Selective Immigration Policies and the U.S. Labor Market

JOAN LLULL[*§]
MOVE, Universitat Autònoma de Barcelona, and Barcelona GSE

Preliminary and Incomplete
February 2020

While immigration of unskilled workers often generates controversy in the political arena, there is often more consensus in favor of selective immigration policies. This paper studies the effects of selective immigration policies on the labor market. High skilled immigration introduces two potentially confronting forces on labor market prospects of native workers: first, it increases the competition for skilled jobs, reducing labor market opportunities, and, as a result, reducing native incentives to invest in human capital; second, it increases productivity through spillovers and technological progress. I pose and estimate a labor market equilibrium dynamic discrete choice model that can account for these effects. The estimated model is used to evaluate the labor market consequences of the two most important skill-biased immigration policies in recent U.S. history: the introduction of H-1B visa program in 1990, and the elimination of the National Origins Formula in 1965. I also use the model to simulate the level of selectivity of immigration policy that maximizes native workers' wellbeing.

I. Introduction

The Syrian refugee crisis has put immigration policy back at the core of the political debate in many developed countries. The possibility of exerting a larger control on immigration and of having stronger borders was one of the main arguments used to support Brexit, it was a salient issue in President Donald J. Trump's presidential campaign, who proposed the construction of a wall on the Mexico-

[*] Departament d'Economia i Història Econòmica. Universitat Autònoma de Barcelona. Campus de Bellaterra – Edifici B, 08193, Bellaterra, Cerdanyola del Vallès, Barcelona (Spain). URL: http://pareto.uab.cat/jllull. E-mail: joan.llull [at] movebarcelona [dot] eu.

[§] I would like to thank Jeff Campbell, Ellen McGrattan, Martí Mestieri, Robert Miller, Víctor Ríos-Rull, Luis Rojas, Raül Santaeulàlia-Llopis, Elena T. Aguilar, seminar participants at UPenn, NYU, Chicago Fed, WashU/St.Louis Fed, Georgetown, UCLA, Carnegie Mellon, CReAM (UCL), Toulouse School of Economics, Oxford, IIES-Stockholm, Carlos III, Southampton, Bristol, Geneva, Autònoma de Barcelona, Girona, Bolzano, IEB/Universitat de Barcelona, Illes Balears, and conference participants at SAEe and Barcelona GSE Winter Workshop, for enlightening comments and discussions. Daniele Alimonti, Maja Garbulinska, Rim Mesraoua, Minyoung Rho, and Anqi Shan provided excellent research assistance. This work has been supported by the European Research Council (ERC), through Starting Grant n. 804989, the Spanish Ministry of Economy and Competitiveness, through grant ECO2014-59056-JIN, and the Severo Ochoa Programme for Centers of Excellence in R&D (SEV-2015-0563), by the Fundació Bancària "La Caixa" (CG-2017-05), and by the Generalitat de Catalunya (2017-SGR-1765).

United States border, and it has been central in several (some of them successful) presidential campaigns in many European countries.[1] Unlike with general immigration, the policies that favor the inflow of highly skilled workers are generally less challenged. Immigration policies in Canada, Australia, and the United Kingdom are mostly based on points systems that favor immigration of skilled workers, and Japan recently implemented a similar policy. Canada recently launched as well the Global Skills Strategy, a program that gives Canadian employers faster access to highly skilled foreign workers. In the United States, the H-1B visa program and some of the executive actions approved by President Barack Obama in November 2014 aimed at facilitating immigration of highly skilled workers. And even President Trump, who has questioned the efficacy of the H-1B program in bringing in high quality immigration, and has announced changes to prevent fraud and to make it more selective, has not threatened its existence.[2]

Are there economic gains of selective immigration policies relative to general immigration? Whose labor market prospects are improved by skilled immigration, whose are worsened, and by how much? Does high skilled immigration discourage native investments in human capital? Are there knowledge spillovers from skilled immigration? What level of selectivity in immigration policy maximizes natives' utility? The answers to these questions are fundamental for the design of immigration policy. In this paper, I provide answers to these questions by posing and estimating a labor market equilibrium dynamic discrete choice model. The estimated model is used to simulate a different set of positive and normative counterfactual exercises that provide the relevant answers. The positive analysis focuses on the first four questions. These simulations evaluate the economic impact of the two most important skill-biased immigration policies in recent U.S. history: the introduction of the H-1B visa program in 1990, and the elimination of the National Origins Formula in 1965.[3] In the normative analysis, I quantify

[1] For example, Jörg Haider and Norbert Hofer in Austria, Gábor Vona and Viktor Orban in Hungary, Geert Wilders in the Netherlands, Nikos Michaloliakos in Greece, Timo Soini in Finland, Frauke Petry in Germany, Kristian Thulesen Dahl in Denmark, Gianluca Iannone in Italy, Björn Söder in Sweden, and Marine Le Pen in France.

[2] In April 2017, President Trump signed the executive order "Buy American and Hire American" which dictates, among other things, that "in order to promote the proper functioning of the H-1B visa program, the Secretary of State, the Attorney General, the Secretary of Labor, and the Secretary of Homeland Security shall, as soon as practicable, suggest reforms to help ensure that H-1B visas are awarded to the most-skilled or highest-paid petition beneficiaries". (The White House Press Office, April 18, 2017).

[3] The H-1B visa program is a guest program targeted to attract workers in Science, Technology, Engineering and Mathematics (STEM) fields for a once-renewable period of three years, after which the employer can sponsor the worker for permanent residency. The main users of this program are immigrants from India, China, and other Asian countries. The National Origins Formula was a quota system in place between 1924 and 1965 that selected immigrants on the

the level of selectivity of the immigration policy that maximizes the wellbeing of native workers. On the one hand, attracting more skilled workers enhance native productivity through externalities and knowledge spillovers. On the other hand, skilled immigration poses competition on skilled natives, which may discourage investments in the first place.

My modeling approach builds on Llull (2018a), who shows the importance of accounting for native human capital and labor supply adjustments to quantify labor market effects of the increase in (mostly unskilled) immigration in the United States over the last four decades. Unlike Llull (2018a), the estimated model accounts for two confronting forces of high skilled immigration in the labor market. First, the model allows for a potential externality through knowledge spillovers and endogenous technological progress. The model features the generation of ideas as a consequence of using skilled labor (and equipment capital) in production. This generation of ideas and knowledge spillovers endogenously produce neutral and skill-biased technological change, which affects the productivity of other labor market inputs (and capital). The second one, is a competition effect through the labor market equilibrium. The inflow of skilled workers puts downward pressure on wages of competing workers (the closest substitutes are skilled natives), reducing their incentives to invest in human capital in the first place. These two confronting forces make the determination of the level of selectivity in the immigration policy that native workers are willing to accept an empirical question.

The estimated model is a labor market equilibrium model with knowledge spillovers. On the labor demand side, a representative firm (which represents the behavior of a continuum of atomistic firms) combines three types of labor (blue collar, white collar, and STEM) and two types of capital (structures and equipment) to produce a single output. In doing so, the firm generates ideas as a by-product of using equipment capital and STEM labor in production. Ideas generate productivity spillovers changing the relative demand for different inputs, and also fostering factor neutral technological progress. As a representation of atomistic firms, the representative firm does not take into account these productivity enhancements in its labor demand, and, therefore, this constitutes an externality. On the labor supply side, heterogeneous individuals decide on education, participation, and occupation over their life cycle. Individuals are heterogeneous

basis of national origin in order to preserve the ethnic mix of the U.S. population. The removal of the National Origins Formula in 1965 completely reshaped the skill composition of immigrants in the U.S., switching from a relatively educated Western immigration to a less educated Latin American and then Asian one (Borjas, 1993, 1995; Antecol, Cobb-Clark and Trejo, 2003).

in many dimensions, including, in the case of immigrants, national origin, which is essential both to determine the prevalence of H-1B visas, and to simulate the National Origins Formula.

The model is estimated using U.S. micro-data from both the Current Population Survey (CPS) for survey years 1993-2016 (March Supplements, linked over two consecutive years) and the Survey of Income and Program Participation (SIPP) for survey years 1984-2014, along with national-level data for some of the aggregate variables. One of the central aspects of this paper is to credibly identify the knowledge externality. Beyond functional form assumptions, and econometric implementation, it is fundamental to understand the variation from the data that identifies the spillovers. To this end, it is crucial to have a measurement of the stock of ideas. In this paper, I use two alternative variables to measure it: the accumulation of intellectual property products (IPP) capitalized in the National Income and Product Accounts (NIPA) recently revised by the Bureau of Economic Analysis (BEA), and the number of patents generated in the U.S. in a given year. Armed with such measurements, the model interprets the data as follows. Changes in wages that follow an exogenous change in STEM labor supply (or capital equipment) holding fixed the stock of ideas (e.g. a negative "ideas shock" offsets the change in labor inputs) are interpreted as elasticities of substitution across labor inputs. On the contrary, changes in the stock of ideas that do not follow any change in labor or capital inputs identify the externality.

The estimation of the model using full solution methods (like in Lee and Wolpin, 2006, 2010; Llull, 2018a) is computationally too demanding. Alternatively, I estimate the model using conditional choice probability (CCP) estimation, combining techniques and arguments developed by Hotz and Miller (1993), Hotz, Miller, Sanders and Smith (1994), Altuğ and Miller (1998), Aguirregabiria and Mira (2002), and Arcidiacono and Miller (2011) (see Arcidiacono and Ellickson (2011) for a review). CCP estimation avoids solving for the value functions in each iteration of the parameter search, which is very costly. Additionally, it has the advantage of transparently presenting parameter identification, as well as allowing for sensitivity analysis to different functional form assumptions and different datasets.

Altuğ and Miller (1998) is the first (and, to my knowledge, only) paper that applies CCP estimation methods to models that feature aggregate shocks. They present a labor supply (and consumption) decision model that allows for aggregate conditions that are consistent with Pareto optimal allocations. As these authors note, the implementation of this class of CCP estimators requires estimates for the CCPs for different (counterfactual) realizations of the aggregate shocks in order to correctly specify expectations about the future. Altuğ and Miller (1998) obtain these CCPs exploiting the variation in the shadow value of consumption, heterogeneous across individuals, along with stationarity. For example, one can predict the behavior of a wealthy individual living in an economic slump by observing the behavior of a poorer individual living in a prosperous world. They can use this argument in identification because they have data on consumption. Absent consumption information in my data, I show that one can still exploit the stationarity of the model along with the equilibrium structure to infer the counterfactual CCPs from time series variation in aggregate conditions. Given stationarity, calendar time only affects labor supply decisions through skill prices (driven by the aggregate shock), and thus constitutes a sufficient statistic for the aggregate shock in the observed baseline economy. This allows me to recover equilibrium skill prices by estimating the wage equations using current period baseline CCPs. Having recovered them, I reestimate the CCPs (now conditional on skill prices). In doing so, I interpret periods with high skill prices as counterfactuals for periods with low prices, had these prices been high.

This paper contributes to the growing literature on the consequences of skilled immigration.[4] Different strands of this literature analyze: the effect of skilled immigration on patenting and entrepreneurship (Hunt and Gauthier-Loiselle, 2010; Kerr and Lincoln, 2010; Hunt, 2011); the displacement effect on native STEM employment and wages (Borjas, 2009; Kerr and Kerr, 2013; Bound, Braga, Golden and Khanna, 2015; Peri, Shih and Sparber, 2015; Kerr, Kerr and Lincoln, 2015; Doran, Gelber and Isen, 2016; Bound, Khanna and Morales, 2017; Ma, 2017); the career prospects of science and technology immigrants relative to similar natives (Gaulé and Piacentini, 2013; Hunt, 2015); the crowding out effects of H-1B visas on native students (Kato and Sparber, 2013; Orrenius and Zavodny, 2015); and the competition and spillover effects in the space of ideas, exploiting evidence on massive inflows of foreign professors (Borjas and Doran, 2012, 2015a,b, and Moser, Voena and Waldinger, 2014).

Among these strands of the literature, my paper is mostly related to the first two.[5] Except for Bound et al. (2015), Bound et al. (2017), and Ma (2017), all these papers use a reduced form approach. The structural approach allows me to

[4] More generally, it also contributes to the general literature on immigration on wages (Borjas, 2003; Card, 2009; Ottaviano and Peri, 2012; Dustmann, Frattini and Preston, 2013; Llull, 2018a,c; see Dustmann, Schönberg and Stuhler, 2016 for a recent survey).

[5] It is also related to the last strand, even though this group studies a narrower population of interest, namely Soviet mathematicians migrating to the U.S. after the collapse of the Soviet Union, and Jewish chemists migrating from Nazi Germany. The model is also able to speak to the remaining two strands, but they are less central to the main contribution of the paper.

expand this literature in several dimensions. First, I take into account potential displacement effects not only on employment of natives, but also on their decisions to invest in education and to become STEM workers. Second, the structural model allows me to identify the spillover effects on wages through the generation of knowledge, measured as the accumulation of patents, or IPP capital. For example, while Hunt and Gauthier-Loiselle (2010) and Kerr and Lincoln (2010) analyze the effect of skilled immigration on patenting; I additionally measure how this extra intangible capital maps into higher wages for natives, and how that affects their incentives in the labor market. Third, I quantify the heterogeneous effects of size and selectivity of immigration policy across different groups of native workers and I use the estimated model to characterize the level of selectivity of the immigration policy that maximizes native workers' wellbeing. Four, I take into account capital-skill complementarity, which is important to correctly measure the wage impact of skilled (and unskilled) immigration (see Lewis, 2011, 2013, and Llull, 2018a), and I allow for skill-biased technological spillovers from high skilled immigration. And five, on top of the H-1B visa policy, the estimated model is also used to evaluate the National Origins Formula as a selective immigration policy.

Bound et al. (2015) present a macro-calibrated model of partial equilibrium for the market of computer scientists. Even though they focus on a much narrower market (computer scientists), their paper is more related to mine than the others above as it takes into account the change in the supply of prospective workers with computer science majors (education). It also allows, in a reduced form way, for spillovers of immigration on overall productivity, even though their mechanism is not endogenous. Bound, Khanna and Morales (2017) expand their previous model to endogenize technological change, linking productivity increases in the U.S. during the 1990s to increase in the utilization of computer scientists in the economy. Ma (2017) goes beyond the computer scientist market, and estimates a model in which natives and immigrants can work in either computer science or in other STEM jobs. She studies the potential crowding out of natives into other STEM fields, which is the differentiation mechanism that allows them avoid competition from immigrants. She, however, abstracts from knowledge spillovers, and from effects on non-STEM occupations.

More broadly, my paper is also related to other literatures. First, it is related to the literature that estimates dynamic labor market equilibrium models of career choices (Altuğ and Miller, 1998; Heckman, Lochner and Taber, 1998; Lee, 2005; Lee and Wolpin, 2006, 2010; Llull, 2018a). Moreover, it contributes to the literatures that analyze skill-biased technical change and capital-skill complementarity (e.g., Krusell, Ohanian, Ríos-Rull and Violante, 2000; see Acemoglu, 2002; Acemoglu and Autor, 2011 for surveys), and knowledge spillovers in aggregate production (Romer, 1986; Lucas, 1988; see Klenow and Rodríguez-Clare, 2005 for a review), providing a model that features endogenous neutral and skill-biased technological change. Finally, it relates to the macro literature that explores the role of intangible capital in explaining the recent evolution of labor productivity and the labor share (e.g., McGrattan and Prescott, 2010, 2014; Koh, Santaeulàlia-Llopis and Zheng, 2016).

The rest of the paper is organized as follows. Section II introduces some policy background and descriptive statistics. Section III presents the model. Section IV discusses identification. Section V introduces the estimation procedure. Section VI presents the estimated parameters of the model and evaluates the goodness of fit. Section VII presents simulation results for the different policy experiments. And Section VIII concludes.

II. Immigration in the United States: Policy and Facts

A. U.S. Immigration Policy Background: From the National Origins Formula to the H-1B Visa Program

Throughout its history, the United States has been a nation of immigrants. From colonial times to mid-nineteenth century Western European immigrants (especially from Britain and Ireland, but also from Germany and Scandinavia) kept entering the U.S. without any federal legislation (Ewing, 2012). Beginning in 1850s, the so-called "new immigration" brought in immigrants from Eastern and Southern Europe as well as from Asia and Russia. Americans' preference for "old" rather than "new" immigration reflected a sudden rise in conservatism and the appearance of the first nativist movements. In 1875 the first federal immigration law was passed; this law prohibited the entrance of criminals and convicts, prostitutes, and Chinese contract laborers. This law paved the road for the 1882 Chinese Exclusion Act, which almost prohibited Chinese workers to enter the United States.[6] It was the first of many laws that targeted specific ethnic groups, starting a bias against Asian that lasted until 1952.[7]

The Immigration Act of 1917 defined a "barred zone" of nations in the Asia-Pacific triangle from which immigration was prohibited. In 1921 the U.S. Congress

[6] Later on, Chinese were issued Japanese passports to enter the United States. In 1907, a "Gentleman's Agreement" with Japan effectively ended with Chinese and Japanese immigration.

[7] The Immigration and Naturalization Act of 1952 was the first step towards removing racial distinctions from U.S. immigration policies.

passed the Emergency Quota Act, which limited the annual number of immigrants to be admitted from any country to a maximum of the 3% of the number of persons from that country living in the U.S. in 1910; in 1924, the share was reduced to 2% and the reference year was switched to 1890. It was the birth of the National Origins Formula. The Immigration and Nationality Act of 1952 consolidated this system setting the quotas for each country to one sixth of one percent of the number of persons of that ancestry living in the United States as of 1920. These restrictions, aimed at preserving the ethnic composition of U.S. population, reserved most immigration slots for immigrants from the United Kingdom, Ireland, and Germany (Ewing, 2012).

The 1965 Amendments to the Immigration and Nationality Act radically changed U.S. immigration policy. The National Origins Formula was abolished, and replaced by aggregate limitations (initially by hemisphere, worldwide from 1976) with a maximum amount per country (common to all of them). The new policy also allowed to issue an unlimited amount of visas to immediate relatives (parents, spouses and children) of U.S. citizens and legal immigrants.[8]

The 1965 Amendments were not aimed at fostering immigration or changing the ethnic composition of immigrant inflows. They were rather a reaction to the civil rights movements during 1960s. According to the speech by President Lyndon B. Johnson when he signed the legislation into law, the reform was not "revolutionary": "it does not change the lives of millions" he said. Several of the promoters of the reform defended it in the debate at the U.S. Senate. Senator Edward M. Kennedy enumerated what, according to his view, the policy would not do: "First, our cities will not be flooded with a million of immigrants annually. [...] Secondly, the ethnic mix of this country will not be upset". Several representatives expressed the same opinion.[9] Many of them, emphasized that "the proposed

[8] The quotas from the National Origins Formula were not of application for the Western Hemisphere (Canada, Latin America, and the Caribbean). However, the legal immigration process from Latin America and the Caribbean was very costly. These large costs kept immigration from these countries not very far from the levels that would be implied the quotas. In 1942, the U.S. government introduced a large-scale program of temporary immigration of Mexican workers, the so-called *bracero* program. The costly process of immigrating permanently to the U.S. fostered an important increase in unauthorized immigration (through *braceros*' overstays). In 1954, under the "Operation Wetback", about one million Mexican immigrants were deported (Ewing, 2012). Even though this program ended in 1964, the introduction of the family reunification visa completely transformed the immigration process from Mexico.

[9] Among others, interventions along these lines included those from Senators Edward and Robert Kennedy, Hart, Fong, Scott, Pell, Williams, Kuchel, Bartlett, Inouye, McCarthy, McNamara, Moss, Proxmire, the Secretary of Labor Willard Wirtz, and the Secretary of State Dean Rusk. Their interventions are transcribed in the Senate Part 1, Book 1 as made available at `http://vdare.com/articles/so-much-for-promises-quotes-re-1965-immigration-act`, accessed March 23th, 2017.

legislation would not greatly increase the number of immigrants" (Senator Eugene McCarthy) and highlighted that the ethnic mix would not change: "the people from that part of the world [the Asia-Pacific Triangle] probably will never reach 1 percent of the U.S. population" (Senator Hiram Fong). However, as we discuss in Section II.B, removing the National Origins Formula not only changed the ethnic mix of the country drastically, but also precluded the radical change in the skill composition of immigrant inflows that followed.

