mathbb{E}[Y_{ji,t}|X] = \Phi\Big(\alpha + \beta_1 \log(\text{LCA}_{i,t}) + \beta_2 \log(\text{PS}_{i,t}) + \beta_3 \text{Age}_{j,t} + \delta_t + \eta_i + \epsilon_{ji,t}\Big)$$

where $Y_{ji,t}$ is equal to 1 if I observe the inventor j patenting in the firm i in the years after t. I identify incumbent inventors as the ones who have already patented with the assignee before t. Inventors who patented for the first time in year t are discarded.

$\text{Age}_{j,t}$ is the time between t and the first patent of the inventor j. If the firm is substituting old inventors with H-1B workers, this could bias the coefficient estimates in one direction or the other so I add $\text{Age}_{j,t}$ as control variable. The model is not linear so in order to capture the firm fixed effect, I use the Chamberlaine-Mundlak correction by including as regressors the time average of $\log(\text{LCA}_{i,t})$ and $\log(\text{PS}_{i,t})$.

I continue by testing the impact of H-1B workers on incumbent patenting activity at the intensive margin:

$$\log(P_{ji,t}) = \alpha + \beta_1 \log(\text{LCA}_{i,t}) + \beta_2 \log(\text{PS}_{i,t}) + \beta_3 \text{Age}_{j,t} + \delta_t + \eta_{ji} + \epsilon_{ji,t}$$

As before, since I am interested only in incumbent inventors, I discard all the inventors who patented for the first time in year t.

Discerning the difference between escape competition and collaboration effect is not easy. I try to exploit the different ranking of inventors within the same patent. The ranking is not arbitrary: the first inventor is really the one bringing the main contribute to the invention. The incumbent, willing to prove his value against newcomers, is more likely to concentrate his activity on patents where he is the first inventor and so he can gain higher recognition. Therefore, I interpret the effect of LCAs on this subset of patents as escape competition effect.

Conversely, collaboration with H-1B workers may realize through patents where the newcomer is the first inventor and the incumbent is one of the subsequent. In this case, the entrant is bringing the key piece of knowledge for the patent. Hence, I see the effect of H-1B workers on this subset of patents as collaboration effect. I identify an inventor who patent for the first time in year t as newcomer. Since I don't observe the country of birth of inventors, I cannot distinguish between natives and foreigners so I am implicitly assuming all the newcomers are H-1B workers.

As before, all the models explained in this section are also estimated by instrumenting the number of LCAs with the interaction between H-1B quotas and Asian incumbent inventors within the firm.

2.3.3 Citation Analysis

Even if incumbent and newcomer inventors do not work on the same patents, there is still room for potential spillovers generated by the arrival of workers of different nationality, experience and knowledge. Citation data have been used in the literature to detect such effects (Hall *et al.*, 2001). In my case, I expect incumbents to cite more foreign patents as more H-1B workers enter the firm. For all the patents generated by incumbents in the period 2001-2007, I construct a Herfindahl diversity index:

$$H_j = 1 - \sum_k \left(\frac{C_{jk}}{C_j} \right)^2$$

where C_{jk} is the number of patents from country k cited by patent j and C_j is the total number of citations made by patent j. The index has value 0 if all the cited patents are from the same country and it increases as diversity grows. I identify the location of the patent as the country of residence of the first inventor.

The regression model follows:

$$H_{ji,t} = f\Big(\alpha + \beta_1 \log(\mathrm{LCA}_{i,t}) + \beta_2 \log(\mathrm{PS}_{i,t}) + \beta_3 \,\mathrm{Self}_j + \delta_t + \eta_i + \epsilon_{ji,t} \Big)$$

where $H_{ji,t}$ is the diversity index of citations for patent j applied at time t by the assignee i. Self_j is the ratio of self-citations, i.e citations to patents with the same assignee, over total citations for patent j. Firms citing their own patents more often are likely to have a lower diversity index so I want to control for this effect.

Since the dependent variable is a ratio, the linear regression model is not the best alternative. Hence, I apply the fractional logit model by Papke and Wooldridge (1996) and I include the time average of log(H-1B$_{i,t}$) and log(PS$_{i,t}$) to capture firm fixed effects.

Similarly, I construct two other indices: one controlling the diversity of citations by state (where, again, I assumed the state of residence of the first inventor as location of the patent) and the other one controlling the diversity of citations by class. The intuition is the same: if the new H-1B workers generate spillovers by bringing new expertise, I should observe the citations of incumbent patents to be more disperse in terms of state and class.

2.4 Results

Table 2.1 shows the results from the regression of patents over LCAs by firm. Looking at the estimates from the log-log regression, in the first column, the coefficient of log(LCA$_{i,t}$) is positive and significant: a 10% increase in the number of applications for H-1B workers in STEM occupation generates a 0.6% increase in patents. The results for the Poisson model, in the third column, and the Negative Binomial model, in the fourth column, are similar.

The outcome from the 2SLS regression with the instrumental variable, in the second column, shows a coefficient which is almost ten times larger than the previous ones: a 10% increase in LCAs induces a 5% increase in patent activity. The instrument is relevant as proved by the t-statistic of the first-stage regression equal to 2.07. The sample reduces to 15% of the original one, losing all the firms for which we don't observe Asian inventors who patented between 1990 and 2000 but not after. However, the coefficient remains highly significant and may capture even better the effect of new H-1B workers in the firm. Actually, the data about LCA include both new issuance and renewal requests so the number of incoming foreign employees is inflated. Moreover, the approval of the LCA is just the first step for obtaining H-1B visa: we don't observe the final outcome so again, the quantity of new H-1B workers is higher than the real one. In sum, there is evidence of large measurement error leading to the underestimation of the coefficient in the case where we don't use the instrumental variable.

Table 2.1: Impact of H-1B workers on patents by assignee

VARIABLES	(1) log Patent (linear)	(2) log Patent (with instr.)	(3) Patent (Poisson)	(4) Patent (Neg. Bin.)
log LCA STEM	0.062***	0.560***	0.071***	0.053***
	(0.0111)	(0.1950)	(0.0203)	(0.0070)
log Patent Stock	0.069**	0.360***	0.245***	0.347***
	(0.0270)	(0.0584)	(0.0547)	(0.0075)
Constant	2.044***			-0.891***
	(0.0913)			(0.0375)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	16,569	4,160	22,696	22,696
Number of firms	6,844	848	5,381	5,381
R-squared	0.082			
t-stat		2.07		

Robust standard errors in parentheses
*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 2.2 reports the outcome of the regression of new inventors over LCA by firm. The coefficient supports the presence of market size effect: it is positive and significant. According to the log-log regression without instrumental variable, a 10% increase in H-1B workers gives rise to a 0.2% increase in the number of new inventors in the following year. So, we observe firms who require more foreign employees today hiring more inventors even tomorrow. The estimates are robust to the Poisson and

Negative Binomial specification. As before, the coefficient in the 2SLS regression with the instrumental variable is much higher: the same 10% increase in LCA corresponds now to 4% more future inventors.

Table 2.2: Impact of H-1B workers on the number of new inventors

VARIABLES	(1) log New Inv. (linear)	(2) log New Inv. (with instr.)	(3) New Inv. (Poisson)	(4) New Inv. (Neg. Bin.)
log LCA STEM	0.024**	0.399**	0.050**	0.051***
	(0.0126)	(0.1880)	(0.0197)	(0.0087)
log Patent Stock	0.042	0.125**	0.050	0.306***
	(0.0255)	(0.0607)	(0.0428)	(0.0107)
Constant	1.658***			-0.644***
	(0.0933)			(0.0541)
Firm FE	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes
Observations	11,302	3,322	16,031	16,031
R-squared	0.061			
Number of firms	4,853	759	4,005	4,005
t-stat		2.12		

Robust standard errors in parentheses

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

Moving to the analysis about incumbent inventors, we can see from Table 2.3 that the probability of patenting again in the same firm in the subsequent years is positively related to number of LCA. The coefficient is significant in both regressions, with and without instrument.

