Selective Immigration Policies and the U.S. Labor Market

Joan Llull*§
Institute of Economic Analysis (IAE-CSIC)
and Barcelona School of Economics

Preliminary and Incomplete
June 2024

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-United States border, and it has been central in several (some of them successful)

* IAE-CSIC. Campus de Bellaterra, 08193, Bellaterra, Cerdanyola del Vallès, Barcelona (Spain). URL: https://joanllull.github.io. E-mail: joan.llull [at] iae.csic [dot] es.

§ I would like to thank Manuel Arellano, Jeff Campbell, Jinyong Hahn, Ellen McGrattan, Martí Mestieri, Robert Miller, José Víctor Ríos-Rull, Luis Rojas, Raül Santaeulàlia-Llopis, Elena T. Aguilar, and a large number of conference and seminar participants for enlightening comments and discussions. Daniele Alimonti, Maja Garbulinska, Rim Mesraoua, Minyoung Rho, Anqi Shan, and Jacek Barszczewski 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, Grant PID2021-122605NB-I00 funded by MCIN/AEI/10.13039/501100011033 and by ERDF A way of making Europe, 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 (2021-SGR-00571).

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 basis of national origin in order to preserve the ethnic mix of the U.S. population. The removal

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 **?**, 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 **?**, 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 in many dimensions, including, in the case of immigrants, national origin, which

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 (**???**).

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 years 1994–2020 (March Supplements, linked over two consecutive years) and the Survey of Income and Program Participation (SIPP) for years 1987–2007, 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 **???**) is computationally too demanding. Alternatively, I estimate the model using conditional choice probability (CCP) estimation, combining techniques and arguments developed by **?**, **?**, **?**, **?**, and **?** (see **?** 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.

? 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. **?** 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 (**???**); the displacement effect on native STEM employment and wages (**????????**); the career prospects of science and technology immigrants relative to similar natives (**??**); the crowding out effects of H-1B visas on native students (**??**); and the competition and spillover effects in the space of ideas, exploiting evidence on massive inflows of foreign professors (**???**, and **?**).

Among these strands of the literature, my paper is mostly related to the first two.[5] Except for **?**, **?**, and **?**, all these papers use a reduced form approach. The structural approach allows me to 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 **?** and **?** 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

[4] More generally, it also contributes to the general literature on immigration on wages (**?**; **?**; **?**; **?**; **??**; see **?** 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.

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 ??, and ?), 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.

? 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. ? 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. ? 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 (??????). Moreover, it contributes to the literatures that analyze skill-biased technical change and capital-skill complementarity (e.g., ?; see ?? for surveys), and knowledge spillovers in aggregate production (??; see ? 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., ???).

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 (**?**). 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 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 (**?**).

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,

[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.

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 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

[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 (?). 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.

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 (**?**). 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.

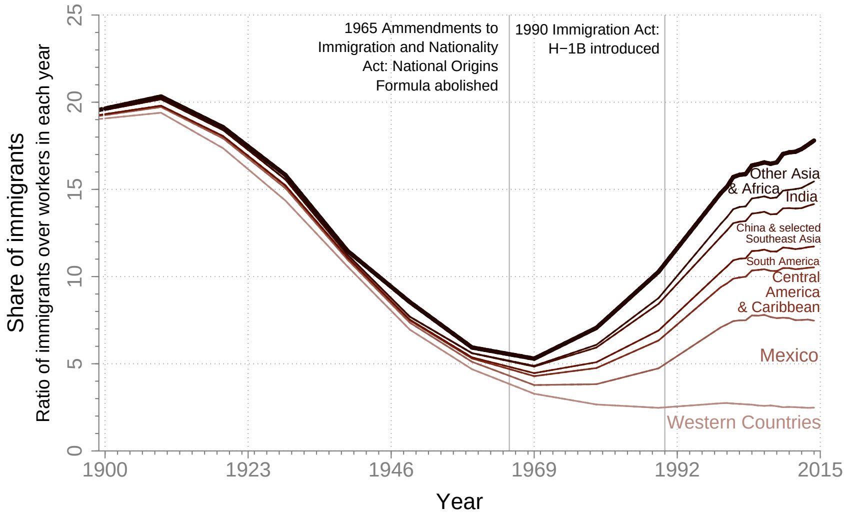
B. Ethnic and Skill Composition of Immigration

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, 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

[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" (**?**). 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.

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).

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 disaggregated by national origin, the average education level of immigrants from each

[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.