After subsequent policies mostly focused on preventing illegal immigration (e.g., the 1986 Immigration Reform and Control Act (IRCA), followed by an amnesty, and the 1996 Illegal Immigration Reform and Immigrant Responsibility Act), one of the most important policy changes after 1965 came with the 1990 Immigration Act which restricted the number of visas to be issued to immediate relatives of previous immigrants and U.S. citizens, and established a preference skilled immigration. This policy introduced the H-1B visa for guest skilled immigrant workers in "specialty occupations". These occupations are defined as requiring theoretical and practical application of a body of highly specialized knowledge in a field of human endeavor, including, but not limited to, architecture, engineering, mathematics, physical sciences, social sciences, medicine and health, education, law, accounting, business specialties, theology, and art (Bound et al., 2015). Applicants have an educational requirement of at least a bachelor's degree. H-1B visas restrict holders to work only for the company that sponsored them. The standard duration of an H-1B visa is for a stay of three years, renewable up to a maximum of 6 years. However, firms can sponsor H-1B visa holders for a permanent resident visa.[10] Since 1990 there has been a cap in the total number of H-1B visas that can be issued, which has been binding since mid 1990s (except for years 2000-2003, in which the cap was threefold increased). Since year 2000, universities and non-profit research facilities are excluded from the cap. The H-1B visa has been instrumental in the recent increase in highly skilled immigration, especially from India, but also, to a lesser extent, from China and other parts of Asia.

B. Ethnic and Skill Composition of Immigration

This section provides descriptive statistics that offer a general picture about the evolution of the skill composition of immigration in the United States over the recent decades. It also shows the evolution of the national origin composition,

[10] Before H-1B, the Immigration and Nationality Act of 1952 introduced the H-1 visa, targeted at guest workers of "distinguished merit and ability" (Bound et al., 2015). However, they never became as popular as H-1B visas because the H-1B program introduced the possibility of transferring to permanent immigrant status, which made it more appealing for workers and firms.

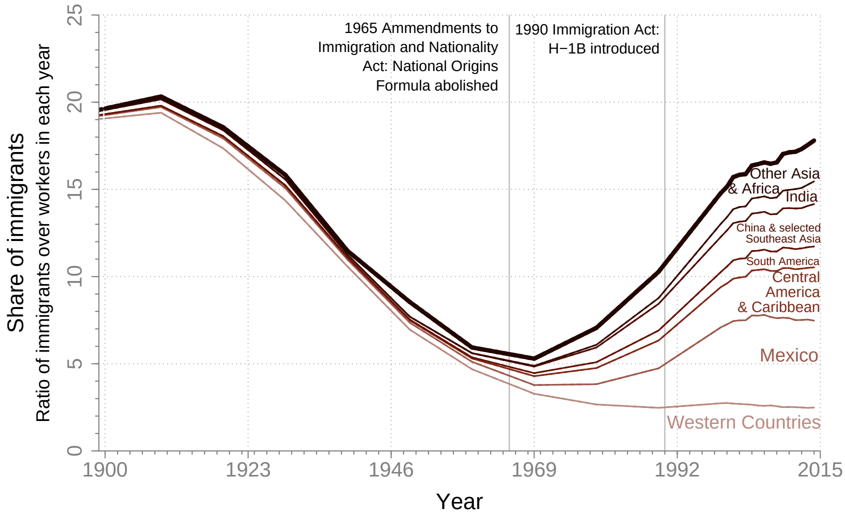
demonstrating that the two are highly associated. Finally, it provides evidence of the increasing incidence of immigration in STEM occupations.

Figure 1 shows the evolution of the share of immigrants in the population working-age by national origin over the last century. The figure shows that the predictions of the promoters of the 1965 Amendments to the Immigration and Naturalization Act were not very accurate. From the time the legislation was enacted until the present day, the stock of immigrants aged 16 to 64 in the U.S. increased by a factor of six, roughly from 6 to 37 million immigrants. This average inflow of about 650,000 immigrants of per year increased the share of immigrants in the population working-age from around 5.5% to 17.8%. Furthermore, the ethnic mix changed substantially. By mid 1960s, the majority of working-age immigrants were from Western countries (70.9%), but the inflows from these countries, which by then had already been decreasing for decades, continued decreasing until today. By 2015, Western immigrants only represent 14% of all immigrants working-age. On the contrary, coinciding with the approval of the 1965 bill, a steady inflow of Mexican and other Central American/Caribbean immigrants increased their presence from 8.2% and 6.3% to 28% and 17.1% respectively. More recently, the inflow of Asian immigrants has increased substantially, in particular, from China and selected Southeast Asia, and, especially, India.[11] In 1990, when H-1B visa program was introduced, Indian immigrants were only 3.1% of all working-age immigrants, while in 2015 they represent 7.3% of the total.

The national origin composition of immigration is closely associated to the skills of immigrants. Table 1 explores the extent to which this is the case. Panel A reports average years of education for natives, for immigrants as a whole, and for immigrants from each national origin. Panel B reports the fraction of individuals in each group that has a college degree. Both panels describe a similar picture. In 1970, natives and immigrants had similar education levels. However, since then, education of immigrants increased at a slower rate than that of natives. Interestingly, the change in the national origin composition of immigration that followed the removal of the National Origins Formula is crucially associated with this slower increase. As noted in Figure 1, a massive increase in immigration from Mexico and Central American and Caribbean countries followed the approval of the 1965 bill. These two groups of countries have, by far, the lowest education levels, which pushes the average education level of immigrants down. Once disag-

[11] The selected set of Southeast Asian countries grouped together with China includes Taiwan, Hong Kong, South Korea, Japan, the Philippines, and Singapore. Hereinafter, I refer to this group as selected Southeast Asia.

Figure 1. Immigrant Share by National Origin (1900-2015)



Note: Areas delimited by plotted lines show the share that immigrants from each national origin represent of all individuals of working-age. Country grouping and inter-census interpolations are described in Appendix A. *Sources:* Census data (1900-2000) and ACS (2001-2015).

gregated by national origin, the average education level of immigrants from each of the origin country groups evolved with a similar slope than natives. Moreover, besides Mexican and Central American immigration, other groups also have education levels that are similar to (or even higher than) natives. Table 1 also shows evidence of the influence of the introduction of H-1B visas on the education level of immigrants. Both at the aggregate level and for the majority of country groups individually, there is an important increase in education (relative to natives) during the 1990s, coinciding with the introduction of these visas. On aggregate, immigrants increased average education by one year, while natives only did by 0.7 years, despite the aforementioned increase in the importance of Mexico and Central American countries over the period. At the country group level, average education increased, over that decade, by 1.2, 1.5, 0.9, 0.9, 0.7, 0.4, and 1.1 years for Western countries, Mexico, Central America and Caribbean, South America, China and selected Southeast Asia, India, and other Asia and Africa respectively.[12] Indian immigrants are the only group that increased education in 1990s by less than natives, but that is a special case, because they already had extremely high education levels to begin with.

Figure 2 further explores the importance of national origin distribution in understanding skill composition of immigration. Compared to Table 1, it provides a sense of the distribution of education by national origin, over and above averages. The figure shows that U.S. immigration is bimodal. In particular, they are mostly

[12] Some groups, like Western countries went from a slightly lower education level than natives until 1990 to higher level than them (with an increasing gap) after that.

Table 1—Education of Natives and Immigrants

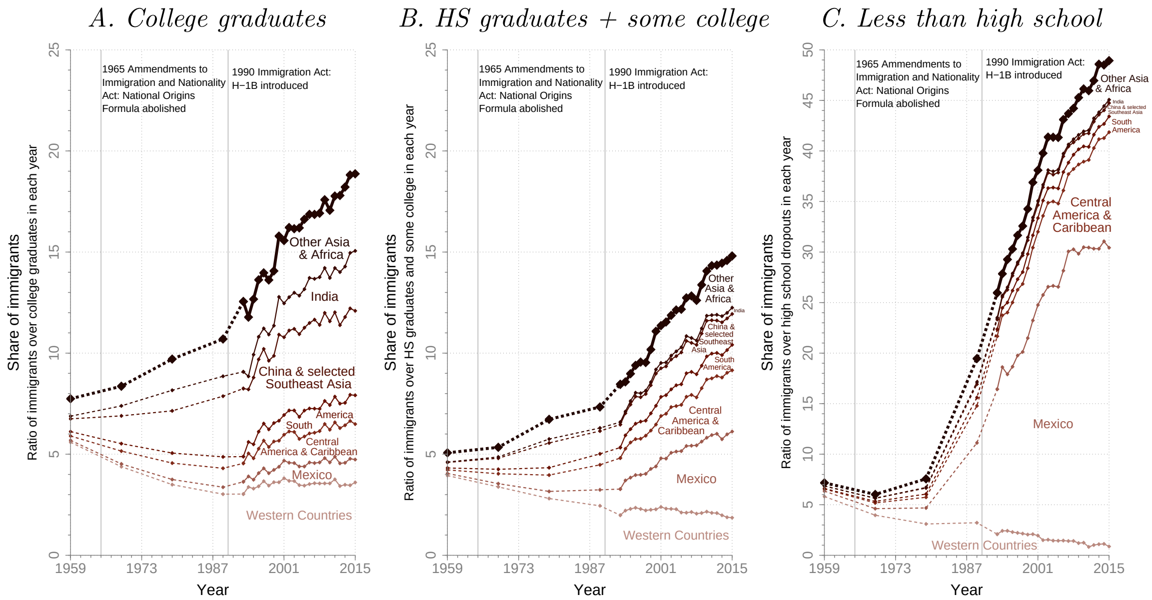
	1970	1980	1990	2000	2010	2015
A. Average years of education:						
Natives	11.2	12.3	13.0	13.7	13.7	13.8
Immigrants	11.0	11.7	12.1	13.1	13.1	13.3
Western countries	10.5	11.8	12.8	14.0	14.3	14.6
Mexico	6.3	7.1	7.6	9.1	9.3	9.6
Central America & Caribbean	10.2	10.9	11.0	11.9	11.5	11.5
South America	11.5	12.1	12.6	13.5	13.4	13.6
China & sel. Southeast Asia	11.7	13.2	13.7	14.4	14.6	14.6
India	15.6	15.2	15.1	15.5	15.4	15.4
Other Asia & Africa	11.2	11.8	12.2	13.3	13.2	13.5
B. Fraction with a college degree (%):						
Natives	23.2	35.3	50.8	57.7	56.1	58.2
Immigrants	23.6	35.5	44.3	47.1	45.3	47.7
Western countries	22.1	35.2	51.3	61.2	64.8	69.1
Mexico	6.6	10.1	13.8	16.1	15.4	17.0
Central America & Caribbean	22.0	29.7	35.0	37.8	34.2	35.0
South America	33.5	40.6	48.4	56.2	52.9	55.9
China & sel. Southeast Asia	42.9	57.3	65.9	71.1	71.9	72.1
India	80.7	78.7	77.4	79.1	80.5	79.3
Other Asia & Africa	32.5	41.4	51.4	59.5	58.1	60.5

Note: Figures in each panel indicate respectively the average years of education and the percentage of individuals with a college degree in each group. The sample is restricted to individuals aged 24-64. *Sources*: Census data (1970-2000) and ACS (2009-2011, and 2014-2015).

present among the highest educated (college graduates, 19% in 2015), and among the lowest educated (less than high school, 49%). It also shows that the relative importance of immigrants from each national origin in each of the education groups vary substantially. While Indian immigrants only represent 1.3% of the U.S. working-age population, they represent more than 3% of college graduates, and a negligible fraction of all individuals with less than high school. On the other extreme, while Mexican immigrants represent 5% of the U.S. population, and they represent 29% of all individuals with less than a high school diploma, they only represent slightly above 1% of all individuals with a college degree. Moreover, Panel A (and, importantly, not Panels B and C) also provides evidence of a change in the slopes after the introduction of H-1B visas in 1990.

Finally, Figure 3 provides a similar picture for occupations. While immigrants are relatively more resent in STEM and blue collar occupations, they are less present in white collar occupations. Mimicking the results for education, the vast majority of STEM occupations held by immigrants are executed by Asian nationals (more than 18.3% of all STEM employment, while they only represent less than 5% of the population), and very few of them are executed by Mexicans and

Figure 2. Immigrant Share by National Origin and Education (1960-2015)



Note: Areas delimited by plotted lines show the share that immigrants from each national origin represent of all individuals aged over 24 with the indicated education. Country grouping and inter-census interpolations are described in Appendix A. *Sources:* Census data (1960-1990), and CPS (1994-2016).

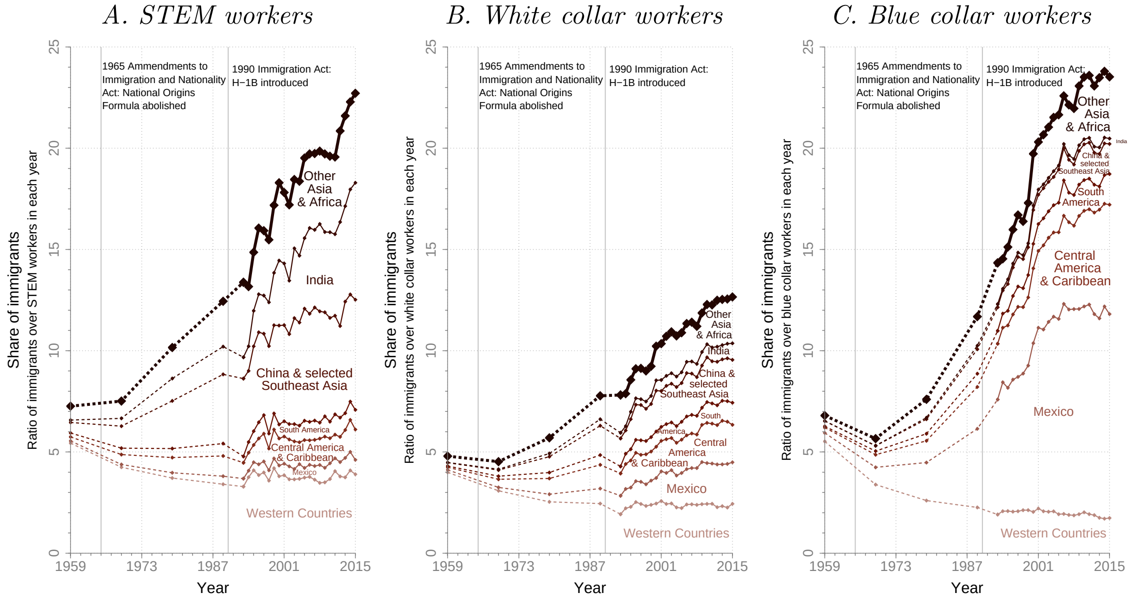
Central Americans/Caribbeans (less than 3%, even though they represent more than 8% of the population). In this figure, the change in slope after the introduction of H-1B visas is more prominent, and it is mostly driven by Indian workers. This claim is clearly supported by more direct evidence since, for example, of all H-1B visas issued in 2016, 70.4% were issued to Indian nationals.[13] Overall, Indian workers represented 9.6% of STEM employment in 1993 and 18.3% in 2016.

C. *Skilled Labor and the Generation of Ideas*

One of the central aspects of the model presented below is the presence of knowledge spillovers and externalities in the production of ideas from the use of STEM labor in production. To illustrate the extent of the association between STEM employment and the production of ideas, Figure 4 plots the spatial correlation between the (log of 1 plus) number of patents per 100,000 workers and different measurements of STEM intensity. To do so, I exploit metropolitan area variation in labor supply (obtained from the American Comunity Survey (ACS) for years 2000–2015) and in the number of patents (from the U.S. Patent and Trademark Office). In Panel A, STEM intensity is measured as the proportion of workers in the metropolitan area that are employed in STEM. The figure shows a very strong and positive correlation.

[13] U.S. Department of State, Bureau of Consular Affairs, `https://travel.state.gov/content/dam/visas/Statistics/Non-Immigrant-Statistics/NIVDetailTables/FY16%`

Figure 3. Immigrant Share by National Origin and Occupation (1960-2015)



Note: Areas delimited by plotted lines show the share that immigrants from each national origin represent of all individuals working in the indicated occupation. Country grouping and inter-census interpolations are described in Appendix A. *Sources:* Census data (1960-1990), and CPS (1994-2016).

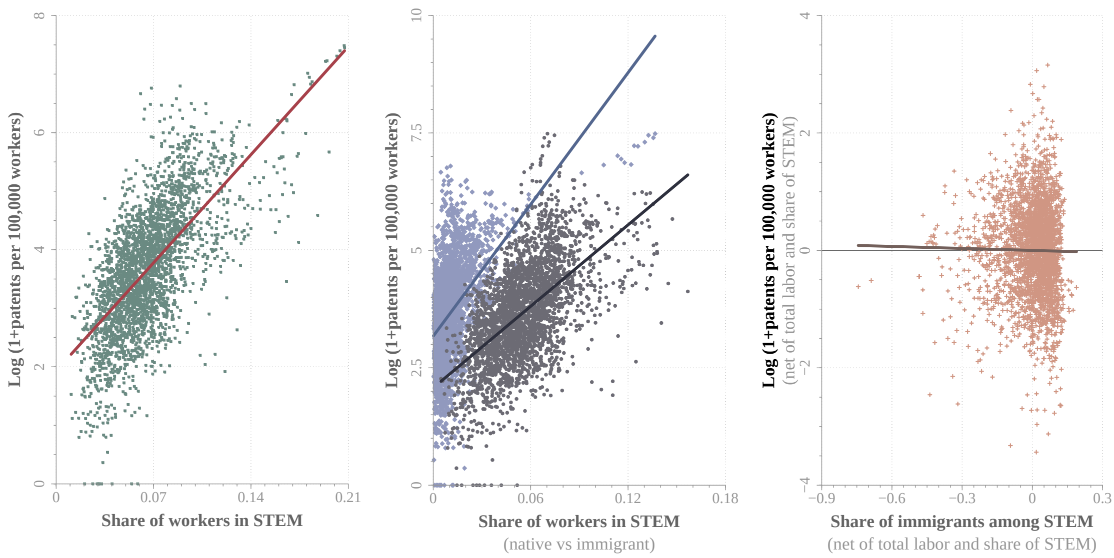
A relevant question for this paper is whether this correlation is driven by natives or immigrants. Panel B separates the STEM intensity in two parts: the one driven by natives, and the one driven by immigrants. In particular, the figure plots the share of all workers that are native (immigrant) STEM workers. The correlation stays strong and positive for both groups, even though regression lines seem to suggest a steeper pattern for natives. In order to corroborate or reject this different correlation, Panel C correlates the residuals of the regressions of the previous measure of patent productivity and the share of STEM workers that are immigrants on the total number of workers in the metropolitan area and the share of these workers that are employed in STEM occupations. The figure shows that, once labor market size and STEM intensity are controlled for, the correlation between immigrant intensity within STEM workers and the productivity in producing patents is essentially zero.

Overall, Figure 4 suggests a strong positive correlation between STEM intensity and productivity in patent production, and this correlation seems to be equally driven by native and immigrant STEM workers. As a final remark, it is important to note these correlations are only meant to show the link between STEM intensity and productivity in patent production, but the direction of causality could go in both directions. The structure of the model below provides a better framework

`20NIV%20Detail%20Table.xls`, accessed March 13th, 2017.

FIGURE 4. SPATIAL CORRELATION BETWEEN STEM LABOR AND PATENTS (2000-2015)

A. Share of STEM workers *B. Native vs immigrant STEM* *C. Share of immigrants in STEM*



Note: The three figures plot the spatial correlation (scatter and regression fit) between different measures of relative STEM labor and the log of (one plus) the number of patents per 100,000 workers. The left figure measures STEM labor as the fraction of all workers that work in STEM. The central figure plots the share of all workers that are native STEM (purple/diamonds) and the share of all workers that are immigrant STEM (gray/circles). The right figure plots (the residuals of) the share of all STEM workers that are migrants (from a regression that controls for the number of workers in the metro area and the share of all workers that are STEM). An individual is defined as a worker if she worked at least 40 weeks in the reference year, and usually worked at least 20 hours per week. She is defined as a STEM worker if she worked in a STEM occupation and has a college degree. Immigrants are defined as foreign born individuals. *Sources:* ACS (2000–2015) for employment and U.S. Patent and Trademark Office for patents.

to address these endogeneity concerns.