H-1B workers have a positive impact on the patenting activity of incumbents also at the intensive margin, as shown in Table 2.4. The first three columns are log-log regressions without instrument while the last three are with the instrument. In general, a 10% increase in LCA translates into 0.2% more patents for incumbents (0.4% if we consider the 2SLS regression outcome). Hence, there is enough evidence supporting the complementarity between existing inventors and foreign newcomers.

Table 2.3: Impact of H-1B workers on incumbent inventors (extensive margin)

VARIABLES	(1)	(2) (with instr.)
log LCA STEM	0.035***	0.344***
	(0.0080)	(0.0518)
log Patent Stock	-0.598***	-0.756***
	(0.0265)	(0.0508)
Inventor Age	-0.0046***	-0.0062***
	(0.0016)	(0.0020)
Constant	-0.633***	-3.748***
	(0.0486)	(0.1770)
Firm FE	Yes	Yes
Year FE	Yes	Yes
Observations	506,515	391,764
Number of inventors	239,249	174,673
Number of firms	6,072	934

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

The distinction between escape competition and collaboration effect is harder to detect. In the second column (fifth for the 2SLS regression), we can see the impact of H-1B workers on the patents where the incumbent is the first inventor and doesn't have any new entry coauthor: a 10% growth in LCA brings in 0.12% (0.9% with the instrument) more patents. In the third column (sixth for the 2SLS regression), we observe the effect on the patents where the newcomer is the first inventor and the incumbent is one of the subsequent: a 10% increase in foreign workers leads to 0.24% (1.63% with the instrument) more patents. So, the estimates suggest the presence of both the escape competition and the collaboration effect. The latter is larger than the former but the difference is not significant.

Finally, we can turn to the citation analysis in Table 2.5. We can see that the increase in LCA gives rise to a higher diversity index by country: the margin is really small, a 10% rise corresponds to more 0.002 in the index, but it is significant at the

10% level. We can't say the same for the diversity index by state: in this case, the effect is positive but not significant. Conversely, a 10% increase in H-1B applications leads to a class diversity index higher by 0.003. Again, the effect is really small but it is significant at the 1% level. So, we can conclude that incumbents working in firms more dependent on H-1B visa cite patents more diverse in terms of country of origin and class. This gives us some evidence of positive spillovers stemming from the arrival of new foreign inventors.

Table 2.4: Impact of H-1B workers on incumbent inventors (intensive margin)

	no instrument			with instrument		
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	log Patents	First Inv.	Sub. Inv.	log Patents	First Inv.	Sub. Inv.
log LCA STEM	0.019***	0.012**	0.024**	0.040*	0.091***	0.163**
	(0.0032)	(0.0056)	(0.0114)	(0.0215)	(0.0313)	(0.0714)
log Patent Stock	-0.058***	0.078***	-0.112***	-0.034**	0.057**	-0.141***
	(0.0105)	(0.0176)	(0.0313)	(0.0142)	(0.0231)	(0.0476)
Inventor Age	-0.070***	-0.053***	-0.0153**	-0.134***	-0.049***	-0.003
	(0.0021)	(0.0036)	(0.0068)	(0.0083)	(0.0047)	(0.0118)
Constant	0.148**	-0.888***	-0.318	-0.057	-1.193***	-0.799**
	(0.0715)	(0.1180)	(0.2090)	(0.1220)	(0.1860)	(0.3820)
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	252,080	111,358	37,976	151,578	89,061	30,917
R-squared	0.037	0.009	0.014	0.036		
N. inventors	129,142	64,687	31,491	48,546	49,651	25,285
N. firms	7,042	6,732	3,518	613	607	519

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

2.5 Conclusions

There has always been great debate about immigration policies. On August 2, 2017 the RAISE (Reforming American Immigration for Strong Employment) Act was introduced in the United States Senate. The bill would not change the legislation with regard to H-1B temporary visas. However, it would introduce a visa-point system for

Table 2.5: Impact of H-1B workers on the diversity index of citations

	country index		state index		class index	
VARIABLES	(1)	(2)	(3)	(4)	(5)	(6)
	coeff.	margins	coeff.	margins	coeff.	margins
log LCA STEM	0.006*	0.002*	0.002	0.001	0.014***	0.004***
	(0.0039)	(0.0009)	(0.0040)	(0.0010)	(0.0043)	(0.0011)
log Patent Stock	0.003	0.001	-0.031***	-0.008***	-0.052***	-0.013***
	(0.0094)	(0.0022)	(0.0102)	(0.0026)	(0.0106)	(0.0026)
Self Citation Rate	-0.446***	-0.106***	0.569***	0.142***	0.325***	0.081***
	(0.015)	(0.004)	(0.020)	(0.005)	(0.018)	(0.004)
Constant	-0.518***		0.479***		-0.024*	
	(0.0113)		(0.0119)		(0.0124)	
Firm FE	Yes	Yes	Yes	Yes	Yes	Yes
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	191,024	191,024	191,024	191,024	191,024	191,024

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

employment-based permanent visas, which are the obvious solution for H-1B workers willing to stay in the United States in the long term. Hence, the RAISE Act might have some impact also in terms of high-skilled temporary workers.

The possible consequences for U.S. innovation are complex and cannot be inferred by reduced-form specifications. However, the outcome of my empirical analysis offers a valid baseline for a future model. Native and immigrant inventors appears to be complementary to each other. The data support the hypothesis of market-size effect. There is some evidence of both escape competition mechanism and positive spillovers. All these results represent a good starting point for designing a model of endogenous growth through which evaluate the economic impact of immigration policies.

Chapter 3

Mass Attrition: An Analysis of Drop Out from Principles of Microeconomics MOOC[29]

3.1 Introduction

A Massive Open Online Course (or MOOC) is a course given over the web for free, to anyone who is interested in signing up and has the basic technical capabilities to do so. The movement started in 2008 and took off in 2012, dubbed by the New York Times as The Year of the MOOC.[30] Some large platforms include Coursera, Udacity and edX. The potential of reaching a wide audience, both geographically and in terms of interests and demographics, is exciting. The chance of offering a lower-cost method of delivering education by the best professors and content creators is transformative (Cowen and Tabarrok, 2014; Acemoglu *et al.*, 2014). MOOCs have already reached millions of students, introducing them to topics as disperse as modern poetry and introduction to artificial intelligence. Champions of the MOOC movement believe that it will transform higher education in the USA by providing low cost alternatives to college credits as well as open academic level education to millions around the world. Some economists see them as direct competition to non-selective colleges (Hoxby, 2014).

A typical MOOC is composed of three elements. First, there are a series of videos

[29]Allione, G. and Stein, R. M. (2016). Mass attrition: An analysis of drop out from principles of microeconomics MOOC, *The Journal of Economic Education*, **47**(2), 174-186

[30]Laura Pappano, "The year of the MOOC," *The New York Times, November 4, 2012.*

that substitute for the typical lecture. The recommendation is to keep each segment short: 5-10 minutes each. The second element is composed of quizzes and/or problem sets that allow students to evaluate their own understanding and demonstrate their knowledge in order to receive recognition (for example, in the form of a certificate of accomplishment). The third element is the discussion forum that allows students to ask for help, on logistical, technology or content related issues, from peers and/or instructors.

The University of Pennsylvania joined as a partner with Coursera in 2012 and our nine-week course, Principles of Microeconomics, which took place during the spring of 2013, was one of the first courses to run as product of this partnership. As was typical of the first generation of MOOCs, we mimicked, subject to a few modifications,[31] a full semester course in length and scope. The course had a firm start date and completion date and though most of the work was asynchronous, students were expected to keep to a weekly schedule in terms of assignments and deliverables. Each week included 4-15 lecture video segments and a quiz of ten multiple-choice questions. There were six short answer assignments during the course and a multiple-choice final exam. Students with a grade of at least 65% earned a certificate of completion and those above 85% got one with distinction.