Table 1—Education of Natives and Immigrants

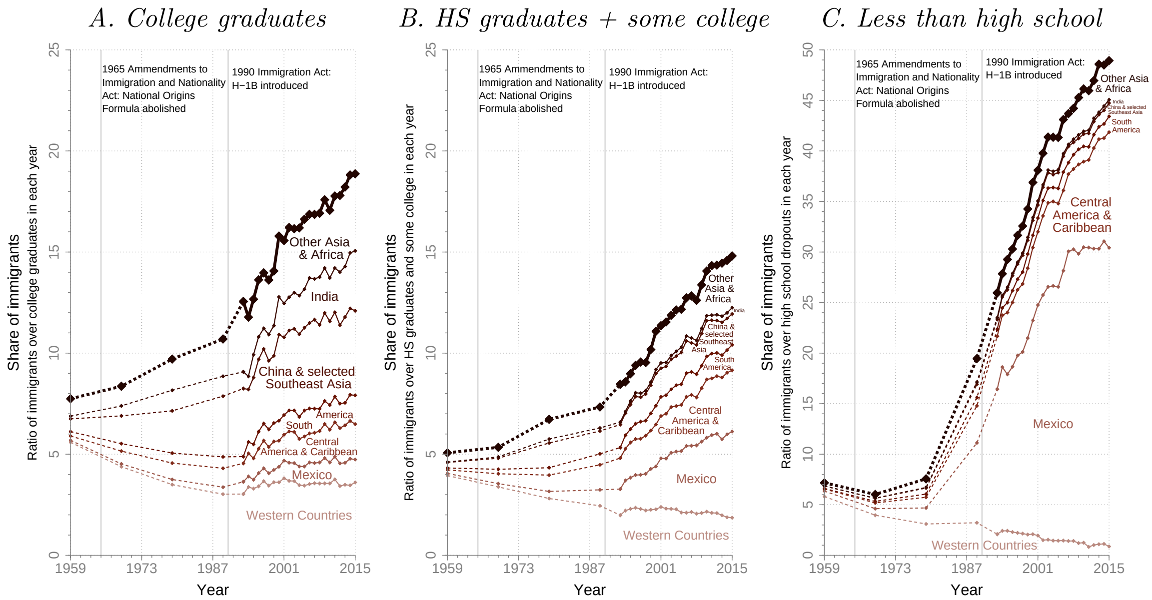
	1970	1980	1990	2000	2010	2015
A. Average years of education:						
Natives	11.2	12.3	13.0	13.7	13.7	13.8
Immigrants	11.0	11.7	12.1	13.1	13.1	13.3
Western countries	10.5	11.8	12.8	14.0	14.3	14.6
Mexico	6.3	7.1	7.6	9.1	9.3	9.6
Central America & Caribbean	10.2	10.9	11.0	11.9	11.5	11.5
South America	11.5	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).

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

[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.

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



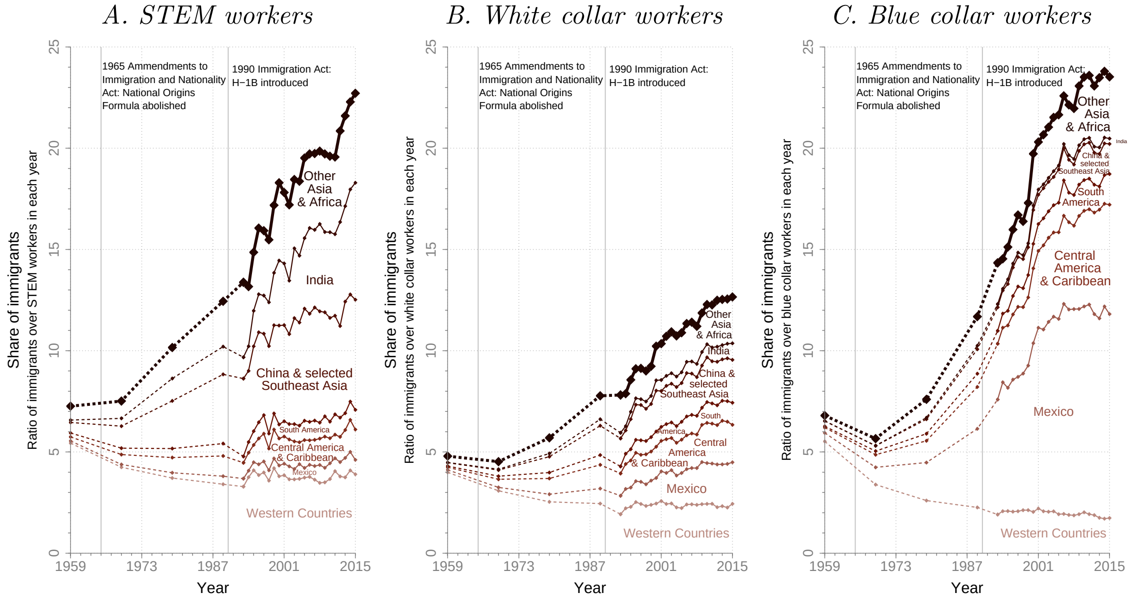
Note: Areas delimited by plotted lines show the share that immigrants from each national origin represent of all individuals aged over 24 with the indicated education. Country grouping and inter-census interpolations are described in Appendix A. *Sources:* Census data (1960-1990), and CPS (1994-2016).

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 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

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).

than 5% of the population), and very few of them are executed by Mexicans and 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.