III. A Labor Market Equilibrium Model with Immigration and Knowledge Spillovers

In this section, I present a labor market equilibrium model with skilled and unskilled immigrants arriving from different countries of origin and competing with natives in three different occupations. The model, estimated with U.S. data, is then used to evaluate the selective immigration policies of interest, and to quantify the selectivity of immigration policy that maximizes native workers well-being. The modeling framework introduces heterogeneity across immigrants from different nationalities, accounts for human capital and labor supply adjustments by natives and previous generations of immigrants, and allows for economy-wide knowledge spillovers from skilled (STEM) workers.

A. *Representative firm*

A representative firm combines capital and labor to produce a single output. Knowledge spillovers are modeled as an externality through the production of ideas, measured as intellectual property products or the number of patents produced in the economy. Innovation is generated endogenously as a by-product of using STEM labor and capital equipment in general production. Replicating the behavior of a continuum of atomistic firms, the representative firm takes the stock of ideas as given, not internalizing the externality produced by equipment and STEM labor through knowledge spillovers.

Innovation. Let I_t denote the stock of ideas in period t, ΔI_t denote the net increase in the stock of ideas with respect to previous period (innovation), S_{Tt} denote the aggregate supply of STEM labor (measured in skill units), and K_{Et} denote equipment capital stock. Also let ξ_t denote an aggregate shock in the production of ideas. Innovation is generated when using equipment capital and STEM labor in general production, as described by the following technology:

$$\Delta I_t = \xi_t K_{Et}^{\chi_1} S_{Tt}^{\chi_2}. \tag{1}$$

The parameters χ_1 and χ_2 are not restricted to sum to one, thus allowing for increasing, constant, or decreasing returns to scale in the production of ideas. The exogenous innovation shock ξ_t is assumed to evolve according to:

$$\Delta \ln \xi_{t+1} = \pi_\xi + \sigma_\xi \upsilon_{\xi t+1}, \tag{2}$$

where $\upsilon_{\xi t}$ is a zero-mean innovation independently and identically distributed over time as a standard normal. The presence of such innovation shock is fundamental in identification, as discussed below, because it provides exogenous variation in innovation even when the inputs stay unchanged.

Production function. Let K_{St} denote capital structures, S_{Bt} and S_{Wt} denote aggregate supplies of blue collar and white collar labor (skill units), and Y_t denote aggregate output. Also let ζ_t denote an aggregate (factor-neutral) productivity shock. Output is produced according to the following nested CES technology:

$$Y_t = (\zeta_t I_r^{\varphi}) K_{St}^{\varsigma_t} \left\{ \alpha_t S_{Bt}^{\rho} + (1-\alpha_t) \left[\theta_t S_{Wt}^{\kappa} + (1-\theta_t) \left(\iota_t S_{Tt}^{\psi} + (1-\iota_t) K_{Et}^{\psi} \right)^{\frac{\kappa}{\psi}} \right]^{\frac{\rho}{\kappa}} \right\}^{\frac{1-\varsigma_t}{\rho}}, \tag{3}$$

where the demand shifters are allowed to evolve over time with the stock of ideas:

$$o_t \equiv \frac{\exp(\tilde{o}_0 + \tilde{o}_1 I_t)}{1 + \exp(\tilde{o}_0 + \tilde{o}_1 I_t)} \qquad \text{for } o \in \{\varsigma, \alpha, \theta, \iota\}. \tag{4}$$

This functional form ensures that the demand shifters lie between zero and one, implying constant returns to scale when I_t is taken as given. The exogenous factor neutral productivity shock ζ_t evolves according to the following process:

$$\Delta \ln \zeta_{t+1} = \pi_\zeta + \sigma_\zeta \upsilon_{\zeta t+1}, \tag{5}$$

where $\upsilon_{\zeta t}$ is a zero-mean innovation independently and identically distributed over time as a standard normal.

In this production function, innovation increases productivity and generates economic growth both in a factor neutral and in a skilled-biased way.[14] The term $\zeta_t I_r^\varphi$ can be interpreted as total factor productivity (TFP). To the extend to which $\varphi > 0$, the generation of ideas produces factor-neutral technological progress by enhancing TFP in the spirit of Romer (1986) and Lucas (1988). Furthermore, innovation also shifts the relative demands of inputs by changing ς_t, α_t, θ_t, and/or ι_t, thus inducing endogenous skill-biased technological change (as long as $\tilde{o}_1 \neq 0$ for some $o \in \{\varsigma, \alpha, \theta, \iota\}$). For example, the invention of computers may foster economic growth (TFP), increase the relative productivity of STEM labor, and/or substitute out blue collar labor. As in Krusell et al. (2000), this production function can also produce skilled-biased technical change through capital-skill complementarity as a result of an exogenous fall in the prices of equipment.[15]

It is also relevant to compare this technology to other production functions specified in the immigration literature. This production function extends the one used in Llull (2018a) to allow for a third input (STEM labor) and to incorporate knowledge spillovers in the way described above. Both production functions differ from the nested CES structure introduced in the immigration literature by Borjas (2003) and Ottaviano and Peri (2012). This is so because, unlike in these papers, the explicit modeling of labor supply can account for occupational decisions of immigrants, which allows to reduce the number of types of imperfect substitutability that I need to model to rationalize the data. Specifically, Ottaviano and Peri (2012) discuss the importance of imperfect substitutability between natives and immigrants with the same observable skills "because they tend to work in different occupations". The equilibrium structure endogenously generates this imperfect substitutability (even though immigrants within a given occupation are perfect substitutes) through endogenous sorting into occupations. In fact, using data simulated from his model, Llull (2018a) finds a "reduced form" elasticity of

[14] The introduction of intangible capital in the aggregate production function has already been discussed in the macroeconomics literature (e.g. McGrattan and Prescott, 2010, 2014). These papers, however, model it as a (rival) intermediate input in which the firm invests, not as a non-rival stock of ideas generated unintendedly in general production.

[15] Lewis (2011, 2013) discusses the importance of capital-skill complementarity in measuring labor market effects of immigration.

substitution between natives and immigrants that fits well within the ballpark of estimates in Ottaviano and Peri (2012).

Profit maximization. The representative firm maximizes profits in a static way. Given the single output production, I normalize output prices to one. Let r_{jt} for $j \in \{T, W, B\}$ denote the market prices of STEM, white collar, and blue collar labor skill units. Let r_{St} and r_{Et} denote the rates of return to structures and equipment capital respectively. The firm's problem is defined by:

$$\max_{S_{Tt}, S_{Wt}, S_{B_t}, K_{Et}, K_{St}} \left\{ \begin{array}{r} Y(\zeta_t, I_t, S_{Tt}, S_{Wt}, S_{B_t}, K_{Et}, K_{St}) - r_{Tt}S_{Tt} \\ -r_{Wt}S_{Wt} - r_{Bt}S_{Bt} - r_{Et}K_{Et} - r_{St}K_{St} \end{array} \right\}. \tag{6}$$

As a representation of atomistic firms, the representative firm takes the stock of ideas in the economy I_t as given. Define the following Armington aggregators: $Q_{1t} \equiv (\iota_t S_{Tt}^{\psi} + (1-\iota_t)K_{Et}^{\psi})^{1/\psi}$, $Q_{2t} \equiv (\theta_t S_{Wt}^{\kappa} + (1-\theta_t)Q_{1t}^{\kappa})^{1/\kappa}$, and, finally, $Q_{3t} \equiv (\alpha_t S_{Bt}^{\rho} + (1-\alpha_t)Q_{2t}^{\rho})^{1/\rho}$. The aggregate demand of STEM skill units is:

$$r_{Tt} = (1-\varsigma_t)(1-\alpha_t)(1-\theta_t)\iota_t \left(\frac{Q_{2t}}{Q_{3t}}\right)^{\rho} \left(\frac{Q_{1_t}}{Q_{2t}}\right)^{\kappa} \left(\frac{S_{Tt}}{Q_{1t}}\right)^{\psi} \frac{Y_t}{S_{Tt}}. \tag{7}$$

The demand of white collar skill units is given by:

$$r_{Wt} = (1-\varsigma_t)(1-\alpha_t)\theta_t \left(\frac{Q_{2t}}{Q_{3t}}\right)^{\rho} \left(\frac{S_{Wt}}{Q_{2t}}\right)^{\kappa} \frac{Y_t}{S_{Wt}}. \tag{8}$$

The demand of of blue collar skill units is given by:

$$r_{Bt} = (1-\varsigma_t)\alpha_t \left(\frac{S_{Bt}}{Q_{3t}}\right)^{\rho} \frac{Y_t}{S_{Bt}}. \tag{9}$$

Finally, the demands for capital structures and equipment are:

$$r_{St} = \varsigma_t \frac{Y_t}{K_{St}}, \tag{10}$$

and:

$$r_{Et} = (1-\varsigma_t)(1-\alpha_t)(1-\theta_t)(1-\iota_t) \left(\frac{Q_{2t}}{Q_{3t}}\right)^{\rho} \left(\frac{Q_{1_t}}{Q_{2t}}\right)^{\kappa} \left(\frac{K_{Et}}{Q_{1t}}\right)^{\psi} \frac{Y_t}{K_{Et}}. \tag{11}$$

It is important to note that, despite the externalities, the representative firm makes zero profits. This is so because, from the point of view of the atomistic firms, the above production function is constant returns to scale, even if $\varphi > 0$ or $\tilde{o}_1 \neq 0$ for some $o \in \{\varsigma, \alpha, \theta, \iota\}$. As such, the factor shares sum to one, as it can be trivially shown from (7) through (11).

B. *Workers*

Workers make life-cycle decisions on education, occupation and participation. Consistent with standard models of human capital (Ben-Porath, 1967) they concentrate education at the beginning of their careers, and then keep accumulating human capital in the form of experience throughout their working life. Given the dynamics of the model, individuals face the trade-off between wages/utility for today and investment for the future. Through equilibrium and spillovers, immigration affects this trade-off by changing relative wages.

Life cycle. Let $a \in \{16, ..., 65\}$ denote age. Native individuals are born with $a = 16$ and a given initial human capital endowment, to be specified below. They make yearly decisions until $a = 65$, when they die with certainty. Immigrants only start making decisions upon entry in the United States, which occurs at a given (individual-specific) age of entry $\tilde{a}$, and there is no return migration. They arrive to the U.S. with a given initial human capital, which is also specified below, and make yearly endogenous decisions over ages $a \in \{\tilde{a}, ..., 65\}$.

Choice sets. Every year, individuals decide one of three to five mutually exclusive alternatives: working in blue collar, white collar, or STEM, attending school, and staying home. The STEM occupation has a college degree requirement, and therefore is not available to individuals with less than 16 years of education. Likewise, consistently with the very low rates of school reentry observed in the data, I assume that leaving school is an absorbing state. Therefore, the choice set depends on previous choice and education level.

Let $\mathcal{D}_{11}$ denote the choice set for individuals with at least 16 years of education (college degree or more) who were in school in the previous period; let $\mathcal{D}_{10}$ denote the choice set for college educated individuals whose previous choice was not school; and let $\mathcal{D}_{01}$ and $\mathcal{D}_{00}$ denote the choice sets for individuals with less than 16 years of education who were and were not in school in the previous period. The first set is formed off five alternatives: blue collar, white collar, STEM, school, and home, that is $\mathcal{D}_{11} \equiv \{B, W, T, S, H\}$. Furthermore, given the college degree requirement for STEM occupations, $\mathcal{D}_{0i} = \mathcal{D}_{1i} \setminus \{T\}$ for $i \in \{0, 1\}$. Finally, the absorbing nature of dropping out from school implies $\mathcal{D}_{i0} = \mathcal{D}_{i1} \setminus \{S\}$ for $i \in \{0, 1\}$.

More compactly, denote the choice set as $\mathcal{D}(h_a)$, where h_a is the vector of individual-specific state variables defined below, which includes education and previous period decision among other variables. The choice of a given individual at age a is denoted by $d_a \in \mathcal{D}(h_a)$, and a set of indicator variables is defined, such that $d_{ja} \equiv \mathbb{1}\{d_a = j\}$ for any $j \in \mathcal{D}(h_a)$, with $\sum_{j \in \mathcal{D}(h_a)} d_{ja} = 1$.[16]

Observable idiosyncratic state variables. Let $h_a \equiv (a, \ell, E_a, d_{a-1}, n_a, \tilde{a})'$ denote the vector of observable individual-specific state variables. This vector is defined by age a, demographic type ℓ, education E_a, lagged choice d_{a-1}, number of preschool children at home $n_a \in \mathcal{C} \equiv \{0, 1, 2+\}$, and, in the case of immigrants, age at entry $\tilde{a}$. I partition the workforce into a finite number of types, subscripted by ℓ, defined by observable characteristics. Natives are classified into six groups, defined by race (Hispanic, non-Hispanic black, non-Hispanic non-black) and gender (male and female). Immigrants are divided into fourteen groups, defined by national origin (Western countries, Mexico, Central America & Caribbean, South America, China & selected Southeast Asia, India, and other Asia and Africa) and gender. These groups identify twenty types of individuals, $\mathcal{L} \equiv (\mathcal{R} \times \mathcal{G}) \cup (\mathcal{O} \times \mathcal{G})$, where $\mathcal{R}$ denotes the set of races for natives, $\mathcal{O}$ denotes the set of national origins of immigrants, and $\mathcal{G} \equiv \{1, 2\}$ denotes the set of genders (male and female are denoted by 1 and 2 respectively). I also define four sets of types, $\tilde{\mathcal{L}}_1 \equiv \{\ell : \ell \in (\mathcal{R} \times \{1\})\}$, $\tilde{\mathcal{L}}_2 \equiv \{\ell : \ell \in (\mathcal{R} \times \{2\})\}$, $\tilde{\mathcal{L}}_3 \equiv \{\ell : \ell \in (\mathcal{O} \times \{1\})\}$, and $\tilde{\mathcal{L}}_4 \equiv \{\ell : \ell \in (\mathcal{O} \times \{2\})\}$, to denote native male, native female, immigrant male, and immigrant female respectively, and the operator $\tilde{\ell}(\ell)$ to index them. All this (observable) heterogeneity is necessary for several reasons. The distinction between the six national origins for immigrants is necessary to capture the different skill composition of immigrant inflows, as described in Section II.B, their different labor supply and occupation propensities, and ultimately the different probability of holding H-1B visas. Furthermore, permanent unobserved heterogeneity is not identifiable in this model. As noted by Llull (2018a), to correctly identify the distribution of permanent unobserved heterogeneity one would need data for the same individuals before and after migration to the U.S. (to my knowledge, unavailable), and the individual migration decisions should be modeled explicitly, which would be intractable. The heterogeneity in the number of children provide an interesting exclusion restriction that is useful to identify the model, as discussed below. Finally, the presence of age at entry in immigrants' state vector also allows for assimilation, as defined in LaLonde and Topel (1992): among two individuals with the same observable skills, the one who spend more time in the U.S. earns more. Capturing this feature, documented in the literature, is important to correctly quantify the size of the labor supply shock induced by immigration under different scenarios.

The initial state vector for natives is given by $h_{16} = (16, \ell, E_{16}, d_{15}, 0, \cdot)'$, where ℓ

[16] The indicator function $\mathbb{1}\{\cdot\}$ is defined to be one if the argument is satisfied, zero otherwise.

and E_{16} are exogenously determined at birth, and $d_{15} = S$ if $E_{16} = 12$ and $d_{15} = H$ otherwise. Immigrants enter into the United States with a given state vector $h_{\tilde{a}} = (\tilde{a}, \ell, E_{\tilde{a}}, d_{\tilde{a}-1}, 0, \tilde{a})'$, where $d_{\tilde{a}-1} = S$ if $E_{\tilde{a}} \geq (\tilde{a} - 6) - 2$, and $d_{\tilde{a}-1} \equiv F$ otherwise.[17] The distribution of initial state variables of natives and immigrants is specified outside of the model. The state vector is updated as follows: a is increased in one unit, ℓ is constant, $E_{a+1} = E_a + d_{Sa}$, the previous period choice d_{a-1} is replaced by the current choice d_a, $\tilde{a}$ is constant, and the number of children is increased stochastically with cumulative distribution function $P_n(n|h_a, d_a)$ with $P_n(n|h_a, j) = P_n(n|h_a, k)$ for any $j, k \neq S$.[18] The set of possible values for the observable idiosyncratic state variables at age $a+l$ when the state variable was h_a at age a is denoted by $\mathcal{H}_{a+l|h_a}$, and the unconditional set of possible values is $\mathcal{H}$.

Idiosyncratic shocks. This model includes two types of idiosyncratic shocks, which are independent and identically distributed across individuals and over time. Let η_a denote a shock to individual productivity, with $\eta_a|h_a \sim \mathcal{N}(0, 1)$. Also let ϵ_{ja} denote a taste shock associated to alternative $j \in \mathcal{D}(h_a)$. I define the vector of combined idiosyncratic shocks, denoted by ε_a, as:

$$\varepsilon_a \equiv (\sigma_{B\ell}\eta_a + \epsilon_{Ba}, \sigma_{W\ell}\eta_a + \epsilon_{Wa}, \sigma_{T\ell}\eta_a + \epsilon_{Ta}, \epsilon_{Sa}, \epsilon_{Ha})', \tag{12}$$

where $\sigma_{j\ell}$ is defined below. The distribution of ϵ_a is such that the combined idiosyncratic shock is generalized extreme value distributed, $\varepsilon_a|h_a \sim F_\varepsilon(\varepsilon_a)$ with:

$$F_\varepsilon(\varepsilon_a) \equiv \exp\left\{-\left[\left(e^{-\varepsilon_{Ba}/\varrho} + e^{-\varepsilon_{Wa}/\varrho} + e^{-\varepsilon_{Ta}/\varrho}\right)^\varrho + e^{-\varepsilon_{Sa}} + e^{-\varepsilon_{Ha}}\right]\right\}, \tag{13}$$

where $\varrho \equiv \sqrt{1 - \mathrm{Corr}(\varepsilon_{ja}, \varepsilon_{ka})}$ for any $j, k \in \{B, W, T\}$.

Wages, skill units, and aggregate state variables. Individual wages in this model are defined as the product of the amount of skill units supplied by the individual in occupation j, and the market price of these skill units. Let r_t denote the vector of skill prices at time t, defined as $r_t \equiv (r_{Bt}, r_{Wt}, r_{Tt})'$. Let $s_j(h_a, \eta_a)$ denote the amount of skill units supplied by an individual with state variables h_a and η_a. The occupation-j wage of this individual is defined as:

$$w_j(h_a, \eta_a, r_t) \equiv r_{jt} s_j(h_a, \eta_a). \tag{14}$$

[17] This assumption implies that immigrants with up to two years of work experience abroad can still enroll in school when they arrive in the United States. Given data availability, I assume that $F = H$, so that the cost of reentry to work is the same if the individual was in the U.S. but not working or she was abroad (whether working or not).

[18] In order to capture the demographic transition, and the subsequent increase in female labor force participation, I assume that the transition function before 1970 was $\tilde{P}_n(n|h_a, d_a)$ instead of $P_n(n|h_a, d_a)$. For tractability, I assume that the affected cohorts do not take into account the change in the fertility process when, before the change, form expectations about the future.

The specification of the wage as the product of skill units and their market price is very explicit about how immigration affects natives. On impact, relative skill prices are changed in equilibrium due to the change in relative supplies induced by immigration. Then natives adjust to this change in incentives by changing their behavior and, as a result, their skill units. Individuals have different mechanisms to adjust to (skilled and unskilled) immigration: they can adjust their education, change their occupation, and decide to stay home. As for skilled immigration, competition effects unambiguously incentivize natives to work in less skilled occupations and discourage investments in education. However, knowledge spillovers can potentially mitigate or even offset these competition effects.

As key determinants of individual choices, aggregate skill prices r_t are included as state variables. Given the presence of aggregate shocks (and uncertain future migration, as discussed below), individuals cannot perfectly predict future skill prices $\{r_{t+l}\}_{l\in\{1,\dots,65-a\}}$. Let ϖ_t denote all information available to individuals at time t to forecast them. The state vector for the worker decision problem is thus expanded to also include ϖ_t.

I assume the idiosyncratic productivity shock is log-additively separable. In particular, the skill unit production function $s_j(h_a, \eta_a)$ is defined as:

$$s_j(h_a, \eta_a) \equiv \exp(\tilde{s}_j(h_a) + \sigma_{j\ell}\eta_a), \tag{15}$$

where $\tilde{s}_j(\cdot)$ is a function of the observable idiosyncratic state vector h_a, and $\sigma^2_{j\ell}$ is the occupation-j-type-ℓ-specific conditional variance. I assume that the fifth element of h_a (number of children) does not enter $\tilde{s}_j(\cdot)$.