Taking a successful live course and translating it into the Mass Online Open environment is time consuming and challenging but overall rewarding. The lectures are pedagogically the easiest part to shift online. In a large class[32] there is minimal interaction with individual students and thus for most students, even in entertaining and engaging lectures, it is, at the core, a passive experience. Though easy to do ped-

[31]Topics that were omitted included consumer theory and factor markets. This was primarily done to keep the course in a nine week format which was set by the University.

[32]The typical lecture in Principles of Microeconomics at the University of Pennsylvania has 150-250 students. Students also attend a weekly recitation of 25 students led by a TA.

agogically, the recordings require both good technical support (we had a designated studio and editor) and a level of comfort talking to the camera and lecturing without the feedback of a live audience. The process is extremely time consuming and exhausting: each hour recorded takes about 4 hours of preparation, retakes, corrections etc. But, of course, this is a one-time fixed cost and future runs of the course will not require re-lecturing as they do in a live setting.

The most difficult aspect of the live course to translate to the MOOC environment was the response, critique and provision of feedback on assessments. Following the model developed in engineering and mathematics courses, we tried a variety of methods. Each week we had a quiz compiled of ten multiple-choice questions. As we were attempting to teach and not just evaluate, students were given feedback on incorrect answers and had two attempts to submit the quiz. We also had a mix of short answer assignments. Three were fill in the blank, mostly numeric, questions, which were easy to administer and evaluate. Three included a graphical application of the models to real world scenarios. To provide feedback and evaluation we chose the peer assessment model, in which each student who submitted an assignment one week, had to review, comment and evaluate five peer answers during the following week. We had some very vocal participants complaining about the time and effort this required, while other, more muted, voices thanked us for the high level of assessment and suggested that both completing the assignment and evaluating their peers were helpful learning experiences.

On-line discussion forums were set up to allow for students interaction with the instructors and with each other. This not only allows for timely communication on logistical and content related issues but also builds a community of learners. Five undergraduate economics students were hired to monitor the online forums, respond and, as need be, alert the teaching team when content clarifications would be helpful.

Each student checked into the site about one hour a day for the duration of the course.

Table 3.1 summarizes some basic course statistics that point to the stark difference between enrollment in a MOOC and a traditional course. In terms of magnitude, there were almost 36 thousand students enrolled. This size, typical of a MOOC, is equivalent of enrolling a whole state university in one course concurrently. However, enrollment numbers are misleading. Only 18 thousand students, one out of two enrollees, logged in even once and only 15 thousand watched, at a minimum, part of one lecture. Initial enrollment, in other words, is more akin to browsing a university brochure than registering for a live semester in a campus setting. By the second week only 6,486 students remained. 2,744 stayed to the end of the course and watched the last video lecture in week nine. 866 students attempted the final exam. Only 740 earned a certificate of accomplishment, which is the recognition that a student successfully mastered the material covered in the course.[33]

Table 3.1: Statistics on course activity

Activity	N. Participants
Enrolled in the course	35,819
Logged into the course at least once	18,009
Watched at least part of one video	15,427
Watched first video of the first week	13,483
Watched first video of the second week	6,486
Watched last video of the last week	2,744
Took the Final Exam	886
Earned the Statement of Accomplishment	740

This pattern of attrition is not unique to our course. It has been clear from the start that though MOOCs enroll tens of thousands of students each, only a very small fraction of them, between 5 and 12 percent, finishes the course and earns a certificate

[33] 472 students, out of 740 that passed the course, met the distinction criteria (score greater than 85%).

of completion (Koller *et al.*, 2013; Clow, 2013; Perna *et al.*, 2014). In comparison, for-credit academic online courses have retention rates between 88% (Xu and Jaggars, 2013) and 60% (Johnson and Cuellar Mejia, 2014). The summer online course taught by Rebecca Stein at the University of Pennsylvania has a retention rate around 60%. Critics of the MOOC movement cite the low retention rates as a sign of failure and proof of MOOCs limited impact.[34] As these courses are not given for credit, the low completion rate is not surprising – after all, many people pick up a book at the library and do not check it out, let alone read it or write a review of it. Koller *et al.* (2013) explain that the variety of intents of online course takers is very different from the traditional classroom model. In the MOOC framework, some students do indeed plan on completing the course and can be referred to as "Committed Learners". However, others may just want to understand what the topic is about or find out more about a particular professor or even about MOOCs as a whole. Those are the so-called "Browsers". Banerjee and Duflo (2014) note that attrition is larger for students that enrolled one day late, compared to students who enrolled on time. They interpret the result as an indication that less organized learners do worse in this new environment. Jordan (2014) explores enrollment and completion rates of mass online courses over the first 18 months of their existence and finds that though MOOCs size did not change, the completion rates, in most cases defined by earning a certificate of accomplishment, increased over this period (though the majority of them still have completion rates below 10% with a median of 6.5%). Completion rates are negatively correlated with course length and are lower when the course requires relatively more assignments, especially if they involve programming or peer assessment exercises, as indeed our course did.

[34] Maria Konnikova, "Will MOOCs be flukes?" *The New Yorker*, November 7, 2014. `http://www.newyorker.com/science/maria-konnikova/moocs-failure-solutions`.

We wanted to explore better the distinction between the Committed Learners and Browsers and learn whether the patterns of engagement are set early on, so that within a short period we can distinguish statistically between these two groups of students. Students had the opportunity to engage in the course not only by watching videos but also by participating in online forums (posting questions, responding to other posts, voting on posts to move them up or down on a priority list), solving the weekly quizzes and peer-assessed assignments, grading other students assignments and taking the final exam. Table 3.2 summarizes course engagement data during the first week of the course.

Table 3.2: Statistics on first-week activity

Activity	N. Participants
Watched the First Video in the First Week	13,483
Took First Quiz	4,571
Submitted First Peer Assessment	1,860
Graded First Peer Assessment (during 2nd week)	1,457
Participated in Forum Activity	684

Only about a third of the students who watched at least one video took the first quiz and only 18% submitted the first peer assessment. As quizzes and assignments were a major component of the grade in the course, failing to take these suggests that most of the enrollees behaved as passive learners or auditors rather than the fully engaged students we expect in a live, credit bearing course. In general, for-credit academic online courses perform worse than traditional courses in terms of retention rates (Xu and Jaggars, 2013; Johnson and Cuellar Mejia, 2014). There is an extensive literature studying the reasons behind this phenomenon (Hart, 2012). Some of the factors mentioned in this literature as associated with persistence are more easily applicable to the MOOC environment. These include general satisfaction with online learning, motivation, time management skills and family support. Others may be more challenging to create in a MOOC, including peer support, the sense of belonging to a learning community and increased communication with instructor. Our study does not test these factors directly. However, we included a voluntary opt-in survey that was available for completion throughout the course that collected both demographic information as well as a question asking for the students initial reason for taking this course.

As summarized in Table 3.3, out of the 3,215 students who submitted the survey, 65% percent were male and the predominant age group was 22-35 years old. Around 50% had taken a Coursera course before. Almost half of the participants were working full time and the vast majority were participating in the course in addition to other big time commitments such as studies, part time work etc. Students were also asked the reason why they enrolled in this course. About 60% knew nothing or little about Microeconomics and wanted to expand their knowledge. 25% of students chose to enroll in the course because it would help them in their job or life skills. 11.5% used the course as preparation for a future class in Microeconomics. Finally, 3.5% of participants were specifically interested in taking a course with the University of Pennsylvania and/or Rebecca Stein. Students represented 147 countries with the largest five being United States, India, China, Brazil, Spain. In our tables, we divide the countries into three categories: the USA, non-USA English speaking[35] countries and non-USA non-English speaking countries. We also matched each country with GDP per capita,[36] as a measure of economic development. Using the survey data, we are able to highlight key demographic characteristics such as age, gender, country of origin that are correlated with retention.

[35] Defined as countries where English is one of the official languages. The World Factbook. Central Intelligence Agency.

[36] GDP per capita, PPP (current international $). World Bank, International Comparison Program Database.