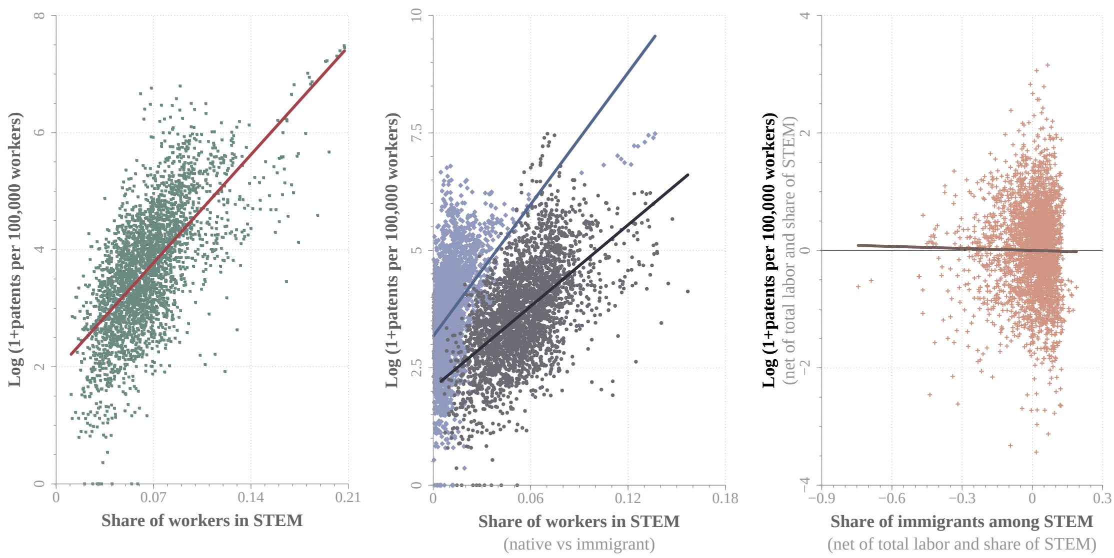
C. *Skilled Labor and the Generation of Ideas*

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

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

strong and positive correlation.

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 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 **?** and **?**. 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

[14] The introduction of intangible capital in the aggregate production function has already been discussed in the macroeconomics literature (e.g. **??**). 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.

(TFP), increase the relative productivity of STEM labor, and/or substitute out blue collar labor. As in **?**, 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 **?** 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 **?** and **?**. 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, **?** 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, **?** finds a "reduced form" elasticity of substitution between natives and immigrants that fits well within the ballpark of estimates in **?**.

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:

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

[15] **??** discusses the importance of capital-skill complementarity in measuring labor market effects of immigration.

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 (**?**) 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 re-

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

spectively, 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 **?**, 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 **?**: 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 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.

[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.

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 $\varrho \equiv \sqrt{1 - \text{Corr}(\varepsilon_{ja}, \varepsilon_{ka})}$ 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:

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

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 $\{r_{t+l}\}_{l\in\{1,\ldots,65-a\}}$. 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 **?**, but differs from **??**, **?**, **??**, or **?**.

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}:

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

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

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

where the terminal value is defined to be zero, $V(h_{65}, \varepsilon_{65}, r_t, \varpi_t) \equiv 0$. Following ?, 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:

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

[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.

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:

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

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 (**?????**). In the immigration context, **?** finds that an autoregressive process in first differences and the contemporaneous change in the aggregate shock can explain 99.9% of the variation in the level of skill prices. Further results presented in **?** 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 **?**, 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.

[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.

C. *Capitalists*

The comparison of results in **?**, **?**, and **?** 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): **?** assumes that capital does not adjust to immigration; in **?**, interest rates do not react (long run small open economy); **?** 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.

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:

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

[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.

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:

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

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

[22] **?** 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.

$S_{jt}^{(S)}(r_t)$, 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, 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 (**????**). 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, $\tilde{\varsigma}_0$, $\tilde{\varsigma}_1$, $\tilde{\alpha}_0$, $\tilde{\alpha}_1$, $\tilde{\theta}_0$, $\tilde{\theta}_1$,

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

[24] **?** 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.

$\tilde{\iota}_0$, $\tilde{\iota}_1$, 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 **?**, **?**, **?**, **?**, **?**, **??**, and **?**.

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.

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 **?** 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 **?** 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 **?**. 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 (**?**):

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

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:

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

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:

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

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:

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

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 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 **?** (with the reformulation in **?**), 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 **?**). Substituting (45) into (22) yields:

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

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 **?**. 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

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

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:

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

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:

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

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, $\hat{\hat{p}}(h_a, t)$, 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, $\hat{p}_S = \hat{\hat{p}}_S = 0$, when $d_{a-1} \neq S$, and the probability of working in STEM is set to zero, $\hat{p}_T = \hat{\hat{p}}_T = 0$, 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 (**?**):

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

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 $\widehat{\ln r_{jt}}$, $\hat{\tilde{s}}_j(h)$, 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.

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:

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

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:

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

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

[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.

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 **?** and **?** 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 $\{r^{(m_i)}_{t_i+l_i}\}_{l_i=1,...,65-a_i}$, 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:

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

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:

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

and:

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

$$\sum_{\ell\in\mathcal{L}} \tau_{0\ell}\,\mathbb{1}\{\ell_i=\ell\} + \tau_1\,\mathbb{1}\{12\le E^{(i)}<16\} + \tau_2\,\mathbb{1}\{E^{(i)}\ge 16\}, \tag{60}$$

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 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)).

? 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 **?** estimator (no iteration) and the full solution estimation, which is obtained

[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.