The model allows for different possibilities of adjustment at different points of the life cycle. Young individuals may be more likely to change occupations and, if still in school, to adjust their education decision, while older individuals, who devoted all their career to a given occupation, may exhibit lower occupational mobility. The skill units production function, specified in Equation (15) allows for different transition costs at different ages through the interaction of previous period choice and age, which allows for such behavior. Furthermore, the error structure (along with these costs) gives freedom to the model to fit at the same time the persistence in choices and wages in the data and the observed variance in wages. Correctly reproducing the transition probability across alternatives is crucial to credibly identify labor supply and human capital adjustments to immigration. Given that the model is estimated with the CPS (which offers a one-year panel dimension), I abstract from introducing accumulation of occupation-specific work experience. This approach is in the spirit of Heckman et al. (1998), but differs from Keane and Wolpin (1994, 1997), Lee (2005), Lee and Wolpin (2006, 2010), or Llull (2018a).

Alternative-specific period utility functions. I assume that individuals are not allowed to save or borrow. Individuals are thus assumed to consume all earned income when they work (wages), and I assume the utility functions for non-working alternatives (school and home) include a given amount of consumption embedded in parameter values. The utility of consumption is assumed to be logarithmic, and other non-pecuniary elements enter additively. The utility of working in occupation j is given by the sum of log-consumption (log-wage), a type-specific amenity value of working in occupation j, denoted by Λ_{0j}, an occupation-specific re-entry cost if the individual was at home in the previous period Λ_{1j}, and the occupation-specific taste-shock ϵ_{ja}:

$$\begin{aligned} u_j(h_a, \varepsilon_a, r_t) &\equiv \ln w_j(h_a, \eta_a, r_t) + \Lambda_{0j}(\ell) - \Lambda_{1j} d_{5a-1} + \epsilon_{ja} \\ &= \ln r_{jt} + \tilde{s}_j(h_a) + \Lambda_{0j}(\ell) - \Lambda_{1j} d_{5a-1} + \varepsilon_{ja}, \quad j \in \{B, W, T\}, \end{aligned} \tag{16}$$

The net utility of attending school is assumed to depend on individual type ℓ, educational level E_a, and the idiosyncratic taste shock ε_{Sa}:

$$u_S(h_a, \varepsilon_a, r_t) \equiv \tau(\ell, E_a) + \varepsilon_{Sa}. \tag{17}$$

Finally, the utility of staying home depends on individual type, the number of children n_a, and the taste shock ε_{Ha}:

$$u_H(h_a, \varepsilon_a, r_t) \equiv \vartheta(\ell, n_a) + \varepsilon_{Ha}. \tag{18}$$

Intertemporal decisions. Let β denote the subjective discount factor. Also let $\tilde{u}_j(\cdot)$ denote the deterministic part of the period utility function, defined as $\tilde{u}_j(h_a, r_t) \equiv u_j(h_a, \varepsilon_a, r_t) - \varepsilon_{ja}$.[19] An individual with a state vector h_a and idiosyncratic shock ε_a, observed at time t (that is, when equilibrium prices are r_t and the information available to predict their future values is ϖ_t), chooses $\{d_{ja}\}_{j\in\mathcal{D}(h_a)}$ to sequentially maximize the expected discounted sum of payoffs:

$$\mathbb{E}_t\left[\sum_{l=0}^{65-a} \beta^l \left(\sum_{j\in\mathcal{D}(h_{a+l})} d_{ja+l}[\tilde{u}_j(h_{a+l}, r_{t+l}) + \varepsilon_{ja+l}]\right)\right], \tag{19}$$

where $\mathbb{E}_t[\,\cdot\,] \equiv \mathbb{E}[\,\cdot\,|h_a, \varepsilon_a, r_t, \varpi_t]$ denotes the conditional expectation of the argument given the information available to the individual at time t, including h_a,

[19] Given that ε_{ja} enters $u_j(\cdot)$ additively, and that ε_{ka} for any $k \neq j$ does not enter $u_j(\cdot)$, the above expression does not depend on ε_a.

ε_a, r_t, and ϖ_t. Appealing to the Bellman's principle (Bellman, 1957), I express worker's decision problem recursively as:

$$V(h_a, \varepsilon_a, r_t, \varpi_t) = \max_{\{d_{ja}\}_{j\in\mathcal{D}(h_a)}} \sum_{j\in\mathcal{D}(h_a)} d_{ja} \left\{ \tilde{u}_j(h_a, r_t) + \varepsilon_{ja} + \beta\, \mathbb{E}_t[V(h_{a+1}, \varepsilon_{a+1}, r_{t+1}, \varpi_{t+1})] \right\}, \tag{20}$$

where the terminal value is defined to be zero, $V(h_{65}, \varepsilon_{65}, r_t, \varpi_t) \equiv 0$. Following Arcidiacono and Miller (2011), I define two additional objects derived from the above expression. First, define the ex-ante value function (the continuation value of being in state (h_a, r_t, ϖ_t), just before ε_a is revealed) as:

$$\bar{V}(h_a, r_t, \varpi_t) \equiv \int V(h_a, \varepsilon, r_t, \varpi_t) dF_\varepsilon(\varepsilon). \tag{21}$$

Second, define the alternative-specific conditional value function as:

$$v_j(h_a, r_t, \varpi_t) \equiv \tilde{u}_j(h_a, r_t) + \beta \int \sum_{h\in\mathcal{H}_{a+1|h_a}} \bar{V}(h, r, \varpi) P_h(h|h_a, j) dF_r(r, \varpi|\varpi_t, r_t), \tag{22}$$

where the function $F_r(r, \varpi|\varpi_t, r_t)$ is the distribution of aggregate conditions in period $t+1$ given information available at time t, and $P_h(h|h_a, j)$ is the transition probability mass function for h_a discussed above, which is degenerate for all elements of h except for n. Thus, optimal choices, denoted by $d^*_{ja}(h_a, \varepsilon_a, r_t, \varpi_t)$ for $j \in \mathcal{D}(h_a)$, are given by:

$$\{d^*_{ja}(h_a, \varepsilon_a, r_t, \varpi_t)\}_{j\in\mathcal{D}(h_a)} = \arg\max_{\{d_{ja}\}_{j\in\mathcal{D}(h_a)}} \sum_{j\in\mathcal{D}(h_a)} d_{ja}[v_j(h_a, r_t, \varpi_t) + \varepsilon_{ja}]. \tag{23}$$

Expectations Rational individuals use $F_\varepsilon(\varepsilon_a)$, $P_h(h_{a+1}|h_a, d_a)$, and $F_r(r_{t+1}|r_t, \varpi_t)$ to form expectations about future state variables. All these functions are specified in the model except for $F_r(r_{t+1}|r_t, \varpi_t)$. Given the presence of aggregate and idiosyncratic shocks, rational expectations imply that ϖ_t includes the entire distribution of state variables in period t, which is intractable.[20] This is a well known problem in macroecnomics and applied microeconometrics that has been addressed by approximating rational expectations by simpler forecasting rules based on equilibrium outcomes (Krusell and Smith, 1998; Altuğ and Miller, 1998; Lee and Wolpin, 2006, 2010; Llull, 2018a). In the immigration context, Llull (2018a) finds that an autoregressive process in first differences and the contemporaneous change in the aggregate shock can explain 99.9% of the variation in

[20] Despite this complication, aggregate shocks are very necessary in the immigration context to avoid making the unrealistic assumption that workers can perfectly predict the future evolution of economic conditions including the future inflow of migrants.

the level of skill prices. Further results presented in Llull (2018b) show that the innovation in the aggregate shock alone can still explain 99.9% of the variation in levels (along with current skill price), and 71–76% of the first differences. Given this, and following a similar approach as in Altuğ and Miller (1998), I assume skill prices are forecasted using the following rule:

$$\ln r_{jt+1} = \ln r_{jt} + \Xi_{0j} + \Xi_{1j}\sigma_\zeta \upsilon_{\zeta t+1} + \Xi_{2j}\sigma_\xi \upsilon_{\xi t+1} + \sigma_{\Upsilon j}\Upsilon_{jt+1}, \quad j \in \{B, W, T\}, \tag{24}$$

where $\Upsilon_{jt+1} \sim \mathcal{N}(0,1)$ is an independent and identically distributed approximation error uncorrelated with Υ_{kt} for any $k \neq j$, $\sigma_\xi \upsilon_{\xi t+1}$ and $\sigma_\zeta \upsilon_{\zeta t+1}$ are defined in (2) and (5) respectively, and Ξ_{0j}, Ξ_{1j}, Ξ_{2j}, and $\sigma_{\Upsilon j}$ are not parameters, but, instead, implicit functions of the fundamentals derived as part of the solution of the model. Equation (24) implicitly assumes that r_t is a sufficient statistic of all the information that individuals have at time t to predict r_{t+1} (as $\upsilon_{\zeta t+1}$ and $\upsilon_{\xi t+1}$ are unknown at t and i.i.d. over time). This implies ϖ_t is a redundant state variable and, thus, I omit it hereinafter.

C. Capitalists

The comparison of results in Borjas (2003), Ottaviano and Peri (2012), and Llull (2018a) suggests that, even though the model described so far is informative on distributional effects of immigration, the overall labor market effects depend on our assumption of how capital reacts to immigration. These papers take one of the two most extreme assumptions (or both): Borjas (2003) assumes that capital does not adjust to immigration; in Ottaviano and Peri (2012), interest rates do not react (long run small open economy); Llull (2018a) provides results with both assumptions noting that reality should probably be somewhere in between.

In this paper, I opt for closing the economy with the simplest possible specification of capital supply, so that counterfactuals can provide a more credible measurement of the overall effects of selective immigration policies.[21] In particular, I assume that capital is supplied by a continuum of infinitely lived capitalists that only make consumption and savings decisions and live out of the return to their assets. These capitalists are homogeneous, and therefore are characterized by a representative consumer model. Unlike workers, I assume they have perfect foresight about future aggregate shocks and, therefore, future interest rates.

[21] Given that I take (equilibrium) capital from the data, and that I do not impose any orthogonality condition to aggregate variables with respect to aggregate shocks, this part of the model is irrelevant for a consistent estimation of the rest of the model.

Let C_t denote aggregate consumption of capitalists in year t, and let A_t denote their asset position (decided in period $t-1$). Let r_{At} denote the interest rate paid to assets, which is connected to r_{St} and r_{Et} in the way described below. The problem of the representative capitalist is given by:

$$\max_{\{C_{t+\tau},A_{t+1+\tau}\}_{\tau=0}^{\infty}} \sum_{\tau=0}^{\infty} \beta^{\tau} \ln C_{t+\tau} \tag{25}$$

subject to:

$$C_{t+\tau} + A_{t+1+\tau} \leq (1 + r_{At+\tau})A_{t+\tau}, \tag{26}$$

$$A_{t+1+\tau} \geq 0. \tag{27}$$

The solution of this problem is obtained from the following Euler equation:

$$\frac{\beta(1 + r_{At+1})}{(1 + r_{At+1})A_{t+1} - A_{t+2}} = \frac{1}{(1 + r_{At})A_t - A_{t+1}}. \tag{28}$$

Define $\Delta A_t \equiv \frac{A_{t+1}}{(1+r_{At})A_t}$. Equation (28) can be rewritten, upon rearrangement, as the following differential equation:

$$\Delta A_t = \frac{\beta}{1 - \Delta A_{t+1} + \beta}. \tag{29}$$

Assuming that the transversality condition holds, the solution of the forward recursion of this expression yields:

$$\Delta A_t = \lim_{T=\infty} \beta \frac{1 - \beta^{T-t}}{1 - \beta^{T-t+1}} = \beta, \tag{30}$$

which implies:

$$A_{t+1} = \beta(1 + r_{At})A_t. \tag{31}$$

There is not a perfect mapping between assets and capital in this model: firms can use a unit of assets to buy q_{St} units of capital structures, or q_{Et} units of equipment.[22] These prices are exogenous in the model. I normalize $q_{St} \equiv 1$.

In the model there is a zero-profit intermediary that transforms assets into capital at the beginning of the period, and capital into assets at the end. The budget constraint of this intermediary is:

$$A_t \geq K_{St} + q_{Et}K_{Et}, \tag{32}$$

[22] Krusell et al. (2000) link the fall in prices of equipment capital and the increase in the college-high school wage gap. It is important to keep this ingredient in the model so that it has room for exogenous skill-biased technical change not driven by the accumulation of ideas.

and the revenue function is:

$$A_t(1 + r_{At}) = (1 - \delta_s + r_{St})K_{St} + q_{Et}(1 - \delta_s + r_{Et})K_{Et}, \tag{33}$$

where δ_S and δ_E are, respectively, the depreciation rates of structures and equipment. Given the linear objective function, there are infinite many interior solutions if $r_{St} - \delta_S = r_{Et} - \delta_E$, and unique corner solutions otherwise.

D. Equilibrium

The market structure in this model is as follows. Immigrant inflows are specified outside of the model. Interest rates are such that capital markets clear. Returns to skill units are such that supply equals labor demand.

The equilibrium capital (as a function of aggregate skill units) is determined as follows. An interior solution of the intermediary's problem (which is the only candidate for equilibrium, given the firm's problem described above) requires $r_{St} - \delta_S = r_{Et} - \delta_E$. Recursively substituting (33), (10) and (11) into (31) and imposing this condition yields a system of equations that determines the sequence of equilibrium levels of capital.

The labor market prices of skill units are also determined by market clearing conditions. Aggregate supply of skills in occupation $j \in \{B, W, T\}$, denoted by $S_{jt}^{(S)}(r_t)$, is given by the aggregation over all skill units supplied by individuals working in occupation j:

$$S_{jt}^{(S)}(r_t) = \iint \sum_{h \in \mathcal{H}} d_{jt}^*(h, \epsilon + \eta, r_t, \varpi_t) s_j(h, \eta) P_h(h) dF_\epsilon(\epsilon) d\Phi(\eta), \tag{34}$$

where $P_h(\cdot)$ is the probability mass function for each point of the (idiosyncratic) state space, $\Phi(\cdot)$ denotes the standard normal cumulative distribution function, and $F_\epsilon(\cdot)$ is the cumulative distribution function of the taste shock ϵ_a.[23] Aggregate labor demands, denoted by $S_{jt}^{(D)}(r_t)$ for $j \in \{B, W, T\}$, are given by the solution of the system of equations defined by (7) through (9), replacing capital by the equilibrium conditions described above.

Even though migration decisions are not modeled, immigration inflows (both size and composition) and capital supplies are endogenously determined by processes specified outside of the model. This is so because, as noted below, no orthogonality condition is assumed between the aggregate productivity shocks, ζ_t and ξ_t, and the aggregate variables (migration process, cohort sizes,...). Thus,

[23] Equation (34) assumes that there is a measure 1 of workers. In the empirical application, this measure is scaled by population size.

immigrant inflows are allowed to react to changes in the economic conditions in the U.S. and other aggregate factors, like endogenous immigration policies.[24]

IV. Identification

The subjective discount factor β is assumed to be equal to 0.95. The capital depreciation rates, δ_I, δ_E, and δ_S are assumed to be 20.88%, 11.93%, and 2.88% respectively (see Appendix A for details). The transition function for preschool children $P_n(\cdot)$ is directly identified from observed transitions in the data. And the distribution of idiosyncratic and aggregate shocks are specified above. Thus, in line with the literature, I assume that these objects, often summarized with the notation (β, F, G), are known (Rust, 1994; Magnac and Thesmar, 2002; Arcidiacono and Miller, 2011, 2015). The remaining parameters (and functions) to be identified are: the skill unit production function $\{\tilde{s}_j(h_a)\}_{j\in\{B,W,T\}}$, the variance parameters for wages $\{\sigma_{j\ell}\}_{j\in\{B,W,T\}}^{\ell\in\mathcal{L}}$, the re-entry costs for the three working alternatives $\{\Lambda_{kj}(\ell)\}_{j\in\{B,W,T\}}^{\{k\in\{0,1\}\}}$, the parameter associated with the correlation across idiosyncratic shocks ϱ, the deterministic part of the schooling utility function $\tau(\ell, E_a)$, the deterministic component of the home utility $\vartheta(\ell, n_a)$, the equipment capital and STEM parameters in the generation of IPP capital technology χ_1 and χ_2, the parameters of the production function φ, $\tilde{\varsigma}_0$, $\tilde{\varsigma}_1$, $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, $\tilde{\theta}_0$, $\tilde{\theta}_1$, $\tilde{\iota}_0$, $\tilde{\iota}_1$, ρ, κ, and ψ, the parameters from the aggregate shock processes π_ξ, π_ζ, σ_ξ, and σ_ζ, and the reduced form parameters of the skill-price expectation function $\{\Xi_{0j}, \Xi_{1j}, \Xi_{2j}, \sigma_{\Upsilon j}\}_{j\in\{B,W,T\}}$ (which are not fundamentals of the model but part of the solution, as noted above). The discussion on their identification builds on Lee (1983), Hotz and Miller (1993), Rust (1994), Altuğ and Miller (1998), Magnac and Thesmar (2002), Arcidiacono and Miller (2011, 2015), and Llull (2018a).

The data consist of one-year panels with information on choices, idiosyncratic state variables, and wages for a sample of individuals that is representative of the United States between 1993 and 2015. Additional aggregate data are needed for identification: aggregate output, capital stocks (equipment, structures, and IPP), and native and immigrant cohort sizes. The distribution of initial skills (at age 16 for natives, and at entry for immigrants) and of observable types are necessary for simulation, but not used in identification and estimation.

[24] Llull (2018a) provides evidence that, in equilibrium, aggregate flows seem to correlate with aggregate shocks, while composition remains rather invariant. The counterfactual evolution of these variables are precisely determined by the design of the policy experiments simulated below.

A. CCPs

Let $\{p_j(h_a, r_t)\}_{j\in\mathcal{D}(h_a)}$ denote the CCPs. They are not directly identified from the data, as r_t is not observed. In order to recover them, I first note that calendar time t is a sufficient statistic for r_t^*, the vector of equilibrium skill prices at time t. I also note that $\{\tilde{p}_j(h_a, t)\}_{j\in\mathcal{D}(h_a)}$ is non-parametrically identified from observed choices by individuals with state vector h_a at time t. Having identified them, I use $\{\tilde{p}_j(h_a, t)\}_{j\in\mathcal{D}(h_a)}$ to recover equilibrium skill prices r_t^* as described below. Finally, I use recovered skill prices to identify $\{p_j(h_a, r_t)\}_{j\in\mathcal{D}(h_a)}$ from observed choices by individuals when skill prices are r_t. Specifically, I exploit that the only source of non-stationarity in the worker's problem are skill prices, and use $\{p_j(h_a, r)\}_{j\in\mathcal{D}(h_a)}$ as the counterfactual CCPs for an individual with state vector h_a at time t if skill prices were equal to r instead of r_t^*.

B. Wage function and aggregate skill prices

The skill production function $\{\tilde{s}_j(h_a)\}_{j\in\{B,W,T\}}$ is identified from individual wage data. Taking logs to (14) yields:

$$\ln w_j(h_a, \eta_a, r_t) = \ln r_{jt} + \tilde{s}_j(h_a) + \sigma_{j\ell}\eta_a. \tag{35}$$

In the absence of self-selection, skill prices would be identified as time dummies, and $\tilde{s}_j(h_a)$ would be identified as a non-parametric function of h_a. The scale of $\ln r_{jt}$ and $\tilde{s}_j(h_a)$ is not separately identified, so $\tilde{s}_j(h_*)$ for some point of the state space h_* is normalized to zero. Even subject to this normalization, these functions are not identified from least squares regression on Equation (35), because $\mathbb{E}[\eta_a|d_{jt} = 1, h_a, t] \neq 0$ (self-selection). Alternatively, I follow Lee (1983) to express (35) as:

$$\ln w_j(h_a, \eta_a, r_t) = \ln r_{jt} + \tilde{s}_j(h_a) + \sigma_{j\ell}\omega_{j\ell}\lambda(\tilde{p}_j(h_a, t)) + \nu_a, \tag{36}$$

where $\omega_{j\ell}$ is a nuisance parameter associated to the degree of endogenous self-selection in wages, $\lambda(\tilde{p}_j(h_a, t)) \equiv \phi(\Phi^{-1}(\tilde{p}_j(h_a, t)))/\tilde{p}_j(h_a, t)$ is the selection correction term (where $\phi(\cdot)$, $\Phi(\cdot)$, and $\Phi^{-1}(\cdot)$ are the standard normal density, cumulative distribution function, and its inverse respectively), and ν_a is an error term that is orthogonal to h_a and t. As $\tilde{p}_j(h_a, t)$ is identified, $\lambda(\tilde{p}_j(h_a, t))$ is identified, and, thus, r_t, $\tilde{s}_j(h_a)$, and $\sigma_{j\ell}\omega_{j\ell}$ are identified from least squares regression on (36).