Table 3.3: Demographic data

Characteristic	Percentage
Gender	
Male	65.07%
Female	34.93%
Total	100.00%
Coursera Experience	
First time Coursera User	49.11%
Experienced Coursera User	50.89%
Total	100.00%
Country of Residence	
USA	23.64%
Non USA – English-speaking Country	24.63%
Non USA – Non English-speaking Country	51.73%
Total	100.00%
Occupational Status	
Student (pre-college)	5.27%
Student (undergraduate)	16.74%
Student (post-graduate)	10.19%
Employed full-time	46.82%
Employed part-time	3.24%
Self-employed	7.54%
Unemployed	6.20%
Other	3.99%
Total	100.00%
Age	
Under 18	4.00%
18 - 21	12.23%
22 - 35	56.52%
36 - 50	18.57%
51 - 70	7.99%
71 90+	0.69%
Total	100.00%
Reason for taking the course	
Learn the principles of Microeconomics	23.06%
Expand my knowledge about Microeconomics	37.45%
Preparation for future course	11.39%
Help with my job	11.24%
Help with my life skills	13.39%
Interested in a course with Penn and /or Rebecca Stein	3.46%
Total	100.00%

Source: Voluntary Survey (N = 3,215)

As we document in the results section, we find that those younger than 22 and college students are more likely to abandon the course and those who are over 35 are the most likely to stay. The strong correlation between age and retention is striking, given that the course was designed for a younger audience in the choice of examples and tone. This raises some doubts about the substitutability of MOOCs to more guided, individualized, coursework that forms that backbone of both online and in-class teaching even at non-selective colleges. Female enrollees are more likely to drop out as measured by video watching, though the difference is true mainly for full-time worker females against their male counterparts. Participants from USA are more likely to drop out. Retention is positively correlated with GDP per capita: it may be that logistical and technical constraints, such as limited bandwidth, impede those from poorer countries. Interestingly, among students who took the first quiz, neither Coursera experience nor being male predicts retention. This suggests that submitting the first quiz is a strong signal of commitment, since some of the differences between categories found in video retention disappear when looking at quiz retention, at least for those who took the survey.

As this is, to our knowledge, one of the first papers to attempt a sophisticated statistical analysis of retention in MOOCs, we take care to explain some technical issues regarding the Cox proportional hazard model, often used in duration analysis. Though many of the details have been relegated to footnotes, we hope the brief summary included here will be helpful to this developing literature. Full specifications of all regressions are available from the authors.

3.2 Model and Results

Compared to probability models, which focus on whether the event occurs, duration models target the time at which the event would possibly occur. This is the appropriate framework for our research question since we are not only interested in whether students finished the course but also in when they abandon it. Duration analysis allows us to capture the entire path of retention and not only the final result. In this way, we take full advantage of our individual level time-stamped data. In particular, we chose the Cox proportional hazard model because it delivers results that are easy to interpret.

The hazard rate is the probability of dropping out at time t conditional on surviving until that point. The proportionality assumes that there is an underlying baseline probability $\lambda_0(t)$ of dropping the course (due, for example, to random life events that interrupt the students ability to continue) which is not directly affected by the independent variables. Rather, the impact of a unit increase in the independent variable is multiplicative with respect to the hazard or dropout rate. The Cox proportional hazard model has the form:

$$\lambda(t|X) = \lambda_0(t)\exp(\beta_1 X_1 + \cdots + \beta_n X_n) = \lambda_0(t)\exp(\beta' X)$$

for the hazard rate at time t and an individual covariate vector X. The independent variables in vector X, were taken to be either measures of course connectivity, identified by the students activity during the first week of class, or demographic data.

As the name suggests, the Cox proportional hazard model assumes that the covariates affect dropout in a proportional manner that is time invariant. The proportionality assumption can be tested by regressing the scaled Schoenfeld residuals on functions of time. We can think of Schoenfeld residuals as the observed minus the expected values of the covariates at each failure time. If the proportionality assumption holds, the slope of the Schoenfeld residuals with respect to time should not be significantly different from zero. In cases where the zero-slope hypothesis is rejected, the proportionality assumption does not hold so we adjust our model by including the interaction term between the covariate and time.

We measure dropping out of the course in two ways: no longer watching videos, conditional on having watched the first one, and no longer submitting quizzes, conditional on having taken the first one. Therefore, in the model the period n is the nth video segment or the nth quiz. Obviously, the sample size in the latter specification is smaller, since not all the students that watched the first video took the first quiz.

In all the tables below we provide the hazard ratio instead of the coefficient. For example, a hazard ratio of 1.25 on the dummy variable "female" would suggest that females are 25% more likely to drop out in any given period compared to males; a hazard ratio of 0.7 would suggest that they are 30% less likely to drop out.

To see whether engagement during the first week impacts the hazard of dropping out, we run a model where the independent variables are measures of the first week activity, including quiz participation, completion of peer-assessed assignment and participation in a forum. Table 3.4 summarizes the results of these specifications. We reject the proportional hazard assumption[37] so we provide both time invariant and time variant coefficients.[38]

We can see that students taking the first quiz, whether or not they participated

[37]The proportionality assumption is rejected with probability $p < 0.01$, $\chi 2(5) = 662.52$ in the model of video retention while it holds in case of quiz retention, $\chi 2(3) = 2.86$. The test on each separate covariate is consistent with the global test. Results are available from the authors.

[38]We ran some robustness checks by including the survey participation indicator and the demographics variables. The coefficients on the first-week activity variables stay significant and qualitatively unchanged.

Table 3.4: Video and quiz retention by first-week activity

	Video		Quiz	
	(1)	(2)	(3)	(4)
Video Only - Forum	1.278***	1.043		
	(0.102)	(0.099)		
Quiz1 - No PA1[a] - No Forum	0.690***	0.502***		
	(0.018)	(0.017)		
Quiz1 - No PA1 - Forum	0.689***	0.520***	0.908	0.945
	(0.066)	(0.063)	(0.083)	(0.133)
PA1 - No Forum	0.359***	0.207***	0.384***	0.411***
	(0.013)	(0.011)	(0.015)	(0.026)
PA1 - Forum	0.320***	0.169***	0.338***	0.328***
	(0.023)	(0.019)	(0.025)	(0.040)
(Video Only Forum) * Time		1.019***		
		(0.005)		
(Quiz1 - No PA1 - No Forum) * Time		1.021***		
		(0.001)		
(Quiz1 - No PA1 Forum) * Time		1.019***		0.979
		(0.004)		(0.055)
(PA1 - No Forum) * Time		1.027***		0.972
		(0.001)		(0.020)
(PA1 Forum) * Time		1.029***		1.007
		(0.003)		(0.035)
N.	13,483	13,483	4,571	4,571

Notes: Hazard ratios (Standard Errors in Parentheses)

Omitted variables: Video Only - No Forum; Quiz1 - No PA1 - No Forum

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

[a] Peer Assessment 1.

in a forum, were about 30% less likely to drop the course compared to students who only watched the videos during the first week. Students taking the first peer-assessed assignment, whether or not they participated in a forum, were about 65% less likely to drop the course compared to students who only watched the videos during the first week. This suggests that participants in a MOOC belong indeed to different types and that activity in the first few days indicates their commitment to the course. The results are consistent with Banerjee and Duflo (2014) who find that students that enroll one day late are 17% less likely to earn a certificate than students that enroll on time, confirming that patterns of retention can be detected by early on behavior.

The hazard ratios in column (2) can be interpreted as follows: at the beginning of the course, the students that took the first quiz (with no participation in the forum and no peer assessment) were 50% less likely to drop out than those that only watched the videos. However, the difference shrinks by 2.1% in each period. This suggests that over time the first weeks activity loses its importance as an indicator of attachment to the course, as we should expect since those least committed to the course have dropped out.

Looking at participation in the online forum, we observe in column (1) that students who posted in the forum were 28% more likely to drop out of the course. However, once we add time interactions, we see that this effect is mainly due to their large drop off between the first and the second week. Why are these students, who took the time to participate in the forum during the first week, dropping out of the course is worth pursuing further. Our conjecture is that they used the forum to voice their frustration about the platform or the peer assessment and thus their postings and their defection are both expressions of the same barriers to learning. This pattern recalls the distribution of online product reviews, observed by some studies: there is a high concentration of reviews at the two extremes but very few in the middle (Hu *et al.*, 2009; Li and Hitt, 2008).