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 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

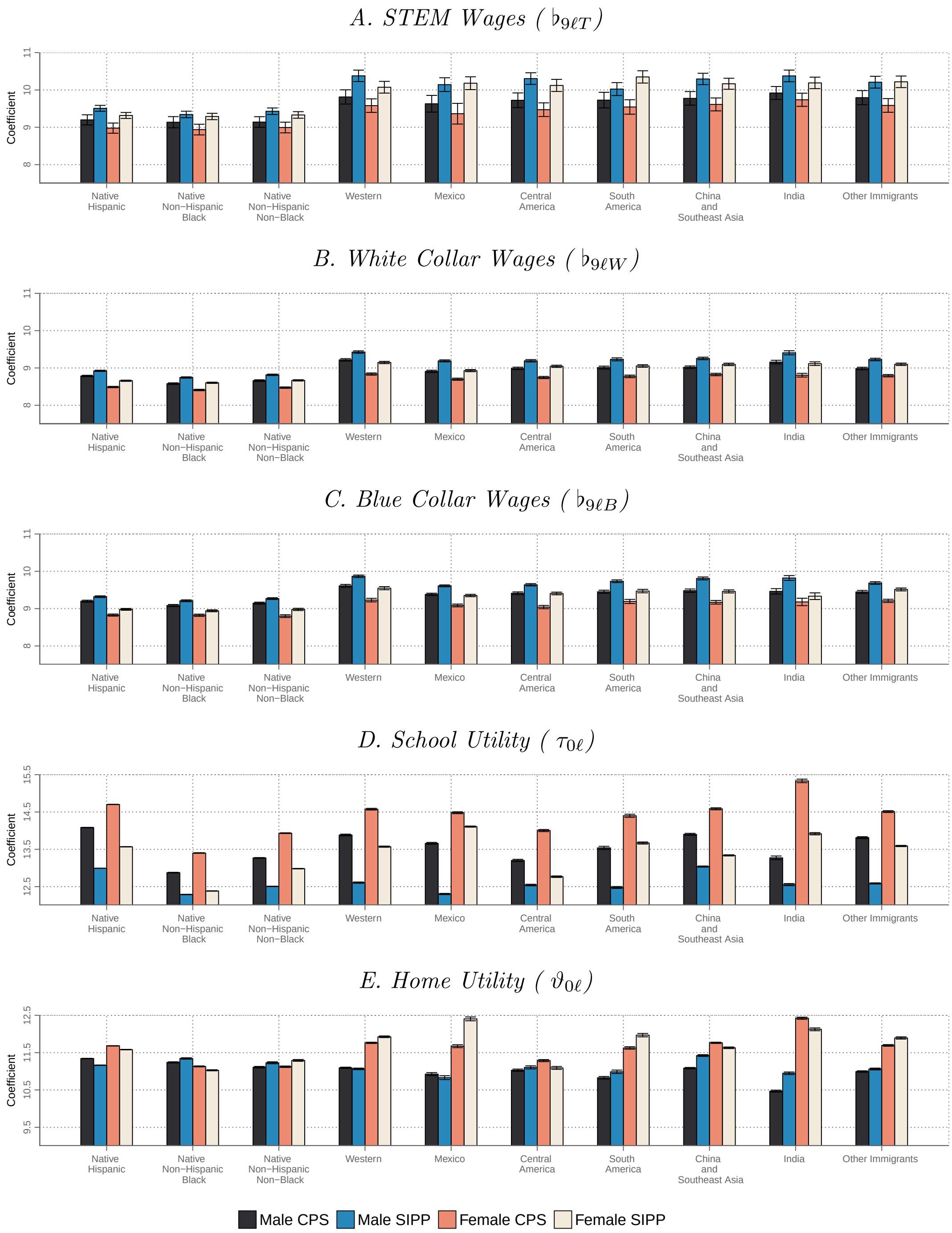
[28] The baseline specifications include the estimation with the CPS (21 observations) and an additional estimation with CPS data extrapolated from 1993 to 1989 using SIPP estimates of skill prices and aggregate skill units (25 observations). Estimation with only SIPP was deemed too imprecise (14 observations). 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).

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 5. 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)

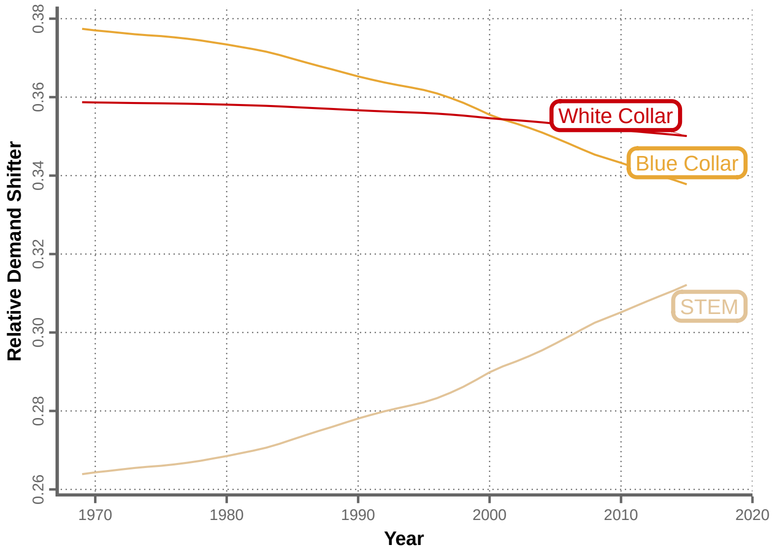
Note: Standard errors (not corrected for estimation error in CCPS and other regressors) in parenthesis.