As discussed in the literature (see Vella (1998) for a survey), credible identification requires an exclusion restriction. In this model, the number of children in the household, n_a affects the utility to stay home, but does not affect wages.

Given this exclusion restriction, the normality assumption could be relaxed, and $\lambda(\tilde{p}_j(h_a,t))$ could still be identified nonparametrically, as noted by Das, Newey and Vella (2003). In the estimation below, I check the stability of the parameter estimates to this assumption.

Finally, $\sigma_{j\ell}$ is identified from the conditional (residual) variance of wages, which has the following form (Lee, 1983):

$$\mathbb{E}[\nu_a^2|j,h_a,t] = \sigma_{j\ell}^2 - (\sigma_{j\ell}\omega_{j\ell})^2 \left[\Phi^{-1}(\tilde{p}_j(h_a,t)) + \lambda(\tilde{p}_j(h_a,t))\right] \lambda(\tilde{p}_j(h_a,t)). \quad (37)$$

By inspection, $\sigma_{j\ell}$ is identified as all other elements of Equation (37) are identified.

C. Production function and expectation parameters

Combining identified individual skill units with aggregate data on cohort sizes, equilibrium aggregate skill units are identified as the aggregation of individual skill units over all individuals employed in each occupation. Given identified aggregate skill units and skill prices, aggregate data on capital (K_{Et}, K_{St}, and I_t) and output (Y_t), and assumed depreciation rates for strutures and equipment, interest rates are identified for the specified production function because equipment, structures, STEM, white collar, and blue collar shares add to one. Let Γ_{lt} denote the share of output devoted to compensate input l. The three labor shares are identified given that skill prices and aggregate skill units are identified. Imposing the condition $r_{Et} - \delta_E = r_{St} - \delta_S$, interest rates are identified solving for r in:

$$1 - \Gamma_{Tt} - \Gamma_{Wt} - \Gamma_{Bt} = \frac{rK_{Et}}{Y_t} + \frac{(r + \delta_S - \delta_E)K_{St}}{Y_t}. \quad (38)$$

Production function parameters are identified from demand equations (7) through (11) as follows. Rewriting Equation (10) as a factor share (Γ_{St}), deriving the analogous expression for $1-\Gamma_{St}$, dividing the former by the latter, and taking logs yields:

$$\ln \frac{\Gamma_{St}}{1 - \Gamma_{St}} = \ln \frac{\varsigma_t}{1 - \varsigma_t} = \tilde{\varsigma}_0 + \tilde{\varsigma}_1 I_t. \quad (39)$$

Thus, the parameters $\tilde{\varsigma}_0$ and $\tilde{\varsigma}_1$ are identified in the above expression as regression coefficients of log relative shares on a constant and IPP capital. Combining Equations (7) and (11) and taking logs to the resulting expression gives, upon rearrangement:

$$\ln \frac{\Gamma_{Tt}}{\Gamma_{Et}} = \tilde{\iota}_0 + \tilde{\iota}_1 I_t + \psi \ln \left(\frac{S_{Tt}}{K_{Et}} \right). \quad (40)$$

Equation (40) provides the basis for identification of $\tilde{\iota}_0$, $\tilde{\iota}_1$, and ψ, which can be obtained as regression coefficients. Having identified these parameters, Q_{1t} is identified. Combining Equations (7) and (8), taking logs to the resulting expression, and rearranging gives:

$$\ln \frac{\Gamma_{Wt}}{\Gamma_{Tt}} + \ln \iota_t + \psi \ln \frac{S_{Tt}}{Q_{1t}} = \tilde{\theta}_0 + \tilde{\theta}_1 I_t + \kappa \ln \left(\frac{S_{Wt}}{Q_{1t}} \right). \tag{41}$$

Proceeding analogously with (8) and (9), $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, and ρ are identified as regression coefficients from the following expression:

$$\ln \frac{\Gamma_{Bt}}{\Gamma_{Wt}} + \ln \theta_t + \kappa \ln \frac{S_{Wt}}{Q_{2t}} = \tilde{\alpha}_0 + \tilde{\alpha}_1 I_t + \rho \ln \left(\frac{S_{Bt}}{Q_{2t}} \right). \tag{42}$$

The parameters associated to IPP capital, φ, χ_1, and χ_2, and those associated to the aggregate shock processes π_ξ, π_ζ, σ_ξ, and σ_ζ are identified as follows. Having identified ς_t, α_t, ρ, θ_t, κ, ι_t, and ψ, the term $\zeta_t I_t^\varphi$ is identified as the residual in Equation (3), which I denote by z_t. Taking logs and first differences, and substituting Equation (5) into the resulting expression yields:

$$\Delta \ln z_t = \pi_\zeta + \varphi \Delta \ln I_t + \sigma_\zeta \upsilon_{\zeta t}. \tag{43}$$

Even though $\Delta \ln I_t$ is correlated with $\upsilon_{\zeta t}$ because $\ln I_t$ is, $\ln I_{t-1}$ is a valid instrument because it is correlated with $\Delta \ln I_t$ but not with $\upsilon_{\zeta t}$. Thus, π_ζ and φ are identified as (instrumental variable) regression coefficients, and σ_ζ is identified as the variance of the residual. Similarly, taking logs and first differences to (1), and substituting Equation (2) into the resulting expression, we obtain, upon rearrangement:

$$\Delta \ln(\Delta I_t) = \pi_\xi + \chi_1 \Delta \ln K_{Et} + \chi_2 \ln S_{Tt} + \sigma_\xi \upsilon_{\xi t}. \tag{44}$$

Thus, π_ξ, χ_1, χ_2, and σ_ξ are identified in an analogous way using $\ln K_{Et-1}$ and $\ln S_{Tt-1}$ as an instrument for $\Delta \ln K_{Et}$ and $\Delta \ln S_{Tt}$.[25]

Finally, identifying $\sigma_\zeta \upsilon_{\zeta t}$ and $\sigma_\xi \upsilon_{\xi t}$ as the residuals in Equations (5) and (2) respectively, the reduced form parameters of the expectation rule in the baseline equilibrium, Ξ_{0j}, Ξ_{1j}, and Ξ_{2j} for $j \in \{B, W, T\}$ are identified in Equation (24) as regression coefficients and $\sigma_{\Upsilon j}$ for $j \in \{B, W, T\}$ is identified as the variance of the residual.

D. Utility parameters

The remaining paramters to identify are the re-entry costs for the three working alternatives $\{\Lambda_j\}_{j \in \{B,W,T\}}$, the deterministic parts of the utilities of school and

[25] The approach described by Equations (43) and (44) exploits similar variation than the popular estimation method proposed by Olley and Pakes (1996), except that Olley and Pakes (1996) also exploit the panel dimension of their firm microdata.

home alternatives, $\{\tau_i(\ell)\}_{i\in\{0,1,2\}}$ and $\vartheta(\ell, n_a)$, and the Generalized Extreme Value parameter ϱ. Define the vector of CCPs as $p \equiv (p_B, p_W, p_T, p_S, p_H)'$. Appealing to Hotz and Miller (1993) (with the reformulation in Arcidiacono and Miller, 2011), there exist mappings $\mu_j(p)$ such that:

$$\mu_j[p(h_a, r_t)] \equiv \bar{V}(h_a, r_t) - v_j(h_a, r_t), \text{ for any } j \in \mathcal{D}(h_a). \tag{45}$$

Given the distributional assumption for $F_\varepsilon(\cdot)$ in this model, $\mu_j(p)$ specializes to:

$$\mu_j(p) = \begin{cases} \gamma - \varrho \ln p_j - (1-\varrho)\ln\left(\sum_{k\in\{B,W,T\}} p_k\right) & \text{if } j \in \{B, W, T\} \\ \gamma - \ln p_j & \text{if } j \in \{S, H\}, \end{cases} \tag{46}$$

where γ is the Euler's constant, approximately equal to 0.5772 (see Lemma 3 in Arcidiacono and Miller, 2011). Substituting (45) into (22) yields:

$$v_j(h_a, r_t) = \tilde{u}_j(h_a, r_t) + \beta \int \sum_{h\in\mathcal{H}_{a+1|h_a,j}} [v_k(h, r) + \mu_k(p(h, r))]\, P_h(h|h_a, j) dF_r(r|r_t), \tag{47}$$

for an arbitrary $k \in \mathcal{D}(h_{a+1})$. Since n is the only element of h whose transition probability is not degenerate, $\mathcal{H}_{a+1|h_a,j}$ includes only three elements.

The three sets of parameters are identified differently. For the re-entry and home parameters, I exploit finite dependence, as in Arcidiacono and Miller (2011). In particular, consider two sets of sequential choices for a given state vector. The first one is to stay home this period, and then stay home again in the next period. The second one is to work in one of the three occupations this period, and stay home in the next period. These two sequences provide identical (expected) continuation values after the second period, which, thus, cancel out when subtracting one alternative-specific value function to another. Given this, I set $k = H$. Evaluating (47) for $j \in \{B, W, T\}$ and $j' = H$, substituting (46) into the resulting expressions, and subtracting one result to the other yields, upon rearrangement:

$$\begin{aligned} v_j(h_a, r_t) - v_H(h_a, r_t) = {} & \ln r_{jt} + \tilde{s}_j(h_a) + \Lambda_{0j} - \Lambda_{1j} d_{5a-1} - \vartheta(\ell, n_a) \\ & - \beta \int \sum_{n\in\mathcal{C}} \ln \frac{p_H(h_j(n, h_a), r)}{p_H(h_H(n, h_a), r)} P_n(n|h_a, H) dF_r(r|r_t), \end{aligned} \tag{48}$$

where $h_j(n, h_a) \equiv (a+1, \ell, E_a, j, n, \tilde{a})'$, and where I exploit that $P_n(n|h_a, H) = P_n(n|h_a, j)$ by assumption, since only choosing education affects the fertility transition. Finally, solve for $v_j(h_a, r_t)$ in (45), substitute the resulting expression (evaluated for j and H) into the left-hand side of (48), and obtain:

$$\varrho \ln \frac{p_j(h_a, r_t)}{\sum_{k \in \{B,W,T\}} p_k(h_a, r_t)} + \vartheta(\ell, n_a) = \ln \frac{p_H(h_a, r_t)}{\sum_{k \in \{B,W,T\}} p_k(h_a, r_t)} \tag{49}$$
$$+ \ln r_{jt} + \tilde{s}_j(h_a) - \Lambda_j d_{5a-1} - \beta \int \sum_{n \in \mathcal{C}} \ln \frac{p_H(h_j(n, h_a), r)}{p_H(h_H(n, h_a), r)} P_n(n|h_a, H) dF_r(r|r_t).$$

The right hand side of the above equation is identified from the arguments above. Thus, $\{\Lambda_j\}_{j \in \{B,W,T\}}$, ϱ, and $\{\vartheta(\ell, n)\}_{\ell \in \mathcal{L}}^{n \in \mathcal{C}}$ are identified as a result of evaluating (49) at all points of the state space and occupational choices, which provides an overdetermined system of linear equations.

School parameters cannot be identified exploiting any form of finite dependence because dropping out from school is an absorbing state. However, for the same reason, $v_j(h_a, r_t)$ is identified for $j \in \{B, W, T, H\}$ using the above arguments, because returning to school is not an option (and thus $\tau_k(\ell)$ does not appear in the value functions) and all other parameters are identified. Evaluating (47) for $j = H$ and $j' = S$, substituting in (45), subtracting the resulting expressions, and rearranging gives:

$$\tau_0(\ell)\, \mathbb{1}\{E_a < 12\} + \tau_1(\ell)\, \mathbb{1}\{12 \leq E_a < 16\} + \tau_2(\ell)\, \mathbb{1}\{E_a \geq 16\} = \vartheta(\ell, n_a) \tag{50}$$
$$+ \ln \frac{p_S(h_a, r_t)}{p_H(h_a, r_t)} - \beta \int \left\{ \begin{array}{l} \sum_{h \in \mathcal{H}_{a+1|h_a,S}} \left[\begin{array}{c} v_H(h, r) \\ - \ln p_H(h, r) \end{array} \right] P_h(h|h_a, S) \\ - \sum_{h \in \mathcal{H}_{a+1|h_a,H}} \left[\begin{array}{c} v_H(h, r) \\ - \ln p_H(h, r) \end{array} \right] P_h(h|h_a, H) \end{array} \right\} dF_r(r|r_t).$$

By inspection, $\{\tau_k(\ell)\}_{k \in \{0,1,2\}}^{\ell \in \mathcal{L}}$ is identified because the right hand side of the above equation is identified. A similar argument could be done evaluating (47) at any $j \in \{B, W, T\}$ instead, and subtracting it again to the same function evaluated at $j = S$. As I discuss below, in estimation I use both the expression with $j = H$, and those with $j \in \{B, W, T\}$.

V. Estimation

I proceed with estimation following a stepwise procedure that closely mimics the identification arguments above. To do so, I combine aggregate data from different sources with use two different micro-datasets: the March Supplements of the CPS linked over two consecutive years for the period 1993-2015, and the SIPP panel also matched over two consecutive years for the period 1988-2007. First, I estimate the CCPs. Second, I estimate the parameters of the wage equation. Third, I proceed with the estimation of the representative firm problem. And fourth, I estimate the remaining utility parameters. Variable definitions and sample selection are specified in Appendix A.

A. CCPs and transition functions

The nonparametric estimates of the CCPs, $\hat{\tilde{p}}(h_a, t)$, are obtained from running flexibly specified multinomial logit models on different subsamples. Because previous choice and having a college degree determine the choice set, these two characteristics always define subsamples. I further divide subsamples by types ℓ, but several types are often grouped due to sample size concerns. After $\hat{r}_t$ is obtained (as described in Section V.B), $\hat{p}(h_a, \hat{r}_t)$ are obtained with an analogous approach, replacing calendar time by the estimate of skill prices. The probability of attending school is set to zero, $\hat{p}_S = \hat{\tilde{p}}_S = 0$, when $d_{a-1} \neq S$, and the probability of working in STEM is set to zero, $\hat{p}_T = \hat{\tilde{p}}_T = 0$, if the worker does not possess a college degree, $E < 16$.

The transition functions are all degenerate except for $P_n(n|h_a, d_a)$. As noted above, this function is assumed to depend on the current choice only if this is schooling. Furthermore, I assume that the dependence on h_a is through type, education level, age, and current number of children. The transition probability matrix is estimated nonparametrically using Census data for 1970-2000, and ACS data for 2001-2015. The probabilities are estimated on subsamples determined by type, education level, and current number of children. The dependence on age is obtained by means of a logit that includes a flexible polinomial in age.

B. Wage function

Even though $\tilde{s}_j(h_a)$ is nonparametrically identified, I parametrize it to obtain more precise estiamtes. Let $X_a \equiv a - \max\{16, E+6\}$ denote potential experience, and $\tilde{X}_a \equiv \max\{0, \tilde{a} - E_a - 6\}$ denote potential experience abroad. The function $\tilde{s}_j(h_a)$ specializes to the following Mincerian regression (Mincer, 1974):

$$\begin{aligned}\tilde{s}_j(h_a) \equiv\ & (\flat_{1j} + \flat_{2j}\, \mathbb{1}\{\tilde{\ell}(\ell) \in \tilde{\mathcal{L}}_3 \cup \tilde{\mathcal{L}}_4\} + \flat_{3j}\, \mathbb{1}\{d_{a-1} \neq j\}) \times E_a \\ & + (\flat_{4j} + \flat_{5j}\, \mathbb{1}\{d_{a-1} \neq j\}) \times X_a \\ & + (\flat_{6j} + \flat_{7j}\, \mathbb{1}\{d_{a-1} \neq j\}) \times X_a^2 \\ & + \flat_{8j}\tilde{X}_a + \sum_{k \in \mathcal{L}} \flat_{9kj}\, \mathbb{1}\{k = \ell\},\end{aligned} \tag{51}$$

where $\tilde{\ell}(\ell) \in \tilde{\mathcal{L}}_3 \cup \tilde{\mathcal{L}}_4$ denotes that the type ℓ is included in the immigrant subset. Substituting this expression into Equation (36), and evaluating $\lambda(p)$ at the CCPs $\tilde{p}_j(h_a, t)$ estimated in the previous step, the wage function parameters are estimated by least squares estimates on the resulting expression. The equilibrium skill prices are obtained as the coefficients associated to calendar time dummies.

This procedure provides estimates for $\widehat{\ln r}_{jt}$, $\hat{\hat{s}}_j(h)$, and $\widehat{\sigma_{j\ell}\omega_{\ell j}}$. The variance parameter $\sigma_{j\ell}$ is obtained from the sample analogs of moment conditions implied by (37). In particular, by the law of iterated expectations, (37) implies:

$$\sigma_{j\ell}^2 = \mathbb{E}\left[\nu_a^2 + (\sigma_{j\ell}\omega_{j\ell})^2\left[\Phi^{-1}(\tilde{p}_j(h_a,t)) + \lambda(\tilde{p}_j(h_a,t))\right]\lambda(\tilde{p}_j(h_a,t))\middle| d_{ja}=1,\ell\right]. \tag{52}$$

A consistent estimator is provided by the sample analog of this expression using the estimated CCPs and $\widehat{\sigma_{j\ell}\omega_{\ell j}}$. To obtain more precise estimates, I only allow $\sigma_{j\ell}$ to differ by gender and immigrant status (native/immigrant), this is, I assume $\sigma_{j\ell} = \sigma_{j\ell'}$ if $\tilde{\ell}(\ell) = \tilde{\ell}(\ell')$.

C. Production function and expectation parameters

Let $\Pi^{(i,t)}$ denote the population elevation factor for individual i for year t, and let N denote sample size. Having recovered $\widehat{\ln r}_t$, aggregate skill units in occupation j are obtained aggregating individual skill units over all individuals working in occupation j:

$$\hat{S}_{jt} = \sum_{i=1}^{N} \Pi^{(i,t)} d_j^{(i)} \exp\left\{\ln w_j^{(i)} - \widehat{\ln r}_{jt_i}\right\}. \tag{53}$$

Production function parameters are obtained combining these with data on structures, equipment, and IPP capital, and output. In particular, they are obtained sequentially, from Equations (38) through (44). Interest rates are recovered solving for r_{Kt} in (38). Then, I estimate $\tilde{\varsigma}_0$ and $\tilde{\varsigma}_1$ from a linear regression in (39), and $\tilde{\iota}_0$, $\tilde{\iota}_1$, and ψ from a regression in (40).[26] Using these estimates, I construct $\hat{A}_{1t}$ and estimate $\tilde{\theta}_0$, $\tilde{\theta}_1$, and κ from (41). Similarly, I construct $\hat{A}_{2t}$ and estimate $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, and ρ from (42). Using $\ln K_{It-1}$ as an instrument for $\Delta \ln I_t$, I estimate π_ζ and φ as the instrumental variables (IV) coefficients of Equation (43) (and σ_ζ is obtained as the estimated residual variance). Analogously, π_ξ, χ_1, χ_2, and σ_ξ are obtained from IV estimation of (44) using using $\ln K_{Et-1}$ and $\ln S_{Tt-1}$ as instruments for $\Delta \ln K_{Et}$ and $\Delta \ln S_{Tt}$. Finally, I obtain the baseline values of Ξ_{0j}, Ξ_{1j}, and Ξ_{2j} from a least squares regressions on (24) for each occupation using the predicted values for skill prices and aggregate shocks obtained from previous regressions, obtaining $\sigma_{\Upsilon j}$ as the residual variances.

[26] Equations (39) through (42) should hold exactly in the population. However, because some elements of the equations are estimated in the sample, I allow for measurement error in these expressions.