We next turn to analyzing student characteristics using data from an opt-in survey. Table 3.5 shows that survey participation is not random. Using the Cox proportional hazard model, we see that in terms of video watching, students taking the survey are 27% less likely to drop the course in each given period, while they are 35% less likely in terms of quiz completion. The same qualitative results still hold when we adjust the model to include the time interaction term.[39]

Table 3.5: Video and quiz retention by survey participation

	Video (1)	Quiz (2)
Survey = 1	0.732***	0.649***
	(0.017)	(0.022)
N.	13,483	4,571

Notes: Hazard ratios (Standard Errors in Parentheses)
* p < 0.1; ** p < 0.05; *** p < 0.01

There might be two possible reasons underlying the different behavior between those who did and did not complete the survey. It is likely to reflect self-selection of the more determined students into the set of survey takers, but it could also be the result of an endowment effect created by the time cost of filling out the survey. If indeed it is the act of engaging in this simple activity that reduces attrition, this suggests that even small changes in the early interaction with the students could greatly improve the effectiveness of MOOCs.

[39] The proportional hazards assumption is rejected with probability $p < 0.01$ in both specifications. $\chi^2(1) = 39.08$ for video retention and $\chi^2(1) = 15.55$ for quiz retention. Including the time interaction term, we find that the gap between those taking the survey and those not taking the survey evolves differently over time between the two measures of course retention. The difference shrinks by 0.6% per video while it increases by 7% per quiz. We also ran some robustness checks by including the first week activity variables. The coefficients show that participation in the survey predicts retention for students who submitted the first quiz or peer assessment but it predicts attrition for those who only watched videos in the first week. This confirms our hypothesis: students that did not actively engage in the first week activities are browsers rather than committed learners and they might have taken the survey for curiosity rather than commitment to the course.

The selection issue involving the survey participation does not imply that the demographic information is unusable; rather, we should understand that survey participants are positively self-selected so any result in this sense applies better to the population of "Committed Learners" rather than "Browsers".

Tables 3.6 and 3.7 summarize the impact of demographic characteristics on the hazard of dropping the course as measured by video watching.[40] All tables include prior Coursera experience, gender, whether the participant is from the USA or another English speaking country and GDP measures. Unfortunately, we do not have the actual age values since students were only asked about their age range. In all specifications, the omitted age category is 22-35 and the omitted occupation is full-time employment.

Students with Coursera experience are approximately 20% less likely to drop out. Various reasons may drive this result. Some of the Coursera beginners probably joined the course due to a curiosity towards the platform itself so their motivation for completion was lower from the start. It is also possible that, using their past experience, prior Coursera users tailor their expectations of the course leading to a lower risk of disappointment. Females are 15% more likely to drop out. The literature documents higher GPA and college completion among girls than boys overall (Di Prete and Buchmann, 2013) but our results are of a specific course in economics, where the gender gap is less surprising (Siegfried, 1979; Williams *et al.*, 2014; Ballard and Johnson, 2005). Our finding is robust to different specifications and we explore it further below, trying to identify its underlining causes. In the United States, at the time of the course, MOOCs were widely publicized in the popular press as well as in the academic world. We conjecture that some of the enrollment by American students

[40]There are very few cases where students took the survey but did not answer to all the questions. The observation was dropped if the missing variable was part of the specification.

Table 3.6: Video retention by demographics (focus on Reasons for taking the course)

	(1)	(2)
Experienced Coursera =1	0.785***	0.807***
	(0.032)	(0.033)
Female = 1	1.163***	1.154***
	(0.050)	(0.050)
USA = 1	1.322***	1.368***
	(0.077)	(0.082)
Non-USA English speaking = 1	1.011	1.04
	(0.053)	(0.055)
log(GDP per capita)	0.97	0.998
	(0.033)	(0.035)
log(GDP per capita)*Time	0.997***	0.997***
	(0.001)	(0.001)
Learn the principles of Microeconomics	1.067	1.054
	(0.058)	(0.058)
Preparation for future course	0.961	0.883*
	(0.067)	(0.063)
Help with my job	0.931	0.975
	(0.066)	(0.070)
Help with my life skills	1.066	1.094
	(0.068)	(0.071)
Interested in a course with Upenn/ Dr Stein	1.077	1.097
	(0.124)	(0.126)
Age Dummies	No	Yes
Occupation Dummies	No	Yes
N.	3,176	3,166

Notes: Hazard ratios (Standard Errors in Parentheses)

Omitted variable: Expand my knowledge about Microeconomics

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

Table 3.7: Video retention by demographics (focus on Age and Occupation)

	(1)	(2)	(3)
Experienced Coursera =1	0.805***	0.801***	0.807***
	(0.033)	(0.033)	(0.033)
Female = 1	1.136***	1.161***	1.154***
	(0.049)	(0.05)	(0.05)
USA = 1	1.354***	1.365***	1.368***
	(0.08)	(0.081)	(0.082)
Non-USA English speaking = 1	1.021	1.039	1.040
	(0.054)	(0.055)	(0.055)
log(GDP per capita)	1.021	1.039	1.040
	(0.054)	(0.055)	(0.055)
log(GDP per capita)*Time	0.997***	0.997***	0.997***
	(0.001)	(0.001)	(0.001)
Under 18	1.328***		1.005
	(0.135)		(0.176)
18-21	1.314***		1.077
	(0.082)		(0.096)
36-50	0.826***		0.844***
	(0.047)		(0.05)
51-70	0.862*		0.921
	(0.071)		(0.078)
71-90+	1.008		1.369
	(0.241)		(0.352)
Student Pre-College		1.411***	1.316*
		(0.129)	(0.212)
Student Undergraduate		1.337***	1.229**
		(0.078)	(0.099)
Student Graduate		0.995	0.959
		(0.072)	(0.07)
Employed Part-Time		0.945	0.947
		(0.112)	(0.112)
Self-Employed		0.916	0.946
		(0.076)	(0.08)
Unemployed		0.885	0.880
		(0.08)	(0.08)
Other		0.712***	0.672***
		(0.084)	(0.086)
Reasons for taking the course Dummies	Yes	Yes	Yes
N.	3,170	3,171	3,166

Notes: Hazard ratios (Standard Errors in Parentheses)

Omitted age variable: 22-35. Omitted occupation variable: Employed Full-Time

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

was to explore the MOOC phenomenon rather than to sample a specific course, so we expect attrition to be higher for this sub-sample. Indeed, participants from the USA are 30% more likely to quit the course. This fact is not driven by language because there is no significant difference between English vs non-English speakers among non-USA participants. A selection story is consistent with the USA hazard ratio. Non-American people are less exposed to advertising and information about Coursera, so gaining knowledge of the course required a greater search effort, justified by stronger motivation. In other words, it is reasonable to expect positive selection from other countries and thus a higher dropout rate from USA participants. On the other hand, we also need to take into account the presence of closer substitutes to Coursera in the USA than in the rest of the world. For example, American college students may attend a similar Microeconomics course in their own university but the same cannot be said of other countries.[41] This consideration leads us to include GDP per capita, in logarithmic term, in the regression. If the previous statement about substitutability holds, we would expect enrollees from poorer countries, where high-quality tertiary education is more scarce,[42] to stay longer in the course. As GDP per capita fails the proportionality test,[43] we include a time interaction on this variable. We find that as the course unfolded, participants from poorer countries became more and more likely to drop out. This suggests that factors such as technological gap, for example the speed of broadband connection, or income constraints come into play and prevail over the benefits that the course provides to participants of low GDP countries as a unique opportunity to access high-level tertiary education.

[41]In a separate specification, we included interactions between the USA dummy and other categories. We found that young students from the USA are extremely likely to drop out and they are the ones driving the result on the pooled coefficients.

[42]The Social Progress Index 2015, created by the nonprofit organization Social Progress Imperative, shows that access to advanced education is highly correlated with GDP per capita.