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 6. 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

Appendix A: Sample Selection and Variable Definitions

The model is estimated using micro data from the Current Population Survey (CPS), and the Survey of Income and Program Participation (SIPP). Additionally, some aggregate macro data are used in the estimation and solution of the model, as described in the main text. While my estimation period is 1987–2020, in my simulations I initialize the model starting in 1860 in order to eliminate the influence of initial conditions.[29] In particular, I simulate the first 40 years, i.e. the period 1860–99, using aggregate data for 1900. Then I simulate the remaining years with actual macro data. As a result, two entire generations go by before the first year of estimation. This appendix provides a detailed description of the main data sources and the main data cleaning procedures.

A1. Aggregate data

Output. Output is measured as Real Gross Domestic Product at chained U.S. dollars of 2012. The raw data is provided by the Bureau of Economic Analysis (BEA) in the National Income and Product Accounts (NIPA), Table 1.1.6 "Real Gross Domestic Product, Chained Dollars" (**?**). Given that the original series starts in 1929, I use the average annual growth rate for the period 1929–2021 to extrapolate backwards to year 1900.

Capital stock. There are three types of capital in the model: structures and equipment, and intellectual property products (IPP) capital, which is my main proxy for the stock of ideas in the economy. The three series are extracted from the Fixed Assets Accounts Tables from the BEA. In all the analysis, I exclude residential assets.[30] For each of the three series, I multiply the chain-type quantity indexes reported in Table 1.2 "Chain-Type Quantity Indexes for Net Stock of Fixed Assets and Consumer Durable Goods" (**?**), which equal 100 in the base year 2012, with the current cost stocks of 2012 obtained from Table 1.1 "Current Cost Net Stock of Fixed Assets" (**?**). In all cases, I take the aggregation of private and government values, reported in rows 18–20 of each of the tables.

[29] In some refinements, my estimation period is 1967–2020.

[30] **?** shows that housing is a major responsible for the documented decreasing in the labor share. The comment by Robert Solow published with the article suggests that an implication of Rognlie's results is that "for estimating an economy-wide elasticity of substitution, it would be better to eliminate the housing stock and associated land on the capital-input side and the rents on the output side." Excluding the part of GDP associated to housing for such a long time series is more complicated: disaggregated in the GDP measures above starting in year 2002. However, not excluding this information from the GDP accounts is not particularly problematic in the context of my paper because wages are deflated by a GDP to CPS/SIPP index, which nets out such aggregate discrepancies in the predicted level of wages by design.

The resulting series are expressed in chained U.S. dollars of 2012. Given that the original series start in 1925, I extrapolate them backwards to 1900 using average annual growth rates.

Capital depreciation rates. The capital depreciation rates are computed combining the Fixed Assets Account Table 1.1 described above with the Table 1.3 "Current-Cost Depreciation of Fixed Assets and Consumer Durable Goods" **?**. In particular, annual depreciation rates are obtained dividing the depreciation series (Table 1.3) for each type of capital by the current cost stock of capital (Table 1.1). Given that the original series start in 1925, I impute the 1925–1929 average to the period 1900–1928.

Cohort sizes. Cohort sizes are extracted from the Integrated Public Use Microdata Series (IPUMS) of the U.S. Census (**?**). In particular, I the largest available sample from each decennial census for 1900–2000, and the American Community Survey (ACS) for 2001–2020. A person is classified as an immigrant if born abroad. Individuals born in Puerto Rico and other outlying areas are categorized as natives. I compute aggregates using person weights provided in the data.[31] Native and immigrant inter-census cohort sizes are estimated following different procedures. For natives, I distribute the decade's cohort size decrease across the years between censuses using annual mortality rates by age from Vital Statistics of the U.S.[32] For immigrants, I use a similar procedure to distribute the net increase in the cohort sizes, but using the estimates of the entry age distribution described below instead of mortality rates.

Age at entry distribution. 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 (2001–2020), 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 1991–2020. Finally, in order to obtain the joint distri-

[31] The 1970 Census includes the six available files, each representative of the whole U.S. Therefore, weights in that year are divided by six to represent the correct U.S. aggregates.

[32] Mortality rates are obtained from different sources: Table 6 in **?** for 1900–1940; Table 55 in **?** for 1941–1960; Tables 1.3 in **?** for 1961–1970 and in **?** for 1971-1980; Table 1.4 in **?**; and Tables 2 in **?**, **?**, **?**, **?**, **?**, and **?** for each year between 1994 and 1999 respectively. The raw data is presented in pdfs, and were manually tabulated. Age is grouped in different categories. For each age, I apply the value of the corresponding aggregated age cell from the original data. In the period 1900–1940, reported rates refer to the subset of U.S. states that reported death rates.

bution 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 compute the ratio of conditional and unconditional distributions from the Census 2000, and then I multiply this relative distribution by the time varying unconditional distribution of age at entry.[33]

Regions of origin distribution. The share of immigrants from each region of origin is obtained from U.S. Census IPUMS 1900–2000 and ACS 2001–2020. I consider seven regions of origin for immigrants: Western Countries; Mexico; Central America and the Caribbean; South America; China and selected Southeast Asia; India; and Other Asia and Africa, which includes all the remaining immigrants from these two continents. Inter-census estimates of the distribution of immigrants from each region of origin are obtained by linear interpolation.

Ethnicity distribution. The share of natives in each ethnic group is obtained from the observed distributions in each U.S. Census IPUMS 1900-2000 and ACS 2001–2020 for individuals aged 16 to 25. I consider three ethnic groups for natives: Hispanic, Non-Hispanic Black, and Non-Hispanic Non-Black. Inter-census estimates of these shares are linearly interpolated.