D. Utility parameters

Home utility parameters and the correlation parameter of the GEV distribution are obtained from moment conditions specified based on (49). To do so, I first compute, for each individual, the following expression:

$$\hat{\Omega}_{Hj}^{(i)} \equiv 0.95 \iiiint\!\!\int \sum_{n \in \mathcal{C}} \left\{ \begin{array}{l} \ln \dfrac{\hat{p}_H\left(h_j(n,h^{(i)}), \hat{r}(\upsilon_\zeta, \upsilon_\xi, \Upsilon_T, \Upsilon_W, \Upsilon_B, \hat{r}_t)\right)}{\hat{p}_H\left(h_H(n,h^{(i)}), \hat{r}(\upsilon_\zeta, \upsilon_\xi, \Upsilon_T, \Upsilon_W, \Upsilon_B, \hat{r}_t)\right)} \\ \times \hat{P}_n(n|h^{(i)}, H)\phi(\upsilon_\zeta)\phi(\upsilon_\xi)\phi(\Upsilon_T)\phi(\Upsilon_W)\phi(\Upsilon_B) \end{array} \right\} \begin{array}{l} d\upsilon_\zeta d\upsilon_\xi \\ d\Upsilon_T d\Upsilon_W, \\ d\Upsilon_B \end{array} \tag{54}$$

where $\hat{r}(\upsilon_\zeta, \upsilon_\xi, \Upsilon_T, \Upsilon_W, \Upsilon_B, \hat{r}_t)$ is the vector of predicted skill prices from Equation (24). Furthermore, I parametrize $\vartheta(\ell, n_a)$ as:

$$\vartheta(\ell, n) \equiv \vartheta_{0\ell} + \sum_{k=1}^{4} \vartheta_{1k}\, \mathbb{1}\{\tilde{\ell}(\ell) \in \tilde{\mathcal{L}}_k\} n. \tag{55}$$

In words, $\vartheta_{0\ell}$ denotes the type-specific intercept, and ϑ_{1k} for $k \in \{1,2,3,4\}$ denote how this utility is shifted by each children in the household respectively for native male, native female, immigrant male, and immigrant female. The parameters $\{\Lambda_j\}_{j\in\{B,W,T\}}$, ϱ, $\{\vartheta_{0\ell}\}_{\ell\in\mathcal{L}}$, and $\{\vartheta_{1k}\}_{k\in\{1,2,3,4\}}$ are estimated by least squares from:

$$\left(\ln \frac{\hat{p}_H(h^{(i)}, r_{t_i})}{\sum_{k\in\{B,W,T\}} \hat{p}_k(h^{(i)}, r_{t_i})} + \widehat{\ln r}_{jt(i)} + \hat{\hat{s}}_j(h^{(i)}) - \hat{\Omega}_{Hj}^{(i)} \right) = \tag{56}$$

$$\Lambda_j d_{5a-1} + \varrho \ln \frac{\hat{p}_j(h^{(i)}, r_{t_i})}{\sum_{k\in\{B,W,T\}} \hat{p}_k(h^{(i)}, r_{t_i})} + \sum_{\ell\in\mathcal{L}} \vartheta_{0\ell}\, \mathbb{1}\{\ell_i = \ell\} + \sum_{k=1}^{4} \vartheta_{1k}\, \mathbb{1}\{\tilde{\ell}(\ell_i) \in \tilde{\mathcal{L}}_k\} n_i.$$

This regression is estimated on a synthetic dataset generated by expanding the original dataset with three observations per individual, one for each occupation.

Finally, for the estimation of the schooling parameters one needs to compute the value functions that appear in the last term of (50). As, this is costly computationally, I follow Hotz et al. (1994) and Altuğ and Miller (1998) and use simulation methods to approximate them. In particular, for each individual i, I simulate M sequences of skill prices and children (denoted by m_i) for periods $l_i \in \{1, ..., 65 - a_i\}$, drawing from the skill prices and children transition functions at each simulation point (m_i, l_i). The latter is done by using the estimated transition probabilities $\hat{P}_n(n|h_a, d_a)$ to partition the unit interval in three groups (for 0, 1, and 2+ children), and draw from a uniform distribution to assign a particular transition to that individual at age $a_i + l_i$. Thus, in each simulation m_i, the individual is exposed to a sequence of skill prices

$\{r^{(m_i)}_{t_i+l_i}\}_{l_i=1,\dots,65-a_i}$, and a five sequences of state variables, depending on the initial choice: $h_{jl_i}(n^{(m_i)}_{jl_i}, h^{(i)}) \equiv (a_i + l_i, \ell_i, E^{(i)} + \mathbb{1}\{j = S\}, j, n^{(m_i)}_{jl_i}, \tilde{a}_i)$. The difference across sequences is as follows: at $l_i = 1$, the previous choice varies across paths; education increases in one unit if $j = S$ and stays constant otherwise; and there are two sequences of children draws, depending on whether the choice is school or not (the distinction between $n^{(m_i)}_{Sl_i}$ and $n^{(m_i)}_{jl_i}$ is necessary because children transition probabilities vary by education, but $n^{(m_i)}_{jl_i} = n^{(m_i)}_{j'l_i}$ for any $j, j' \neq S$ for the same reason). Using these simulations, the last term of (50) for individual (i) is approximated by:

$$\hat{\Omega}^{(i)}_{Sj} = \frac{1}{M}\sum_{m_i=1}^{M}\sum_{l_i=1}^{65-a_i} 0.95^{l_i}\left[\begin{array}{r} \displaystyle\sum_{k=1}^{4}\hat{\vartheta}_{1k}\,\mathbb{1}\{\tilde{\ell}(\ell_i) \in \tilde{\mathcal{L}}_k\}\left(n^{(m_i)}_{Sl_i} - n^{(m_i)}_{jl_i}\right) \\ \displaystyle - \ln\frac{\hat{p}_H\left(h_{Sl_i}(n^{(m_i)}_{Sl_i}, h^{(i)}), r^{(m_i)}_{t_i+l_i}\right)}{\hat{p}_H\left(h_{jl_i}(n^{(m_i)}_{jl_i}, h^{(i)}), r^{(m_i)}_{t_i+l_i}\right)}\end{array}\right], \tag{57}$$

for $j \in \{B, W, T, H\}$. Furthermore, I parametrize the school utility functions as:

$$\tau_0(\ell) \equiv \tau_{0\ell}; \quad \tau_1(\ell) \equiv \tau_{0\ell} + \tau_1; \quad \tau_2(\ell) \equiv \tau_{0\ell} + \tau_2. \tag{58}$$

The parameters $\{\tau_{0\ell}\}_{\ell\in\mathcal{L}}$, τ_1, and τ_2 are estimated by least squares from:

$$\begin{aligned}\left(\ln\frac{\hat{p}_S(h^{(i)}, r_{t_i})}{\hat{p}_H(h^{(i)}, r_{t_i})} + \sum_{\ell\in\mathcal{L}}\hat{\vartheta}_{0\ell}\,\mathbb{1}\{\ell_i = \ell\} + \sum_{k=1}^{4}\hat{\vartheta}_{1k}\,\mathbb{1}\{\tilde{\ell}(\ell_i) \in \tilde{\mathcal{L}}_k\}n^{(i)} - \hat{\Omega}^{(i)}_{SH}\right) = \\ \sum_{\ell\in\mathcal{L}}\tau_{0\ell}\,\mathbb{1}\{\ell_i = \ell\} + \tau_1\,\mathbb{1}\{12 \leq E^{(i)} < 16\} + \tau_2\,\mathbb{1}\{E^{(i)} \geq 16\},\end{aligned} \tag{59}$$

and:

$$\begin{aligned}\left(\ln\frac{\hat{p}_S(h^{(i)}, r_{t_i})}{\sum_{k\in\{B,W,T\}}\hat{p}_k(h^{(i)}, r_{t_i})} - \hat{\varrho}\ln\frac{\hat{p}_j(h^{(i)}, r_{t_i})}{\sum_{k\in\{B,W,T\}}\hat{p}_k(h^{(i)}, r_{t_i})} + \widehat{\ln r}_{jt(i)} + \hat{\tilde{s}}_j(h^{(i)}) - \hat{\Omega}^{(i)}_{Sj}\right) = \\ \sum_{\ell\in\mathcal{L}}\tau_{0\ell}\,\mathbb{1}\{\ell_i = \ell\} + \tau_1\,\mathbb{1}\{12 \leq E^{(i)} < 16\} + \tau_2\,\mathbb{1}\{E^{(i)} \geq 16\},\end{aligned} \tag{60}$$

for $j \in \{B, W, T\}$.[27] The first expression is obtained from the difference between the school and home conditional value functions. The second one is obtained from the difference between the conditional value functions of working in occupation j and attending school. The two expressions are estimated jointly in a single

[27] Note that the term $\hat{\Lambda}_j d^{(i)}_{5a-1}$ does not appear in the left hand side of the expression because, by construction, $d^{(i)}_{5a-1} = 0$ for all individuals in the sample used to estimate this regression. Additional estimation results, available upon request, provide very similar results estimating the school parameters from (59) alone.

regression on a synthetic dataset generated by expanding the original data with four observations per individual, one for each alternative.

E. Refinements

In order to improve the efficiency of the estimates and to correct for potential biased generated by sampling error in the estimation of the CCPs, I introduce several refinements to the estimation procedure outlined above.

Sampling error in the estimation of the CCPs. The estimation of the CCPs, even with relatively large datasets, is subject to potentially non-trivial sampling error. As a result, the estimation of equations that include regressors formed off estimated CCPs may be subject to the standard attenuation bias. The estimation of the model using with both CPS and SIPP provides a natural method to correct for this measurement error. Given that CPS and SIPP provide two independent measurements of the same population object, they can be used to instrument each other in estimation, thus correcting the measurement error bias. This refinement is used in the estimation of the wage equation (to instrument $\lambda(\tilde{p}_j(h_a, t))$), and in the home utility (to instrument the term associated to ϱ in (56)).

Aguirregabiria and Mira (2002) for the labor supply. These authors propose an iterative procedure that combines CCP estimation and the solution of the model to obtain more precise estimates. Intuitively, their estimator obtains CCP estimates, solves the model with them to obtain updated CCPs, and perform CCP estimation again with the updated CCPs. They prove that this algorithm nests both the standard Hotz and Miller (1993) estimator (no iteration) and the full solution estimation, which is obtained iterating this procedure until convergence. Every intermediate iteration provides consistent estimates that are more efficient than those from the previous iteration. In the context of this model, I treat skill prices as additional parameters, and I iterate over the labor supply estimation.

Production function estimates. One of the key difficulties to obtain precise estimates in this paper is the reduced number of time periods available in the data. Since production function parameters are estimated off time series variation in aggregate variables, they are obtained with less than 30 observations in all baseline specifications.[28] To refine this part of the estimation I iterate over equilibrium

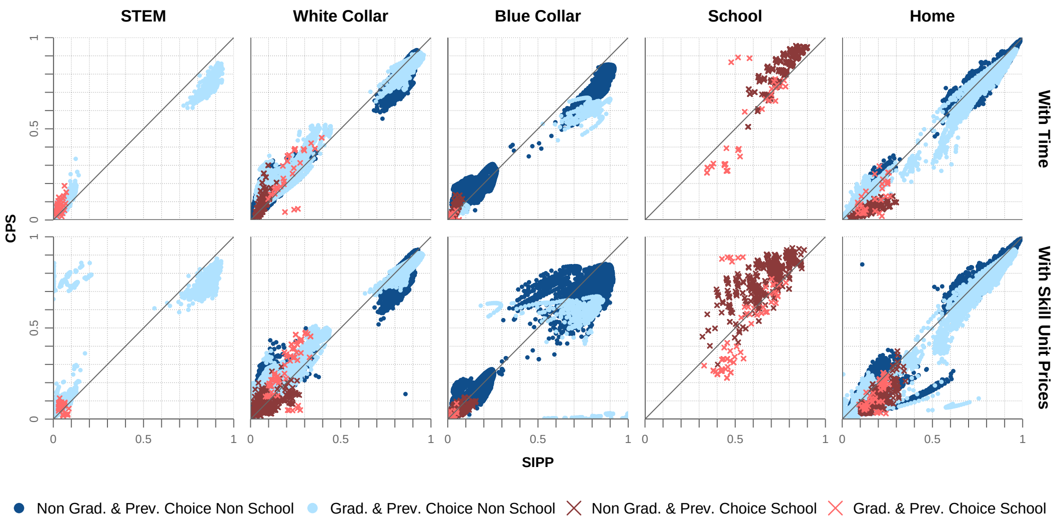
[28] The baseline specifications include the estimation with the CPS (21 observations) and an additional estimation with CPS data extrapolated from 1993 to 1989 using SIPP estimates of skill prices and aggregate skill units (25 observations). Estimation with only SIPP was deemed too imprecise (14 observations). The only exception is Equation (39), which does not require data on skill prices or aggregate skill units and is quite precisely estimated (with 47 observations).

simulations (holding labor supply parameters fixed). This refinement provide two specific improvements. First, it allows me to simulate skill prices and aggregate skill units for the entire period for which I have data on aggregate variables. Second, it obtains production function parameter estimates and a sequence of skill prices that are internally consistent with each other.

Standard errors. Regression standard errors in each step do not take into account that some of the variables included in the regression are themselves estimated. To correct for that, I obtain standard errors through bootstrap.

VI. Parameter Estimates and Goodness of Fit

FIGURE 5. PREDICTED CCPs (CPS VS SIPP)



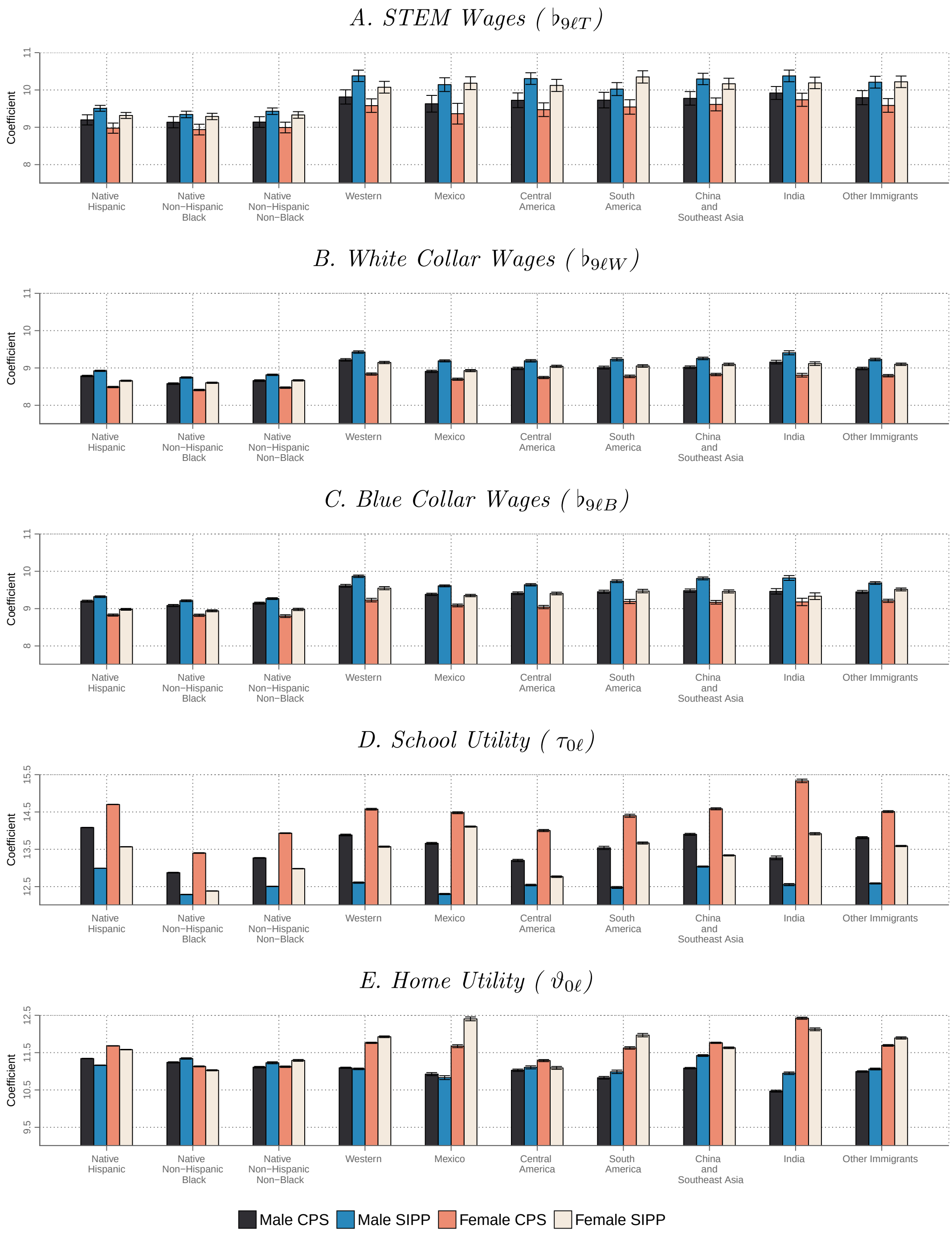
Note: The figure represents the CPS vs SIPP estimated probabilities that an individual with a given set of state variables choses each of the five alternatives. The 45-degree line is represented in each plot. Darker color represents state-space points without college (STEM alternative is not available), whereas lighter colors represent points with a college degree. Dots represent that previous choice was not school (school alternative is not available), while crosses represent state-space points in which previous choice was school. Top panel represents CCPs conditional on calendar time and bottom panel plots estimated CCPs conditional on skill prices. For visibility and file size purposes, a 2 percent random sample of all state points with at least one observation in at least one of the two datasets is represented. Results with the whole set of state points available in at least one of the datasets, available from the author upon request, show a very similar picture.

Table 2—Wage Function Parameters — CCP Estimation

	CPS		SIPP	
	IV to correct for CCP measurement error:			
	No	Yes	No	Yes
A. STEM:				
Education ($\flat_{1T}$)	0.091	0.096	0.079	0.078
	(0.003)	(0.005)	(0.004)	(0.004)
Education × immigrant ($\flat_{2T}$)	-0.031	-0.030	-0.045	-0.045
	(0.006)	(0.006)	(0.009)	(0.009)
Education × prev. choice ≠ STEM ($\flat_{3T}$)	-0.015	-0.024	-0.021	-0.016
	(0.003)	(0.007)	(0.004)	(0.006)
Potential experience ($\flat_{4T}$)	0.034	0.035	0.031	0.031
	(0.002)	(0.002)	(0.002)	(0.002)
Pot. exp. × prev. choice ≠ STEM ($\flat_{5T}$)	0.010	0.007	0.006	0.009
	(0.003)	(0.004)	(0.004)	(0.005)
Potential experience squared ($\flat_{6T}$)	-0.0006	-0.0006	-0.0006	-0.0006
	(0.000)	(0.000)	(0.000)	(0.000)
Pot. exp. sq. × prev. choice ≠ STEM ($\flat_{7T}$)	-0.0002	-0.0002	-0.0002	-0.0002
	(0.000)	(0.000)	(0.000)	(0.000)
Potential experience abroad ($\flat_{8T}$)	-0.009	-0.010	-0.006	-0.007
	(0.002)	(0.002)	(0.002)	(0.002)
B. White collar:				
Education ($\flat_{1W}$)	0.101	0.107	0.097	0.096
	(0.001)	(0.001)	(0.001)	(0.001)
Education × immigrant ($\flat_{2W}$)	-0.026	-0.027	-0.031	-0.031
	(0.002)	(0.002)	(0.002)	(0.002)
Education × prev. choice ≠ WC ($\flat_{3W}$)	-0.007	-0.017	-0.016	-0.012
	(0.001)	(0.001)	(0.001)	(0.001)
Potential experience ($\flat_{4W}$)	0.035	0.040	0.035	0.034
	(0.001)	(0.001)	(0.001)	(0.001)
Pot. exp. × prev. choice ≠ WC ($\flat_{5W}$)	0.007	0.003	-0.007	-0.005
	(0.001)	(0.001)	(0.001)	(0.001)
Potential experience squared ($\flat_{6W}$)	-0.0006	-0.0007	-0.0006	-0.0006
	(0.000)	(0.000)	(0.000)	(0.000)
Pot. exp. sq. × prev. choice ≠ WC ($\flat_{7W}$)	-0.0001	-0.0001	0.0001	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)
Potential experience abroad ($\flat_{8W}$)	-0.009	-0.009	-0.008	-0.008
	(0.001)	(0.001)	(0.001)	(0.001)
C. Blue collar:				
Education ($\flat_{1B}$)	0.065	0.059	0.060	0.060
	(0.002)	(0.002)	(0.001)	(0.001)
Education × immigrant ($\flat_{2B}$)	-0.039	-0.031	-0.038	-0.039
	(0.002)	(0.002)	(0.002)	(0.002)
Education × prev. choice ≠ BC ($\flat_{3B}$)	0.000	-0.024	-0.015	-0.004
	(0.002)	(0.003)	(0.002)	(0.002)
Potential experience ($\flat_{4B}$)	0.034	0.041	0.035	0.033
	(0.001)	(0.001)	(0.001)	(0.001)
Pot. exp. × prev. choice ≠ BC ($\flat_{5B}$)	0.002	0.000	-0.007	-0.006
	(0.001)	(0.001)	(0.001)	(0.001)
Potential experience squared ($\flat_{6B}$)	-0.0005	-0.0007	-0.0006	-0.0005
	(0.000)	(0.000)	(0.000)	(0.000)
Pot. exp. sq. × prev. choice ≠ BC ($\flat_{7B}$)	-0.0000	-0.0000	0.0001	0.0001
	(0.000)	(0.000)	(0.000)	(0.000)
Potential experience abroad ($\flat_{8B}$)	-0.008	-0.007	-0.008	-0.008
	(0.001)	(0.001)	(0.001)	(0.001)

Note: Regression standard errors (not corrected for error in the estimation of CCPS) in parenthesis.