[43]The proportional hazard assumption is rejected with probability $p < 0.05$, $\chi^2(1) = 4.30$.

In Table 3.6 we explore whether the different reasons to take the course have any impact on retention. We run two specifications: one including age and occupation dummies and one without. Without adjusting for age and occupation we find that none of the reasons are associated with retention in a statistically significant way. Once we include age and occupation dummies, those taking the class as "Preparation for a future Microeconomics course" are 11.5% less likely to drop out in any period than those taking the class to "Expand their knowledge about Microeconomics". The result does not hold in the previous specification because most of the students in this category are under 21 and, as shown later, younger participants are, on average, more likely to drop out. In both cases, the F-test cannot reject the null hypothesis that all motivation categories have a coefficient equal to zero.[44]

In Table 3.7 we explore more carefully the impact of age and occupation. Because age and occupation categories tend to overlap, we run three separate specifications. (1) includes only the age categories, (2) only occupations and (3) both age and occupation categories. In each case the omitted age category is 22-35 and the omitted occupation is full-time employment. All the specifications still include the reasons to take the course as dummy variables, but the coefficients are not reported for sake of brevity.

We find that younger participants (specification 1) and those who are full time pre-college or undergraduate students (specification 2) are more likely to drop out. The likelihood to stick with the course is increasing with age. Those between the ages of 36 and 50 are 14% more likely to continue in any given period compared to those 22-35, while those younger than 22 are about 30% less likely to. Once we include both age and occupation dummies in the same regression (specification 3), the coefficients for the age category "Under 18" and "18-21" lose statistical significance. This is not

[44] $\chi^2(5) = 5.47$ for column 1 and $\chi^2(5) = 8.66$ for column 2.

surprising since they highly overlaps with Student Pre College and Student Undergraduate category respectively. The joint test rejects the null hypothesis, confirming our intuition.[45] Despite the fact that Principle of Microeconomics MOOC was closely structured on a successful campus course aimed at mostly freshman 18-year-old students, these age and occupational groups were precisely those least likely to stay in the course. Two reasons could explain this outcome: either this group starts off inherently less motivated, since they have access to their own college classes, or they need the support of the on-campus environment, both socially and pedagogically, more than other groups. If the latter is true, then the online MOOC structure may not be a substitute for in-class teaching, at least for this age category and for those with little college experience. It would be interesting to learn if these results change when allowing the Coursera certificate of completion to fulfill a college credit requirement.

We explore the gender issue in two directions: self-ascribed reason to take the course and occupation.[46] We find that the gender gap disappears in the subsample of students who enrolled because they wanted to take a course with a University of Pennsylvania faculty member or Dr Stein in particular. The outcome suggests that it is motivational for women to take a class taught by a female instructor. Clearly, this study cannot entirely validate the above hypothesis since it refers to a single course. However, it is an idea worth pursuing in a future work that analyzes multiple courses. The factor playing the most important role in the gender gap is occupation. This is not surprising: womens time commitment is very different from men. Comparing occupations within gender, we find that female graduate students, part-time workers and unemployed women are respectively 9%, 13% and 12% less likely to drop out

[45]The null hypothesis is rejected with $p < 0.05$ for "Under 18" joint with "Student Pre-College" and with $p < 0.01$ for "18-21" joint with "Student Undergraduate". $\chi^2(2) = 8.56$ and $\chi^2(2) = 18.16$ respectively.

[46]For sake of brevity, we do not report here the entire results but they are available from the authors upon request.

than female full-time workers.[47] However, these differences are not significant for male students. This is consistent with full-time working females having more time constraints than male counterparts: for the former adding an on-line course is a bigger challenge than for the latter.

Table 3.8 shows the results from a specification identical to Table 3.7, except that the dependent variable is quiz retention.[48] Since many of the coefficients that were significantly different from zero for video retention are not for quiz retention, taking the first quiz appears to be a strong signal of commitment to the course. Specifically, the coefficient on females is no longer statistically significant, suggesting that female students who engaged in the first quiz are not different from their male counterparts. Still, participants from the USA and the young are more likely to drop out as are, in this case, the very old.

3.3 Conclusions

What sets a MOOC apart from a textbook, beyond its being available online for free and the extensive use of videos, are two important aspects. One is the existence of clear deadlines and deliverable assignments; the second is the sense of community built both through the forums and via the knowledge that a set of students is working on a joint schedule. There is a natural tension between making MOOCs flexible and self-paced and building into them structure and a sense of community of learners. The more flexible is the course, the more the sense of community and the power of setting deadlines and short-term goals are in danger of being lost. There is room

[47] All the coefficients are statistically different from zero: the null hypothesis is rejected with $p < 0.1$ for Graduate Student and Part Time Worker females and with $p < 0.05$ for Unemployed females.

[48] We looked at the reasons for taking the course but none of the categories is statistically significant and we fail to reject the null hypothesis of all the coefficients being equal to zero. $\chi^2(5) = 1.31$.

Table 3.8: Quiz retention by demographics (focus on Age and Occupation)

	(1)	(2)	(3)
Experienced Coursera =1	0.940	0.944	0.947
	(0.051)	(0.051)	(0.051)
Female = 1	1.073	1.089	1.090
	(0.061)	(0.062)	(0.063)
USA = 1	1.311***	1.308***	1.302***
	(0.101)	(0.101)	(0.101)
Non-USA English speaking = 1	0.960	0.979	0.972
	(0.066)	(0.068)	(0.067)
log(GDP per capita)	0.977	0.977	0.981
	(0.052)	(0.052)	(0.052)
log(GDP per capita)*Time	0.972*	0.972*	0.972*
	(0.016)	(0.016)	(0.016)
Under 18	1.180	0.000	0.959
	(0.160)		(0.220)
18-21	1.260***		1.154
	(0.103)		(0.138)
36-50	0.934		0.947
	(0.069)		(0.072)
51-70	0.912		0.976
	(0.098)		(0.108)
71-90+	1.450		2.087**
	(0.445)		(0.690)
Student Pre-College		1.241*	1.202
		(0.149)	(0.254)
Student Undergraduate		1.145*	1.037
		(0.088)	(0.112)
Student Graduate		0.919	0.903
		(0.089)	(0.088)
Employed Part-Time		1.012	1.012
		(0.157)	(0.157)
Self-Employed		0.834*	0.839
		(0.092)	(0.094)
Unemployed		0.860	0.855
		(0.100)	(0.099)
Other		0.705**	0.629***
		(0.102)	(0.100)
Reasons for taking the course Dummies	Yes	Yes	Yes
N.	1,988	1,991	1,986

Notes: Hazard ratios (Standard Errors in Parentheses)

Omitted age variable: 22-35. Omitted occupation variable: Employed Full-Time

* $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$

for significant experimentation with course design in this new format to explore this tension. Students who need more flexibility and those who are good self-paced learners will benefit from less structure. On the other hand, many of us are familiar with the role deadlines and deliverables have in motivating us to complete tasks. The demographic data in our paper suggests that such a tension exists in the MOOC format too. As a whole, females in our course were more likely to drop out than males, but among them, those part-time workers were more likely to complete the course than full-time workers. This indicates that firm deadlines may be a problem if other aspects in ones life, such as a full-time job and family care, are nonnegotiable. However, it is not clear if more flexibility for the full-time workers, through, for example, self-paced modules, would assist completion or reduce the student motivation to reach the end, by undermining the same structure that is now the core strength of the course. We also found that young participants and those attending college were more likely to drop out. As the course was given in April through May, this could be due to a time conflict with other coursework or it could reflect young learners being more motivated by incentives such as course credit, which at this point was not given. However, another possibility is that young audiences are less skilled as independent learners and this group may need more individualized support. If the first two explanations are correct, MOOCs are far more likely to grow to be a core component of higher education than if the latter holds.