Initial education distribution. The distribution of initial education is obtained from the U.S. Census IPUMS 1940–2000 and ACS 2001–2020. For natives, I take individuals aged 18 in each census, censoring the number of observed years of education at 10. I focus on 18-years-old individuals instead of 16 to allow for some individuals to temporarily drop out from school, something that is not allowed in the model, and to account for other sources of systematic reporting errors such as those derived from grade retention. The U.S. Census does not contain information on education prior to 1940, so information on earlier cohorts are obtained from the 1940 census, checking, respectively, individuals aged 28, 38, 48, and 58 respectively for years 1930, 1920, 1910, and 1900. For immigrants, I use data starting from 1970, since the 1940–1960 censuses do not include information on years since migration. For each 10-year cohort of arrival of immigrants I compute the observed education distribution, merging the data of different censuses. Because of the grouping of the information of years since migration, the distribution of education of immigrants that arrived before 1940 is assumed to be constant.

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

Fertility process. The fertility process consists of a three times three transition probability matrix, from 0, 1, or 2+ preschool children in period $t-1$ into 0, 1 or 2+ in t, conditional on age, college status, and demographic type. Given the model assumes a stationary distribution for the whole estimation period, I estimate two different processes: one for years before 1970, and and one for the years after 1970 (the transition from the first process to the second is unexpected in the model). The pre-1970 distribution is estimated using Census microdata for census years 1940–1970. The latter is estimated using the censuses of 1980–2000 and the ACS files for 2009–2011 and 2018–2020 to mimic one census point each (the weights of the ACS are divided by three to represent one year's U.S. population). Children are identified using the variables pernum, poploc, and momloc provided by IPUMS, as well as the household identifier. I infer transitions out of cross-sectional data based on the age of the child. In order to smooth probabilities across different ages, I estimate different logits on flexible polynomials of age for every number of children in $t-1$, observable type and college status. Transitions from 0 to 2+ children and from 2+ to 0 are assumed to happen with probability zero.[34] Therefore, the probabilities are estimated using binary logits if the number of children in $t-1$ is 0 or 2+, and a multinomial logit in case it is 1. The logits are specified with up to fourth order polynomials in age. If they do not converge, the order of the polynomial is progressively reduced until convergence is achieved (the lowest specifications are with a quadratic polynomial). The order of the polynomial is also reduced if the largest terms are insignificant. In a few cases, when logits are to be run with less than 100 observations, the probability values are imputed as follows: if the number of children in $t-1$ was 2+, then it reduces to 1 with probability 1/6 and stays at 2+ with probability 5/6; if the number of children at $t-1$ was 1, it reduces to 0 with probability 1/6, it increases to 2+ with a probability that equals the probability that individuals in the same cell have of going from 0 to 1, and it stays at 1 with the remaining probability; if the number of children was 0, it stays at 0 with probability 1. After age 40, the probability of increasing the number of children is not allowed to grow, and, additionally, after age 45 the probability of reducing the number of children is not allowed to decrease.

Wage adjustments. To avoid biases in parameter estimates, I make three adjustments to wages and/or aggregate skill units. On the one hand, CPS and SIPP wages as well as output data include taxes, but individuals make decisions on a net

[34] In practice, when I observe in the data an increase from 0 to 2+, I treat it as if this increase was from 0 to 1, and when a decrease is observed from 2+ to 0, the decrease is treated as if it was observed from 1 child to 0.

income basis. To correct for this discrepancy, in the estimation and simulation of individual choices I deflate gross wages by the ratio of Disposable Personal Income over Personal Income from NIPA Table 2.1 "Personal Income and Its Disposition" (?). On the other hand, there are two reasons why total labor compensation obtained from the aggregation of my CPS or SIPP wage measurements may differ from the observed wages and salaries from the national accounts. First, given the discrete mutually-exclusive choices of the model and the annual frequency, individuals may be assigned to work full-time in an occupation all year when they worked part time or only during a fraction of the year. Likewise, some individuals assigned to school or home may have worked for part of the year. Second, there are forms of labor compensation that are not wages (e.g. some types of bonuses and in-kind payments). These two discrepancies are corrected by adjusting total wage compensation appropriately. To correct for the first, I adjust my predicted aggregate skill units multiplying them by the ratio between the total Wages and Salaries from NIPA Table 2.1 by the predicted aggregate wages from my CPS or SIPP samples. To correct for the second, I multiply my skill prices by the ratio of BEA Compensation of Employees over Wages and Salaries, also from NIPA Table 2.1. The NIPA series starts in 1929. When needed backwards, the ratios are extrapolated backwards assuming them to be constant to the first available year.