Figure 6. Type-Specific Coefficients — CCP Estimation (with IV)



Note: The figure represents the type-specific coefficients estimated using CPS and SIPP using IV to correct for measurement error in the estimation of the CCPs (estimates without the IV correction are available from the author upon request). For each national origin/race, the first two columns indicate male, and the last two are for female. The first columns of each block is estimated from the CPS and the second is estimated from the SIPP. Two regression standard error confidence bands (not corrected for estimation error in the CCPs and other regressors) are displayed.

Table 3—Other Utility Parameters — CCP Estimation

	CPS		SIPP	
	IV to correct for CCP measurement error:			
	No	Yes	No	Yes
A. Variances of wages:				
i. STEM				
Male Native (σ_{1T})	0.530	0.542	0.511	0.509
	(0.005)	(0.005)	(0.006)	(0.006)
Female Native (σ_{2T})	0.493	0.504	0.452	0.450
	(0.006)	(0.006)	(0.006)	(0.006)
Male Immigrant (σ_{3T})	0.532	0.537	0.530	0.528
	(0.010)	(0.010)	(0.016)	(0.016)
Female Immigrant (σ_{4T})	0.534	0.542	0.506	0.504
	(0.012)	(0.012)	(0.017)	(0.017)
ii. White collar				
Male Native (σ_{1W})	0.627	0.625	0.594	0.593
	(0.002)	(0.002)	(0.002)	(0.002)
Female Native (σ_{2W})	0.555	0.548	0.512	0.513
	(0.002)	(0.002)	(0.002)	(0.002)
Male Immigrant (σ_{3W})	0.647	0.653	0.619	0.617
	(0.006)	(0.006)	(0.006)	(0.006)
Female Immigrant (σ_{4W})	0.581	0.586	0.543	0.541
	(0.005)	(0.005)	(0.005)	(0.005)
iii. Blue collar				
Male Native (σ_{1B})	0.547	0.561	0.512	0.514
	(0.002)	(0.002)	(0.002)	(0.002)
Female Native (σ_{2B})	0.563	0.575	0.504	0.503
	(0.006)	(0.005)	(0.005)	(0.005)
Male Immigrant (σ_{3B})	0.519	0.549	0.479	0.475
	(0.005)	(0.005)	(0.005)	(0.005)
Female Immigrant (σ_{4B})	0.470	0.493	0.435	0.434
	(0.010)	(0.009)	(0.009)	(0.009)
B. GEV parameter:				
GEV parameter (ϱ)	0.223	0.286	0.279	0.238
	(0.001)	(0.001)	(0.001)	(0.001)
C. School utility parameters:				
College Shifter (τ_1)	-0.007	-0.006	-0.042	-0.038
	(0.004)	(0.004)	(0.002)	(0.002)
Graduate Shifter (τ_2)	-1.136	-1.132	-0.674	-0.659
	(0.008)	(0.008)	(0.004)	(0.004)
D. Home utility parameters (children shifters):				
Male Native (ϑ_{11})	-0.610	-0.636	-0.570	-0.469
	(0.004)	(0.006)	(0.005)	(0.008)
Female Native (ϑ_{12})	0.412	0.591	0.555	0.908
	(0.004)	(0.006)	(0.005)	(0.008)
Male Immigrant (ϑ_{13})	-0.407	-0.516	-0.334	-0.335
	(0.009)	(0.015)	(0.012)	(0.021)
Female Immigrant (ϑ_{14})	0.598	0.865	0.617	0.846
	(0.009)	(0.015)	(0.012)	(0.022)

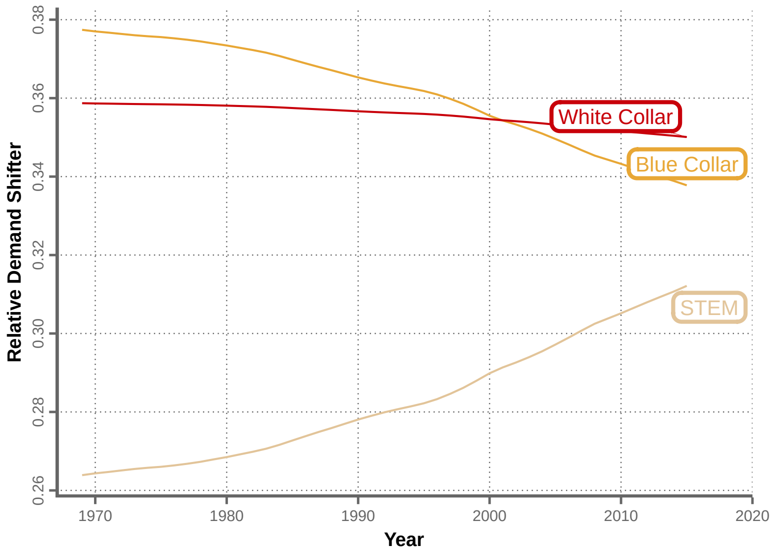
Note: Standard errors (not corrected for estimation error in CCPS and other regressors) in parenthesis.

Table 4—Production Function and Expectation Parameters — CCP Estimation

	CPS		CPS+SIPP	
A. Factor share parameters:				
i. Capital structures				
Constant ($\tilde{\varsigma}_0$)	-1.226	(0.015)	-1.226	(0.015)
IPP Capital/10^{12} ($\tilde{\varsigma}_1$)	0.085	(0.008)	0.085	(0.008)
ii. Blue collar labor				
Constant ($\tilde{\alpha}_0$)	-0.572	(0.028)	-0.576	(0.017)
IPP Capital/10^{12} ($\tilde{\alpha}_1$)	-0.048	(0.024)	-0.060	(0.019)
iii. White collar labor				
Constant ($\tilde{\theta}_0$)	-0.024	(0.190)	0.099	(0.144)
IPP Capital/10^{12} ($\tilde{\theta}_1$)	-0.045	(0.019)	-0.061	(0.014)
iv. STEM labor				
Constant ($\tilde{\iota}_0$)	1.022	(0.246)	1.278	(0.235)
IPP Capital/10^{12} ($\tilde{\iota}_1$)	0.003	(0.014)	0.028	(0.011)
B. Elasticity of substitution parameters:				
Blue collar (ρ)	1.030	(0.093)	0.999	(0.069)
White collar (κ)	1.033	(0.131)	0.916	(0.095)
STEM Equipment (ψ)	0.629	(0.172)	0.849	(0.157)
C. IPP capital parameters:				
Externality (φ)	0.568	(0.526)	0.356	(0.675)
Equipment share (χ_1)	0.956	(0.322)	1.142	(0.626)
STEM share (χ_2)	0.066	(0.097)	0.124	(0.167)
D. Aggregate shocks parameters:				
i. TFP Shock:				
Drift (π_ζ)	-0.017	(0.019)	-0.010	(0.024)
Standard deviation (σ_ζ)	0.015	(0.003)	0.014	(0.002)
ii. IPP Shock:				
Drift (π_ξ)	-0.000	(0.011)	-0.007	(0.024)
Standard deviation (σ_ξ)	0.016	(0.003)	0.016	(0.002)
E. Expectation parameters:				
i. STEM:				
Constant (Ξ_{0T})	0.007	(0.005)	0.007	(0.004)
TFP shock (Ξ_{1T})	0.673	(0.321)	0.721	(0.270)
IPP shock (Ξ_{2T})	-0.148	(0.314)	-0.037	(0.243)
Standard deviation ($\sigma_{\Upsilon T}$)	0.020	(0.003)	0.018	(0.003)
ii. White collar:				
Constant (Ξ_{0W})	0.004	(0.004)	0.003	(0.003)
TFP shock (Ξ_{1W})	0.929	(0.252)	1.032	(0.196)
IPP shock (Ξ_{2W})	-0.015	(0.246)	-0.072	(0.176)
Standard deviation ($\sigma_{\Upsilon W}$)	0.016	(0.003)	0.013	(0.002)
iii. Blue collar:				
Constant (Ξ_{0B})	0.002	(0.004)	0.002	(0.003)
TFP shock (Ξ_{1B})	0.924	(0.248)	0.931	(0.219)
IPP shock (Ξ_{2B})	0.349	(0.243)	0.064	(0.197)
Standard deviation ($\sigma_{\Upsilon B}$)	0.015	(0.003)	0.014	(0.002)

Note: CPS+SIPP indicates that CPS aggregate skill units and skill prices are extrapolated using SIPP estimates. Standard errors (not corrected for estimation error of the regressors) in parenthesis.

Figure 7. Relative Demand Shifters for Each Labor Input



Note: The figure represents three combinations of the estimated values of ς_t, α_t, θ_t, and ι_t that are associated to the relative demand for each of the indicated labor inputs. The statistic associated to blue collar labor is $\frac{\alpha_t}{1-(1-\alpha_t)(1-\theta_t)(1-\iota_t)}$, the one associated to white collar is $\frac{(1-\alpha_t)\theta_t}{1-(1-\alpha_t)(1-\theta_t)(1-\iota_t)}$, and the one associated to STEM is $\frac{(1-\alpha_t)(1-\theta_t)\iota_t}{1-(1-\alpha_t)(1-\theta_t)(1-\iota_t)}$.

VII. Counterfactual Simulations and Policy Analysis

VIII. Conclusions

REFERENCES

Acemoglu, Daron, "Technical Change, Inequality, and the Labor Market," *Journal of Economic Literature*, March 2002, *40* (1), 7–72.

_ **and David H. Autor**, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in Orley C. Ashenfelter and David E. Card, eds., *Handbook of Labor Economics*, Vol. 4B, Amsterdam: North-Holland Publishing Company, 2011, chapter 12, pp. 1043–1171.

Aguirregabiria, Victor and Pedro Mira, "Swapping the Nested Fixed Point Algorithm: a Class of Estimators for Discrete Markov Decision Models," *Econometrica*, July 2002, *70* (4), 1519–1543.

Altuğ, Sumru and Robert A. Miller, "The Effect of Work Experience on Female Wages and Labor Supply," *Review of Economic Studies*, January 1998, *65* (1), 45–85.

Antecol, Heather, Deborah A. Cobb-Clark, and Stephen J. Trejo, "Immigration Policy and the Skills of Immigrants to Australia, Canada, and the United States," *Journal of Human Resources*, Winter 2003, *38* (1), 192–218.

Arcidiacono, Peter and Paul B. Ellickson, "Practical Methods for Estimation of Dynamic Discrete Choice Models," *Annual Review of Economics*, September 2011, *3* (1), 363–394.

_ **and Robert A. Miller**, "Conditional Choice Probability Estimation of Dynamic Discrete Choice Models with Unobserved Heterogeneity," *Econometrica*, November 2011, *79* (6), 1823–1867.

_ **and** _ , "Identifying Dynamic Discrete Choice Models Off Short Panels," mineo Carnegie Mellon University, June 2015.

Bellman, Richard E., *Dynamic Programing*, Princeton, NJ: Princeton University Press, 1957.

Ben-Porath, Yoram, "The Production of Human Capital and the Life Cycle of Earnings," *Journal of Political Economy*, August 1967, *75* (4), 352–365.

Borjas, George J., "Immigration Policy, National Origin, and Immigrant Skills: A Comparison of Canada and the United States," in David E. Card and Richard B. Freeman, eds., *Small Differences that Matter: Labor Markets and Income Maintenance in Canada and the United States*, Chicago: The University of Chicago Press, 1993, chapter 1, pp. 21–43.

_ , "Assimilation and Changes in Cohort Quality Revisited: What Happened to Immigrant Earnings in the 1980s?," *Journal of Labor Economics*, April 1995, *13* (2), 201–245.

_ , "The Labor Demand Curve Is Downward Sloping: Reexamining the Impact of Immigration on the Labor Market," *Quarterly Journal of Economics*, November 2003, *118* (4), 1335–1374.

_ , "The Impact of Foreign Students on the Earnings of Doctorates," in Richard B. Freeman and Daniel Goroff, eds., *Science and Engineering Careers in the United States: An Analysis of Markets and Employment*, Chicago: The University of Chicago Press, 2009, chapter 4, pp. 131–161.

_ **and Kirk B. Doran**, "The Collapse of the Soviet Union and the Productivity of American Mathematicians," *Quarterly Journal of Economics*, April 2012, *127* (3), 1143–1203.

_ **and** _ , "Which Peers Matter? The Relative Impacts of Collaborators, Colleagues, and Competitors," *Review of Economics and Statistics*, December 2015, *97* (5), 1104–1117.

_ **and** _ , "Cognitive Mobility: Labor Market Responses to Supply Shocks in the Space of Ideas," *Journal of Labor Economics*, July 2015, *33* (S1, Part 2), S109–S145.

Bound, John, Breno Braga, Joseph M. Golden, and Gaurav Khanna, "Recruitment of Foreigners in the Market for Computer Scientists in the United States," *Journal of Labor Economics*, July 2015, *33* (S1, Part 2), S187–S223.

— **, Gaurav Khanna, and Nicolas Morales**, "Understanding the Economic Impact of the H-1B Program on the U.S.," in Gordon H. Hanson, William R. Kerr, and Sarah Turner, eds., *High-Skilled Migration to the United States and its Economic Consequences*, Chicago: The University of Chicago Press, 2017, chapter 4.

Card, David E., "Immigration and Inequality," *American Economic Review: Papers and Proceedings*, May 2009, *99* (2), 1–21.

Das, Mitali, Whitney K. Newey, and Francis Vella, "Nonparametric Estimation of Sample Selection Models," *Review of Economic Studies*, January 2003, *80* (1), 33–58.

Doran, Kirk B., Alexander Gelber, and Adam Isen, "The Effects of High-Skilled Immigration Policy on Firms: Evidence from Visa Lotteries," NBER Working Paper No. 20668, February 2016.

Dustmann, Christian, Tommaso Frattini, and Ian Preston, "The Effect of Immigration along the Distribution of Wages," *Review of Economic Studies*, January 2013, *80* (1), 145–173.

— **, Uta Schönberg, and Jan Stuhler**, "The Impact of Immigration: Why Do Studies Reach Such Different Results?," *Journal of Economic Perspectives*, Fall 2016, *30* (4), 31–56.

Ewing, Walter A., "Opportunity and Exclusion: A Brief History of U.S. Immigration Policy," Immigration Policy Center, American Immigration Council, January 2012.

Gaulé, Patrick and Mario Piacentini, "Chinese Graduate Students and U.S. Scientific Productivity," *Review of Economics and Statistics*, May 2013, *95* (2), 698–701.

Heckman, James J., Lance Lochner, and Christopher Taber, "Explaining Rising Wage Inequality: Explorations with a Dynamic General Equilibrium Model of Labor Earnings with Heterogeneous Agents," *Review of Economic Dynamics*, January 1998, *1* (1), 1–58.

Hotz, V. Joseph and Robert A. Miller, "Conditional Choice Probabilities and the Estimation of Dynamic Models," *Review of Economic Studies*, July 1993, *60* (3), 497–529.

— **,** — **, Seth Sanders, and Jeffrey Smith**, "A Simulation Estimator for Dynamic Models of Discrete Choice," *Review of Economic Studies*, April 1994, *61* (2), 265–289.

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

_ , "Are Immigrants the Most Skilled U.S. Computer and Engineering Workers?," *Journal of Labor Economics*, July 2015, *33* (S1), S39–S77.

_ **and Marjolaine Gauthier-Loiselle**, "How Much Does Immigration Boost Innovation?," *American Economic Journal: Macroeconomics*, April 2010, *2* (2), 31–56.

Kato, Takao and Chad Sparber, "Quotas and Quality: The Effect of H-1B Visa Restrictions on the Pool of Prospective Undergraduate Students from Abroad," *Review of Economics and Statistics*, March 2013, *95* (1), 109–126.

Keane, Michael P. and Kenneth I. Wolpin, "The Solution and Estimation of Discrete Choice Dynamic Programming Models by Simulation and Interpolation: Monte Carlo Evidence," *Review of Economics and Statistics*, November 1994, *76* (4), 648–672.

_ **and** _ , "The Career Decisions of Young Men," *Journal of Political Economy*, June 1997, *105* (3), 473–522.

Kerr, Sari Pekkala and William R. Kerr, "Immigration and Employer Transitions for STEM Workers," *American Economic Review: Papers and Proceedings*, May 2013, *103* (3), 193–197.

_ **,** _ **, and William F. Lincoln**, "Skilled Immigration and the Employment Structures of US Firms," *Journal of Labor Economics*, July 2015, *33* (S1, Part 2), S147–S186.

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

Klenow, Peter J. and Andrés Rodríguez-Clare, "Externalities and Growth," in Philippe Aghion and Steven N. Durlauf, eds., *Handbook of Economic Growth*, Vol. 1A, Amsterdam: North-Holland Publishing Company, 2005, chapter 11, pp. 817–861.

Koh, Dongya, Raül Santaeulàlia-Llopis, and Yu Zheng, "Labor Share Decline and Intellectual Property Products Capital," mimeo, February 2016.

Krusell, Per and Anthony A. Smith Jr., "Income and Wealth Heterogeneity in the Macroeconomy," *Journal of Political Economy*, October 1998, *106* (5), 867–896.

— , **Lee E. Ohanian, José-Víctor Ríos-Rull, and Giovanni L. Violante**, "Capital-Skill Complementarity and Inequality: A Macroeconomic Analysis," *Econometrica*, September 2000, *68* (5), 1029–1054.

LaLonde, Robert J. and Robert H. Topel, "The Assimilation of Immigrants in the U.S. Labor Market," in George J. Borjas and Richard B. Freeman, eds., *Immigration and the Work Force: Economic Consequences for the United States and Source Areas*, Chicago: The University of Chicago Press, 1992, chapter 3, pp. 67–92.

Lee, Donghoon, "An Estimable Dynamic General Equilibrium Model of Work, Schooling, and Occupational Choice," *International Economic Review*, February 2005, *46* (1), 1–34.

— **and Kenneth I. Wolpin**, "Intersectoral Labor Mobility and the Growth of the Service Sector," *Econometrica*, January 2006, *74* (1), 1–46.

— **and** — , "Accounting for Wage and Employment Changes in the US from 1968-2000: A Dynamic Model of Labor Market Equilibrium," *Journal of Econometrics*, May 2010, *156* (1), 68–85.

Lee, Lung-Fei, "Generalized Econometric Models with Selectivity," *Econometrica*, March 1983, *51* (2), 507–512.

Lewis, Ethan G., "Immigration, Skill Mix, and Capital-Skill Complementarity," *Quarterly Journal of Economics*, May 2011, *126* (2), 1029–1069.

— , "Immigration and Production Technology," *Annual Review of Economics*, April 2013, *5* (1), 165–191.

Llull, Joan, "Immigration, Wages, and Education: A Labor Market Equilibrium Structural Model," *Review of Economic Studies*, July 2018, *85* (3), 1852–1896.

— , "Online Supplement to "Immigration, Wages, and Education: a Labor Market Equilibrium Structural Model"," 2018.

— , "The Effect of Immigration on Wages: Exploiting Exogenous Variation at the National Level," *Journal of Human Resources*, Summer 2018, *53* (3), 608–622.

Lucas, Robert E., "On the Mechanics of Economic Development," *Journal of Monetary Economics*, July 1988, *22* (1), 3–42.

Ma, Jie, "High Skilled Immigration and the Market for Skilled Labor: The Role of Occupational Choice," mimeo, Colgate University, October 2017.

Magnac, Thierry and David Thesmar, "Identifying Dynamic Discrete Decision Processes," *Econometrica*, March 2002, *70* (2), 801–816.