Any suggested changes should be focused on a well-defined target audience. As previous papers have documented, there are two groups of MOOC users, Committed Learners and Browsers, with different needs. To the extent that we see MOOCs as a Free Library, allowing browsing is efficient, as the marginal cost of the additional user is zero. But the fixed costs of setting up a MOOC are large. If one believes that the core benefits go to the Committed Learners, one needs to explore methods that keep these students engaged. First week activities that create a self-selection between motivated and less motivated learners allow us to focus on those most likely to complete and any evaluation of new course structure, such as increased flexibility, should be limited to and conditional on this group. This is an especially reasonable strategy if we believe that learning is a result of active engagement with the topic through a process of answering questions, making mistakes, getting feedback (be it either automatic or from peers). Watching videos is not tantamount to learning material and we may not want to focus on students who only engage at that level.

As Cowen and Tabarrok (2014) hypothesized, the large market size of MOOCs gives strong incentives for innovation and quality improvements. Though the course analyzed in this paper was given only a couple of years ago, the MOOC platform is already changing, in part to alleviate the low retention issues. For example, some Coursera courses are now given on an on demand schedule, so that students can start them at any given date and take them in shorter modules.[49] It will be interesting to see if the demographic effects, such as the lower retention rates among females and young participants, will stay the same in shorter, on-demand, courses. This new and fast evolving teaching platform is an opportunity to reach a wide audience and a better understanding of its strengths and limitations will help MOOCs achieve their full potential.

[49] "The Hype is Dead, but MOOCs Are Marching On." Knowledge@Wharton, January 5, 2015. Available at `http://knowledge.wharton.upenn.edu/article/moocs-making-progress-hype-died/`

Appendix A

Appendix to Chapter 1

A.1 On multiple migration opportunities

The estimation procedure I use implicitly assumes that each individual gets only one chance of migrating over his entire life. In case of future opportunities, the assumption is irrelevant since I am not interested in workers migrating after 2003. For past opportunities, the issue is more sensitive because, if the worker migrates, his home country past wages enter the moment function and they need to be consistent with the assumption that he didn't migrate before. I correct for the possible bias as follows.

1. In each period, I compute the probability of receiving a US job offer or winning the diversity lottery conditional on not having migrated:

$$P(EB|\text{stay}, \Omega) = \frac{(1 - P(\text{accept US job offer}|\Omega)) \cdot p(S, a)}{P(\text{stay}|\Omega)}$$

$$P(DV|\text{stay}, \Omega) = \frac{(1 - P(\text{apply DV lottery}|\Omega) \cdot q(S, a) \cdot (1 - p(S, a))}{P(\text{stay}|\Omega)}$$

$$\begin{aligned} P(\text{stay}|\Omega) = &(1 - P(\text{accept US job offer}|\Omega)) \cdot p(S, a) \\ &+ (1 - P(\text{apply DV lottery}|\Omega) \cdot q(S, a) \cdot (1 - p(S, a)) \\ &+ (1 - p(S, a)) \cdot (1 - q(S, a)) \end{aligned}$$

2. Then, I define the pool of workers that can be tempted to migrate, i.e. that will receive their chance of migrating in one of the subsequent periods. For them, I simulate the arrival of a US job offer or the winning of the diversity lottery, according to the above conditional probabilities.

3. I simulate the migration shock and, in case of a US job offer, the US income

shock. I draw the home-country income shock from the normal distribution, truncated from below at $\underline{X}$ where

$$\underline{X} = \begin{cases} -\infty & \text{no migration opportunities} \\ \underline{X}^{EB} & \text{US job offer} \\ \underline{X}^{DV} & \text{DV lottery winning} \end{cases}$$

and $\underline{X}^{EB}$ and $\underline{X}^{DV}$ are the respective thresholds for the home country income shock above which the worker decides to not migrate.

In this way, home country wages are consistent with the assumption that current potential migrants have not already migrated.

Bibliography

Acemoglu, D. (2002). Directed Technical Change. *Review of Economic Studies*, **69**(4), 781–809.

Acemoglu, D., Laibson, D., and List, J. A. (2014). Equalizing Superstars: The Internet and the Democratization of Education. *American Economic Review*, **104**(5), 523–527.

Akcigit, U. and Kerr, W. R. (2010). Growth through Heterogeneous Innovations. PIER Working Paper Archive 10-035, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Alesina, A., Harnoss, J., and Rapoport, H. (2013). Birthplace Diversity and Economic Prosperity. NBER Working Papers 18699, National Bureau of Economic Research, Inc.

Alker, J. R. and Kennan, J. (2011). The Effect of Expected Income on Individual Migration Decisions. *Econometrica*, **79**(1), 211–251.

Ballard, C. and Johnson, M. (2005). Gender, Expectations, And Grades In Introductory Microeconomics At A Us University. *Feminist Economics*, **11**(1), 95–122.

Banerjee, A. V. and Duflo, E. (2014). (Dis)organization and Success in an Economics MOOC. *American Economic Review*, **104**(5), 514–18.

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

Bertoli, S. (2010). The informational structure of migration decision and migrants self-selection. *Economics Letters*, **108**(1), 89–92.

Bertoli, S. and Rapoport, H. (2015). Heaven's Swing Door: Endogenous skills, migration networks and the effectiveness of quality-selective immigration policies. *Scandinavian Journal of Economics*, **117**(2), 565–591.

Bertoli, S., Dequiedt, V., and Zenou, Y. (2016). Can selective immigration policies reduce migrants' quality? *Journal of Development Economics*, **119**, 100–109.

Bianchi, M. (2013). Immigration Policy and SelfSelecting Migrants. *Journal of Public Economic Theory*, **15**(1), 1–23.

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

Borjas, G. J. (2004). Do Foreign Students Crowd Out Native Students from Graduate Programs? Working Paper 10349, National Bureau of Economic Research.

Borjas, G. J. (2005). The Labor-Market Impact of High-Skill Immigration. *American Economic Review*, **95**(2), 56–60.

Borjas, G. J. (2009). Immigration in High-Skill Labor Markets: The Impact of Foreign Students on the Earnings of Doctorates. In *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, NBER Chapters, pages 131–161. National Bureau of Economic Research, Inc.

Borjas, G. J., Kauppinen, I., and Poutvaara, P. (2018). Self-selection of Emigrants: Theory and Evidence on Stochastic Dominance in Observable and Unobservable Characteristics. *The Economic Journal*.

Burda, M. C. (1995). Migration and the Option Value of Waiting. SFB 373 Discussion Papers 1, Humboldt University of Berlin, Interdisciplinary Research Project 373: Quantification and Simulation of Economic Processes.

Chassamboulli, A. and Peri, G. (2018). The Economic Effect of Immigration Policies: Analyzing and Simulating the U.S. Case. NBER Working Papers 25074, National Bureau of Economic Research, Inc.

Chellaraj, G., Maskus, K. E., and Mattoo, A. (2008). The Contribution of International Graduate Students to US Innovation. *Review of International Economics*, **16**(3), 444–462.

Chiquiar, D. and Hanson, G. H. (2005). International Migration, Self-Selection, and the Distribution of Wages: Evidence from Mexico and the United States. *Journal of Political Economy*, **113**(2), 239–281.

Chiswick, B. R. and Miller, P. W. (2012). Negative and positive assimilation, skill transferability, and linguistic distance. *Journal of Human Capital*, **6**(1), 35–55.

Chiswick, B. R., Lee, Y. L., and Miller, P. W. (2005). Immigrant earnings: a longitudinal analysis. *Review of Income and Wealth*, **51**(4), 485–503.

Clow, D. (2013). MOOCs and the Funnel of Participation. In *Third conference on learning analytics and knowledge*, Leuven, Belgium.

Colussi, A. (2006). Migrants' networks: An estimable model of illegal Mexican migration.

Constant, A., Zimmermann, K., Kennan, J., and Walker, J. R. (2014). Modeling Individual Migration Decisions. In *International Handbook on the Economics of Migration*, pages 39–54.

Cowen, T. and Tabarrok, A. (2014). The Industrial Organization of Online Education. *American Economic Review*, **104**(5), 519–22.