A2. Microdata

Labor supply parameters are estimated using micro data from the March Supplement of the Current Population Survey (CPS), and the Survey of Income and Program Participation (SIPP). The CPS data are obtained from IPUMS (?). In particular, I selected the longitudinal Annual Social Economic Supplement files, linked by IPUMS across consecutive years. I extract those data from 1994–1995 to 2020–2021 which are the available years with information on immigrant status. The second year corresponds to period t, and the first year is $t-1$. The variables referring to the latter are only used to clean the survey linkage, and to construct the $t-1$ choice d_{a-1}. To clean the linkage, I sequentially drop the observations that satisfy the following criteria: age difference across surveys is more than two years (28,818 observations); the reported sex differs across surveys (2,189 observations); immigrant status differs across surveys (640 observations); race differs across surveys (3,234 observations); and region of origin differs across surveys (373 observations). For linked surveys 1994–1995 and 2020–2021, longitudinal weights are not available. I assume constant weights (equal to 2800, the approximate average of the other years) for those two years. Immigrant status, race, and region of origin are defined as described below. The SIPP data, **?**, was downloaded from the website of the National Bureau of Economic Research (NBER). I downloaded core and the second topical module of panels of 1988, 1990, 1991, 1992, 1993, 1996, 2000, and 2004, which are the ones in which region of origin can be identified with the required granularity. For 1988, I dropped observations in which the interview status is not 1, and I also drop information for reference month number five. For all periods, I drop individuals whom, after cleaning, do not have information on some of the state variables (including those without the second topical module, which I drop at the beginning of the cleaning process).

Age. In line with the timing of the model, I keep individuals aged 16 to 65. To accurately fit with the model's timing, I make some small age adjustments to the reported age as a result of the education adjustments described below. Other than those, in the SIPP I clean the age variable by taking the most frequently reported birth year among the different waves the same individual is interviewed, and computing age accordingly.

Demographic types. Individuals are classified in 20 types depending on gender, nativity, and race. Natives include individuals born in Puerto Rico and other U.S territories, and immigrants are individuals born elsewhere. The regions of origin for immigrants are defined above when describing their aggregate distributions. Natives are classified into nativity groups based on the race and Hispanic variables in the CPS, and race and ethnicity in the SIPP. The classification is done as follows. First, individuals are classified as Hispanic if they have a non-zero value in the hispan variable in the CPS, or, in the SIPP, either declare to be Mexican, Mexican-American, Chicano, Puerto Rican, Cuban, Central or South American, Dominican Republic or Other Spanish/Hispanic in the variables ethnicity (before 1993) or origin (1996 and 2001), or declare to be Hispanic in the variable origin (2004). For the remaining individuals, those indicating black as their single race, as one of their multiple races, or as their ethnicity are classified as non-Hispanic black. Finally, the remaining individuals are classified as non-Hispanic non-black. The variable sex is used to further classify individuals into males and females.[35]

Educational level. The main education variable is defined in years. In the CPS, school is provided in categories. For each category I assign the following years: 0 for "None or preschool"; 2.5 for "Grades 1, 2, 3, or 4"; 5.5 for "Grades 5 or 6"; 7.5 for "Grades 7 or 8"; 9 for "Grade 9"; 10 for "Grade 10"; 11 for "Grade 11" and

[35] In the SIPP, there are a very few discrepancies in the sex and race across waves for a given individual. In these very few instances, the most frequently reported sex/race is considered.

for “12th grade no diploma”; 12 for “High school diploma or equivalent”; 14 for “Some college but not degree”, for “Associate’s degree, occupational/vocational”, and for “Associate’s degree, academic program”; 16 for “Bachelor’s degree”; 17 for “Master’s degree”; 20 for “Professional degree”; and 21 for “Doctorate degree”. I then adjust age, education, and enrollment status to mimic the fact that in the model individuals who quit school are not allowed to return. The adjustments are as follows. If an individual reports to be enrolled in school in either $t-1$ or t, her education level is between 10 and 12 years, and her age is above education level plus six but below education level plus nine, I adjust age to education level plus six and the previous choice is defined to be school. If the individual reports to be enrolled in school in either $t-1$ or t, her education level is 7.5 or 9, and her age is respectively 16 or 16/17, I adjust age to be 16, education level to be 10, and previous choice to be school. If an individual reports the same education level in both $t-1$ and t, that education level is respected. In that case, if age is below what the education level would predict, I adjust as follows. If the individual reports to be enrolled in education last year or this year, I modify the age to reflect the minimum age that would allow for that education level and I ensure that the previous year decision was schooling. If the education level was referring to a range of years (e.g., some college), I additionally adjust exact years of education within the range to fit accordingly. If the gap in education is one education level and the individual reports to be enrolled in education last year or this year, I make analogous adjustments. For education levels in intervals, education in this case is assumed to be in the first year of the interval. For the remaining individuals that declare to be in school in either $t-1$ or t, I make the following choices. If education level is below 9 years or above 12 years I move their previous choice to not school and I keep their education and age. If their education level is 9 and age is below 18, the education level is increased to 10, and previous choice is considered to be school. If education is between 10 and 12 and their age is 20 or below, I keep their education level, I adjust the age as above, and I keep the previous year decision as school. For the remaining individuals, the previous choice is considered to be other than education and, if age is below years of education plus seven, it is increased to that level. Previous choice was set to non-school for all individuals of final ages different from education plus six or if the education level is 21 years.

For the SIPP, I use different variables for the panels before and after 1993. In panels 1988–1993, the highest grade of school attended (higrade) is converted into years of education (codes 1–12 directly correspond to years, and codes 21–26 were directly mapped into 13, 14, 16, 16, 17, and 17 years respectively, which, after the adjustments below, was the combination that best fits education distributions and educational choices conditional on education in the U.S.).[36] After 1993, the highest degree received or completed (eeducate) is used, with the same correspondence with years of education used in CPS. In both cases, when there were discrepancies within a year, the modal educational level reported was selected. Furthermore, age, education, and enrollment status is adjusted as in the CPS. The only differences with the CPS are given by the different panel structure of the two datasets. In the SIPP, an individual is considered enrolled if she reports to be enrolled at least once the given year. Moreover, if the individual is enrolled in t, this supersedes non-enrollment in $t-1$. Finally, if enrollment and education in $t-1$ are unknown (e.g. for the first year in the panel), this is inferred from enrollment at t (if enrolled) or by education level at t and age (if not enrolled).