McGrattan, Ellen R. and Edward C. Prescott, "Unmeasured Investment and the Puzzling US Boom in the 1990s," *American Economic Journal: Macroeconomics*, October 2010, *2* (4), 88–123.

_ **and** _ , "A Reassessment of Real Business Cycle Theory," *American Economic Review: Papers and Proceedings*, May 2014, *104* (5), 177–182.

Mincer, Jacob, *Schooling, Experience and Earnings*, New York: Columbia University Press, 1974.

Moser, Petra, Alessandra Voena, and Fabian Waldinger, "German Jewish Émigrés and US Invention," *American Economic Review*, October 2014, *104* (10), 3222–3255.

Olley, G. Steven and Ariel Pakes, "The Dynamics of Productivity in the Telecommunications Equipment Industry," *Econometrica*, November 1996, *65* (6), 1263–1297.

Orrenius, Pia M. and Madeline Zavodny, "Does Immigration Affect Whether US Natives Major in Science and Engineering?," *Journal of Labor Economics*, July 2015, *33* (S1, Part 2), S79–S108.

Ottaviano, Gianmarco I. P. and Giovanni Peri, "Rethinking the Effect of Immigration on Wages," *Journal of the European Economic Association*, February 2012, *10* (1), 152–197.

Peri, Giovanni, Kevin Shih, and Chad Sparber, "STEM Workers, H1B Visas, and Productivity in US Cities," *Journal of Labor Economics*, July 2015, *33* (S1, Part 2), S225–S255.

Romer, Paul M., "Increasing Returns and Long-Run Growth," *Journal of Political Economy*, October 1986, *94* (5), 1002–1037.

Rust, John, "Structural Estimation of Markov Decision Processes," in Robert F. Engle and Daniel L. McFadden, eds., *Handbook of Econometrics*, Vol. 4, Amsterdam: North-Holland Publishing Company, 1994, chapter 51, pp. 3081–3143.

Vella, Francis, "Estimating Models with Sample Selection Bias: A Survey," *Journal of Human Resources*, Winter 1998, *33* (1), 127–169.

Appendix A: Sample Selection and Variable Definitions

A1. Aggregate data

Aggregate macro data are used in the solution of the model, as described in the main text. The estimation period is 19XX-20XX. However, in order to eliminate the influence of initial conditions, I simulate the model starting in 1860. The model is initialized by simulating the first 40 years (1860-1900) using aggregate data for 1900. Then I simulate the remaining years (1900-20XX) with actual macro data. As a result, two entire generations go by before the first year of estimation.

Output. Output is measured as Gross Domestic Product at chain 2009 U.S. dollars, provided by the Bureau of Economic Analysis (BEA), NIPA Table 1.1.6. Given that the original series starts in 1929, I use the average annual growth rate (1929-2016) to extrapolate backwards to year 1900.

Capital stock. There are three types of capital in the model: structures, equipment and intellectual property products capital. The three series are extracted from BEA, combining flow data from Fixed Assets Tables 1.2 ("Chain-Type Quantity Indexes for Net Stock of Fixed Assets") with year 2009 stock data from Fixed Assets Table 1.1 ("Current Cost Net Stock of Fixed Assets"). Resulting series are expressed in chain 2009 U.S. dollars. Series start in 1925, so I extrapolate them backwards to 1900 using average annual growth rates. The capital depreciation rates are assumed to be 11.93% and 2.88% for equipment and structures respectively, based on mean lives of 14 and 36 years and declining-balance rates of 1.67 and 0.92.[29]

Cohort sizes. Cohort sizes are extracted from Integrated Public Use Microdata Series (IPUMS) of the U.S. Census.In particular, I use information from the decennial Censuses from 1900 to 2000, and from the American Community Survey (ACS) 2001-2015. A person is classified as an immigrant if born abroad; individuals born in Puerto Rico and other outlying areas are categorized as natives. Native and immigrant inter-census cohort sizes are estimated following different procedures. For natives, I distribute the cumulative decade cohort size decrease to each year using annual data on mortality rates by age from Vital Statistics of the U.S. (National Center for Health Statistics). For immigrants, I use a simi-

[29] These mean lives and declining-balance rates are obtained averaging the service life and declining-balance rates provided by the Bureau of Economic Analysis ("Table C. BEA Rates of Depreciation, Service Lives, Declining-Balance Rates, and Hulten-Wykoff categories").

lar procedure, using the estimates of the entry age distribution described below instead of mortality rates.

Age at entry. The distribution of entry age of immigrants is estimated using U.S. Census IPUMS. In order to reduce small sample noise, I average out the distributions for immigrants who arrived at $t-1$, $t-2$,..., $t-5$. Since the exact year of immigration is only available in 1900-1930 and 2000 Censuses, and in the ACS (2001-2015), intermediate years are linearly interpolated. Given that the distribution is stable over the years, I estimate a single distribution for each of the following intervals: 1900-1930, 1931-1940, 1941-1950, 1951-1960, 1961-1970, 1971-1980, 1981-1990 and 1991-2015. Finally, in order to obtain the joint distribution of age at entry and initial education, I estimate the entry age distribution conditional on education. Because of data limitations, I approximate it using the "relative" distribution by educational level, i.e. I compute the ratio of conditional and unconditional distributions from the Census 2000, and then I multiply this relative distribution with the time varying unconditional age at entry distribution.[30]

Regions of origin. I consider seven regions of origin for immigrants: Western Countries, Mexico, Central America & Caribbean, South America, China & selected Southeast Asia, India, Other Asia & Africa includes all the remaining immigrants from these two continents. The stock of immigrants from each of these regions are drawn from U.S. Census IPUMS 1900-2000 and ACS 2001-2015. Inter-census estimates of the stock of immigrants from each region of origin are obtained by combining a linear interpolation of the share of immigrants from each region and the estimated of cohort sizes described above.

A2. Microdata

All micro-data statistics used in the estimation are constructed with data from the March Supplement of Current Population Survey (CPS).[31]

[30] This calculation assumes that the relative distribution is constant over time. Estimates using 1970-1990 Censuses (for which the year of entry is only available by five-year intervals) support this assumption.

[31] CPS data are extracted from IPUMS (Ruggles et al., 2008). The CPS interviews households for 8 eight months: when a household enters the sample for the first time is interviewed four consecutive months, then not interviewed during eight months, and finally interviewed four additional consecutive months (the same four calendar months than in the first spell). Thus, a household in the March sample is interviewed in March for two consecutive years. In most of the survey years it is, hence, possible to match a subset of households for two consecutive years obtaining a small panel. IPUMS data has a recoded individual and household identifier that does not allow to match consecutive surveys. I use samples extracted from the NBER to do the matching. Survey years 1971-72, 1972-73, 1976-77, 1985-86 and 1995-96 can not be matched.

Age. Individuals above 65 and below 16 are not in the model and they are dropped from the samples.

Educational level. I categorize individuals in four education groups: high school dropouts (< 12 years of education), high school graduates (12), persons with some college ($13-15$), and college graduates (16+). In 1992, a methodological change was introduced to CPS regarding education.

Choices. Individuals are assigned to one of the five mutually exclusive year round alternatives: blue collar, white collar and white collar-STEM work, attend school, or stay at home. The procedure to assign individuals follows a hierarchical rule. An individual is assigned to school if she reported that school was her main activity during de survey week (CPS) or if she was attending school at survey date. She is assigned to work in either of the three occupations if she is not assigned to school and she worked at least 40 weeks during the year before the survey date, and at least 20 hours per week.[32] When an individual is assigned to work, she is assigned to the occupation held during the last year (CPS). Blue collar occupations include craftsmen, operatives, service workers, laborers, and farmers, white collar include professionals, clerks, sales workers, managers, and farm managerial occupations, and white collar-STEM include some of the previous occupation for those with at least a degree. Finally, those individuals that are not assigned neither to work nor to attend school are assigned to stay at home.

Wages. Hourly wage is computed for individuals that are assigned to either of the work alternatives according to the previous definition. Workers are assumed to earn their wage entirely in the occupation they are assigned to. Earnings include wage and salary income, and self-employment earnings, deflated to year 2010 U.S.$ using the Consumer Price Index.

Preschool children. Individuals are allowed to have 0,1, or 2+ preschool children (less than five years old). In the data, households are defined as family units; preschool children living in a two family home are only assigned to their parents. In order to link children with their parents, I use IPUMS-created variables momloc and poploc, which identify the position of the mother and father in the household respectively. Parent definition includes biological, step- and adoptive parents. Although they fully comparable over years, there are some minor changes that are listed in the database documentation.

[32]Hours per week are approximated by the number of hours worked in the previous week.

Region of origin. The region of birth is assigned as described above for the aggregate data. A small number of individuals for which the country of birth is unknown are dropped from the corresponding samples. CPS started to ask questions related to immigrant status in survey year 1994. Therefore, statistics that include this information are only used from that survey date onwards.

Years in the U.S. In particular, this variable is defined as age at entry minus the difference between the survey year and years of immigration.

Appendix B: Simulation of Idiosyncratic Shocks

This appendix describes how to simulate idiosyncratic errors ε_a from the distribution described in (13) along with productivity shocks η_a and taste shocks ϵ_{ja} for $j \in \{B, W, T, S, H\}$. Since $\varepsilon_{Sa} = \epsilon_{Sa}$ and $\varepsilon_{Ha} = \epsilon_{Ha}$ are independent of the other shocks and of η_a, they are trivially obtained transforming uniform draws by the quantile function of the Type-I extreme value distribution. That is, let q_S and q_H denote two independent draws from a standard uniform, $\mathbb{U}(0, 1)$, then the corresponding draws of ε_S and ε_H are obtained as:

$$\varepsilon_j = -\ln(-\ln q_j), \quad \text{for } j \in \{S, H\}. \tag{B1}$$

To simulate draws for ε_j for $j \in \{B, W, H\}$, I first derive the marginal distribution of ε_B, the conditional distribution of ε_W given ε_B, and the one of ε_T conditional on the other two. Let $f_{\varepsilon_B \varepsilon_W \varepsilon_T}(\varepsilon_B, \varepsilon_W, \varepsilon_T)$ denote the joint probability density function of ε_B, ε_W, and ε_T, and let $F_{\varepsilon_B \varepsilon_W \varepsilon_T}(\varepsilon_B, \varepsilon_W, \varepsilon_T)$ denote their cumulative distribution function. Equation (13) implies that:

$$\begin{aligned} F_{\varepsilon_B \varepsilon_W \varepsilon_T}(\varepsilon_B, \varepsilon_W, \varepsilon_T) &= \exp\left\{ -\left(e^{-\varepsilon_B/\varrho} + e^{-\varepsilon_W/\varrho} + e^{-\varepsilon_T/\varrho}\right)^{\varrho} \right\} \\ &\equiv \exp\left\{ -\mathcal{Q}(\varepsilon_B, \varepsilon_W, \varepsilon_T)^{\varrho} \right\}, \end{aligned} \tag{B2}$$

and:

$$
\begin{aligned}
&f_{\varepsilon_B\varepsilon_W\varepsilon_T}(\varepsilon_B,\varepsilon_W,\varepsilon_T)\\
&\quad= \frac{\partial^3}{\partial\varepsilon_B\partial\varepsilon_W\partial\varepsilon_T}\exp\{-\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}\}\\
&\quad= \frac{\partial^2}{\partial\varepsilon_W\partial\varepsilon_T}\exp\left\{-\left(\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}+\frac{\varepsilon_B}{\varrho}\right)\right\}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{(\varrho-1)}\\
&\quad= \frac{\partial}{\partial\varepsilon_T}\left[\begin{array}{l}\exp\left\{-\left(\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}+\frac{\varepsilon_B+\varepsilon_W}{\varrho}\right)\right\}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{2(\varrho-1)}\\ \times\left(1-\frac{\varrho-1}{\varrho}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{-\varrho}\right)\end{array}\right]\\
&\quad= \exp\left\{-\left(\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}+\frac{\varepsilon_B+\varepsilon_W+\varepsilon_T}{\varrho}\right)\right\}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{3(\varrho-1)}\\
&\qquad\times\left(1-\frac{\varrho-1}{\varrho}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{-\varrho}\left[3-\frac{\varrho-2}{\varrho}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{-\varrho}\right]\right)
\end{aligned}
\tag{B3}
$$

Given this expression, the marginal density function for ε_B is given by:

$$
\begin{aligned}
f_{\varepsilon_B}(\varepsilon_B) &\equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f_{\varepsilon_B\varepsilon_W\varepsilon_T}(\varepsilon_B,\varepsilon_W,\varepsilon_T)d\varepsilon_W d\varepsilon_T\\
&= \exp\left\{-\left(\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}+\frac{\varepsilon_B}{\varrho}\right)\right\}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{(\varrho-1)}\Big]_{-\infty}^{+\infty}\Big]_{-\infty}^{+\infty}\\
&= \exp\{-(e^{-\varepsilon_B}+\varepsilon_B)\},
\end{aligned}
\tag{B4}
$$

which is the probability density function of a Type-I Extreme Value distribution. Thus, ε_B is drawn using the quantile function in (B1). Having drawn it, we now need to draw from the conditional distribution of ε_W given ε_B. To compute it, we first need to derive the joint distribution of ε_B and ε_W, which is given by:

$$
\begin{aligned}
f_{\varepsilon_B\varepsilon_W}(\varepsilon_B,\varepsilon_W) &\equiv \int_{-\infty}^{\infty} f_{\varepsilon_B\varepsilon_W\varepsilon_T}(\varepsilon_B,\varepsilon_W,\varepsilon_T)d\varepsilon_T\\
&= \left[\begin{array}{l}\exp\left\{-\left(\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}+\frac{\varepsilon_B+\varepsilon_W}{\varrho}\right)\right\}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{2(\varrho-1)}\\ \times\left(1-\frac{\varrho-1}{\varrho}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{-\varrho}\right)\end{array}\right]_{-\infty}^{+\infty}\\
&= \exp\left\{-\left[\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho}+\frac{\varepsilon_B+\varepsilon_W}{\varrho}\right]\right\}\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho-2}\\
&\quad\times\left(\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho}+\frac{1-\varrho}{\varrho}\right),
\end{aligned}
\tag{B5}
$$

where $\mathcal{Q}_2(\varepsilon_B,\varepsilon_W) \equiv \exp(-\varepsilon_B/\varrho)+\exp(-\varepsilon_W/\varrho)$. Thus, the conditional density is:

$$\begin{aligned}
f_{\varepsilon_W|\varepsilon_B}(\varepsilon_W|\varepsilon_B) &\equiv \frac{f_{\varepsilon_B\varepsilon_W}(\varepsilon_B,\varepsilon_W)}{f_{\varepsilon_B}(\varepsilon_B)} \\
&= \exp\left\{e^{-\varepsilon_B}+\varepsilon_B-\left[\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho}+\frac{\varepsilon_B+\varepsilon_W}{\varrho}\right]\right\} \\
&\quad\times \mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho-2}\left(\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho}+\frac{1-\varrho}{\varrho}\right). \qquad \text{(B6)}
\end{aligned}$$

The conditional cumulative function is obtained integrating the above expression with respect to ε_W, which yields:

$$\begin{aligned}
F_{\varepsilon_W|\varepsilon_B}(\varepsilon_W|\varepsilon_B) &\equiv \int_{-\infty}^{\varepsilon_W} f_{\varepsilon_W|\varepsilon_B}(\varepsilon|\varepsilon_B)d\varepsilon \qquad \text{(B7)}\\
&= \exp\left\{e^{-\varepsilon_B}+\varepsilon_B-\left[\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho}+\frac{\varepsilon_B}{\varrho}\right]\right\}\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho-1}.
\end{aligned}$$

Inverting this function with respect to ε_W to obtain the conditional quantile function, and evaluating it on a uniform random draw q_W, we simulate ε_W as:

$$\varepsilon_W = -\varrho\ln\left\{-e^{-\varepsilon_B/\varrho}+\left[\frac{1-\varrho}{\varrho}\mathbb{W}\left(\frac{\varrho}{1-\varrho}\exp\left\{\tfrac{\varrho}{1-\varrho}e^{-\varepsilon_B}-\varepsilon_B\right\}q_W^{\varrho/(\varrho-1)}\right)\right]^{1/\varrho}\right\}, \qquad \text{(B8)}$$

where $\mathbb{W}(\cdot)$ is the Lambert-W function or product logarithm function, which is defined as the inverse of the function $f(x) = xe^x$ (or equivalently as $\mathbb{W}(y)e^{\mathbb{W}(y)} = y$).

Finally, the conditional distribution of ε_T given ε_B and ε_W is given by:

$$\begin{aligned}
&f_{\varepsilon_T|\varepsilon_B\varepsilon_W}(\varepsilon_T|\varepsilon_B,\varepsilon_W) \\
&\equiv \frac{f_{\varepsilon_B\varepsilon_W\varepsilon_T}(\varepsilon_B,\varepsilon_W,\varepsilon_T)}{f_{\varepsilon_B\varepsilon_W}(\varepsilon_B,\varepsilon_W)} \\
&\quad= \exp\left\{\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho}-\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}-\frac{\varepsilon_T}{\varrho}\right\}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{3(\varrho-1)} \\
&\qquad\times\left(1-\frac{\varrho-1}{\varrho}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{-\varrho}\left[3-\frac{\varrho-2}{\varrho}\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{-\varrho}\right]\right) \\
&\qquad\times\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{2(\varrho-1)}\left(1-\frac{\varrho-1}{\varrho}\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{-\varrho}\right)^{-1}. \qquad \text{(B9)}
\end{aligned}$$

Integrating this expression with respect to ε_T we obtain:

$$\begin{aligned}
F_{\varepsilon_T|\varepsilon_B,\varepsilon_W}(\varepsilon_T|\varepsilon_B,\varepsilon_W) &\equiv \int_{-\infty}^{\varepsilon_T} f_{\varepsilon_T|\varepsilon_B\varepsilon_W}(\varepsilon|\varepsilon_B,\varepsilon_B)d\varepsilon \\
&= \exp\left\{\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{\varrho}-\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{\varrho}\right\} \qquad \text{(B10)}\\
&\quad\times\frac{(1-\varrho)\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{2\varrho-3}+\varrho\mathcal{Q}(\varepsilon_B,\varepsilon_W,\varepsilon_T)^{3(\varrho-1)}}{(1-\varrho)\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{2\varrho-3}+\varrho\mathcal{Q}_2(\varepsilon_B,\varepsilon_W)^{3(\varrho-1)}}.
\end{aligned}$$

Finally, the conditional quantile function, which does not have a closed form solution, is used to transform uniform draws into draws from $F_{\varepsilon_T|\varepsilon_B\varepsilon_W}(\varepsilon_T|\varepsilon_B, \varepsilon_W)$.

All the steps so far show how to simulate ε_B, ε_W, and ε_B from its joint distribution. A prior step is needed to complete the simulation. In particular, we need to independently draw ϵ_B from its unknown distribution and $\sigma_{B\ell}\eta_a$ from a normal distribution, and then obtain ε_B as the sum of the two (which by construction will make it a Type-I extreme value). Having done that, we should then proceed simulating ε_{Wa} and ε_{Ta} as described in this Appendix. And finally, the remaining taste shocks are obtained as functions of these draws as $\epsilon_{Wa} = \varepsilon_{Wa} - \sigma_{W\ell}\eta_a$ and $\epsilon_{Ta} = \varepsilon_{Ta} - \sigma_{T\ell}\eta_a$.

The remaining unknown is the distribution of the taste shock ϵ_{Ba}, $f_{\epsilon_B}(\epsilon_B)$. This distribution is obtained as the deconvolution of a Type-I Extreme Value and a normal distribution with zero mean and variance $\sigma_{B\ell}$. This deconvolution is derived using the characteristic functions of the two distributions which are, respectively, $\mathbb{G}(1 - it)$ and $e^{-\frac{t^2}{2}\sigma^2_{B\ell}}$, where $\mathbb{G}(\cdot)$ denotes the complete gamma function, $\mathbb{G}(x) \equiv \int_0^\infty z^{x-z}e^{-z}dz$, and i denotes the imaginary unit. In particular:

$$f_{\epsilon_B}(\epsilon_B) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \mathbb{G}(1 - it)e^{-\frac{t^2}{2}\sigma^2_{B\ell} - it\epsilon_B}dt, \tag{B11}$$

which does not have a closed form. The cumulative density function and the quantile function are then computed numerically.