Di Prete, T. and Buchmann, C. (2013). *The rise of women: The growing gender gap in education and what it means for American schools*. The Russel Sage Foundation.

DiNardo, J. and Card, D. (2000). Do Immigrant Inflows Lead to Native Outflows? *American Economic Review*, **90**(2), 360–367.

Gemici, A. (2011). Family migration and labor market outcomes.

Grogger, J. and Hanson, G. H. (2011). Income maximization and the selection and sorting of international migrants. *Journal of Development Economics*, **95**(1), 42–57.

Hall, B., Jaffe, A., and Trajtenberg, M. (2001). The NBER patent citations data file: Lessons, insights and methodological tools.

Hart, C. (2012). Factors Associated With Student Persistence in an Online Program of Study: A Review of the Literature. *Journal of Interactive Online Learning*, **11**, 19–42.

Hausman, J. A., Hall, B. H., and Griliches, Z. (1984). Econometric Models for Count Data with an Application to the Patents-R&D Relationship. *Econometrica*, **52**(17), 909–938.

Heckman, J. and Kautz, T. (2012). Hard Evidence on Soft Skills. *Labour Economics*, **19**(4), 451–464.

Heckman, J. and Singer, B. (1984). A method for minimizing the impact of distributional assumptions in econometric models for duration data. *Econometrica*, **52**(2), 271–320.

Hoxby, C. M. (2014). The Economics of Online Postsecondary Education: MOOCs, Nonselective Education, and Highly Selective Education. *American Economic Review*, **104**(5), 528–33.

Hu, N., Pavlou, P., and Zhang, J. (2009). Overcoming the J-Shaped Distribution of Product Reviews. *Communications of the ACM*, **52**, 144–147.

Hunt, J. (2011). Which Immigrants Are Most Innovative and Entrepreneurial? Distinctions by Entry Visa. *Journal of Labor Economics*, **29**(3), 417 – 457.

Hunt, J. and Gauthier-Loiselle, M. (2010). How Much Does Immigration Boost Innovation? *American Economic Journal: Macroeconomics*, **2**(2), 31–56.

Jaffe, A. B., Trajtenberg, M., and Henderson, R. (1993). Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations. *The Quarterly Journal of Economics*, **108**(3), 577–98.

Jasso, G. and Rosenzweig, M. R. (2009). Selection Criteria and the Skill Composition of Immigrants: A Comparative Analysis of Australian and U.S. Employment Immigration. In *Skilled Immigration Today: Prospects, Problems, and Policies*. Institute for the Study of Labor (IZA).

Jasso, G., Rosenzweig, M. R., and Smith, J. P. (2002). The earnings of US immigrants: World skill prices, skill transferability and selectivity.

Johnson, H. and Cuellar Mejia, M. (2014). Online Learning and Student Outcomes in California's Community Colleges. Technical report, Public Policy Institute of California.

Jordan, K. (2014). Initial Trends in Enrolment and Completion of Massive Open Online Courses. *International Review of Research in Open and Distance Learning*, **15**, 133–160.

Keane, M. P. and Wolpin, K. I. (1997). The Career Decisions of Young Men. *Journal of Political Economy*, **105**(3), 473–522.

Kerr, S. P. and Kerr, W. R. (2011). Economic impacts of immigration: a survey. *Finnish Economic Papers*, **24**(1).

Kerr, W. R. (2007). The Ethnic Composition of US Inventors. Harvard Business School Working Papers 08-006, Harvard Business School.

Kerr, W. R. (2008). Ethnic Scientific Communities and International Technology Diffusion. *The Review of Economics and Statistics*, **90**(3), 518–537.

Kerr, W. R. and Lincoln, W. F. (2010). The Supply Side of Innovation: H-1B Visa Reforms and U.S. Ethnic Invention. *Journal of Labor Economics*, **28**(3), 473–508.

Koller, D., Ng, A., and Chen, Z. (2013). Retention and Intention in Massive Open Online Courses: In Depth. Technical report, EDUCAUSE Review Online.

Lai, R., D'Amour, A., Yu, A., Sun, Y., and Fleming, L. (2001). Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010).

Lauderdale, D. S. and Kestenbaum, B. (2000). Asian American ethnic identification by surname. *Population Research and Policy Review*, **19**, 283–300.

Lessem, R. (2018). Mexico–U.S. Immigration: Effects of Wages and Border Enforcement. *The Review of Economic Studies*, **85**(4), 2353–2388.

Lessem, R. and Sanders, C. (2013). Decomposing the Native-Immigrant Wage Gap in the United States.

Li, X. and Hitt, L. M. (2008). Self-Selection and Information Role of Online Product Reviews. *Information Systems Research*, **19**(4), 456–474.

Locher, L. (2001). Testing for the Option Value of Migration. *IZA Discussion Papers*, (405).

M. Holland, M. (2016). The Rise of Women: The Growing Gender Gap in Education and What it Means for American Schools by Thomas A. DiPrete and Claudia Buchmann. *NASPA Journal About Women in Higher Education*, **9**(2), 230–232.

Maskus, K., Mobarak, A. M., and Stuen, E. T. (2010). Skilled Immigration and Innovation: Evidence from Enrollment Fluctuations in U.S. Doctoral Programs. CEPR Discussion Papers 7709.

Massey, D. S., Jasso, G., and Espinoza, M. (2017). Weighting for Nonresponse on Round Two of the New Immigrant Survey. Technical report.

Mayer, T. and Zignago, S. (2011). Notes on CEPII's distances measures: The GeoDist database.

Niebuhr, A. (2006). Migration and innovation: Does cultural diversity matter for regional R&D activity? HWWI Research Papers 3-1, Hamburg Institute of International Economics (HWWI).

Ortega, F. and Peri, G. (2013). The effect of income and immigration policies on international migration. *Migration Studies*, **1**(1), 1–35.

Ottaviano, G., Ireo, P., and Peri, G. (2005). Rethinking the Gains from Immigration: Theory and Evidence from the US. CEPR Discussion Papers 5226.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of Applied Econometrics*, **11**(6), 619–632.

Pekkala Kerr, S. and Kerr, W. R. (2013). Immigration and Employer Transitions for STEM Workers. *American Economic Review*, **103**(3), 193–97.

Perna, L., Ruby, A., Boruch, R., Wang, N., Scull, J., Ahmad, S., and Evans, C. (2014). The Life Cycle of a Million MOOC Users.

Rendon, S. and Cuecuecha, A. (2010). International Job Search: Mexicans in and out of the US. *Review of Economics of the Household*, **8**(1), 53–82.

Ruiz, N. G., Wilson, J. H., and Choudhury, S. (2012). The Search for Skills: Demand for H-1B Immigrant Workers in U.S. Metropolitan Areas.

Siegfried, J. J. (1979). Male-Female Differences in Economic Education: a Survey. *The Journal of Economic Education*, **10**(2), 1–11.

Stark, O., Byra, L., Casarico, A., and Uebelmesser, S. (2017). A critical comparison of migration policies: Entry fee versus quota. *Regional Science and Urban Economics*, **66**, 91–107.

Wadhwa, V., Saxenian, A., Rissing, B., and Gereffi, G. (2007). America's New Immigrant Entrepreneurs. Master of Engineering Management Program, Duke University; School of Information, U.C. Berkeley.

Wadhwa, V., Saxenian, A., and Siciliano, F. D. (2012). America's New Immigrant Entrepreneurs: Then and Now, Part VII. Ewing Marion Kauffman Foundation Research Paper.

Williams, M. L., Waldauer, C., and G. Duggal, V. (2014). Gender Differences in Economic Knowledge: An Extension of the Analysis. *The Journal of Economic Education*, **23**, 219–231.

Wooldridge, J. (2002). *Econometric analysis of cross section and panel data*. Massachusetts Institute of Technology 2002 .

Xu, D. and Jaggars, S. S. (2013). The impact of online learning on students' course outcomes: Evidence from a large community and technical college system. *Economics of Education Review*, **37**(C), 46–57.

Yoon, C. (2012). The Decline of the Rust Belt: A Dynamic Spatial Equilibrium Analysis.