Preschool children. Individuals are allowed to have 0,1, or 2+ preschool children (zero to five years old). For the CPS, this variable has been created by IPUMS: nchlt5. For SIPP, the variable is computed as follows. For each individual aged zero to five, I identify parents and legal guardians, as well as their spouses. Before 1996, only one of the parents can be identified. Then, I count how many children are associated to each individual. Finally, I sum the number of children assigned to respondent and spouse, and, to avoid double-counting, I define the number of children as the minimum between the total number of children in the household and the imputed sum of the two spouses. For each year and individual, the mode of the resulting variable across the year is considered.

Age at immigration. This variable is defined, for immigrants, as current age minus the difference between current year and year of immigration minus 16. When the resulting variable is smaller than 0 (i.e., entry before age 16), it is normalized to 0. In both datasets, some of the reported values correspond to (short) intervals. I keep the following values for intervals: 1949 for before 1950; 1951 for before 1952; 1954 fir before 1955; 1959 for before 1960; the center of the interval for periods of an odd number of years; and the first year above the center of the interval for periods of an even number of years.

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. For each individual I need to compute current and lagged choices. The procedure to assign individuals follows a hierarchical rule. First,

[36] In those years, if the individual is aged between 24 and 27, higrade is equal to 26, and is enrolled in either t or $t-1$, the education variable is further adjusted to fit age minus six.

I assign individuals to education. Previous year education status is cleaned as described above when introducing the education level. Current choice is assigned to education if, on top of previous year being education, the individual reports to be enrolled in education this year. Individuals not assigned to education, are assigned to work or home following different criteria in CPS and SIPP. In the CPS, an individual is assigned to work in the preceding year if she worked at least 40 weeks (or at least 20 if last year's choice was school) and at least 20 (usual) hours per week. In the SIPP, an individual is assigned to work in a given month if she worked at least 40/52=76.9% of the weeks for which work information is provided, and at least 20 usual hours per week averaged over the months with available information weighted by the fraction of weeks worked in that month.[37] Individuals assigned to work are then classified into occupations based on their reported main occupation.[38] To classify them into occupations, I first assign to STEM all individuals with a college degree that work in one of the occupations that the Bureau of Labor Statistics lists as STEM in any of its categories.[39] For the remaining workers, blue collar occupations include service workers, agricultural workers, construction workers, operatives, craftmen, and military, and white collar include non-STEM managers, professionals, scientists, technicians, and health workers, artists, sports and sales workers, and clerks.

[37] Weeks and hours information is cleaned as follows. In panels before 1996, I replace zero or missing total weeks worked by the largest number of weeks reported in the first or second job, or first or second businesses. Furthermore, if earnings are positive and the total reported weeks of work is smaller than the of weeks reported in the first or second job, or first or second businesses, I replace the former by the largest among the latter. As for the hours, usual hours worked in the weeks worked are used in panels before 1996. For panels of 1996 and later, I sum hours reported in the first or second job and in the first and second business. Then I confront this with the categorical variable that indicates hours worked per week in each month. If the categorical variable indicates that the individual did not work, I switch usual hours worked per week equal to zero. If the variable indicates that the individual worked 35+ hours all weeks and the sum of hours reported are below 35, I scale them up to 40. If the variable indicates that the individual worked 1–34 hours all weeks and the sum of hours reported are above 35, I reduce them to 20. I proceed analogously for the categories that at least one week 35+ or 1–34 and all others equal to zero. In the few cases in which all hours were missing but the categorical variable was not, I assigned 40 and 20 hours analogously.

[38] In the SIPP, individuals potentially report about two jobs and two business activities. Out of this for, my definition of main occupation is as follows. First, the one in which the individual worked more weeks (before 1996) or earned most income (1996 and after). If there is a tie, the one in which the individual worked more usual hours per week. If there is still a tie (before 1996), I use the one that provided the highest income. Finally, I resolve the very few remaining ties in the following order of priority: job 1, job 2, business 1 and business 2. After this classification, I take the most common job across all months in which the individual reported positive hours in some job, and some weeks of work during the month (if there is more than one mode, I favor occupations in the following order: STEM, then white collar, then blue collar.

[39] https://www.bls.gov/soc/Attachment_B_STEM.pdf, accessed on September 30th, 2022.

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, non-farm business income, and farm income in the CPS, and total personal income in the SIPP. Wages are deflated to 2012 US dollars using the consumer price indices provided by IPUMS with the CPS data. Wages are divided by total hours worked (obtained as the product of usual weeks worked last year and usual hours worked last year), and multiplied by 2080, which is the corresponding amount for year-round full-time work. A few extreme observations earning less than 1 dollar or more than 200 dollars per hour are set to missing.

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:

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

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:

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

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:

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

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:

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

Integrating this expression with respect to ε_T we obtain:

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

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 $e^{-\frac{t^2}{2}\sigma_{B\ell}^2}$, 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:

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

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