Field of Study, Career Choice, and Globalization*

Valerie Smeets
Aarhus University & CEPR

Lin Tian
INSEAD & CEPR

Sharon Traiberman
NYU & NBER

February 22, 2024

PRELIMINARY—PLEASE DO NOT CITE

Abstract

We study how students' field of study shapes the labor market consequences of trade shocks. We develop a dynamic general equilibrium framework that incorporates both a labor market where workers differ in their field of study, and an education margin where students choose their fields in anticipation of future demand. Leveraging uniquely rich Danish administrative data, and features of the Danish university application system, we develop a methodology to jointly estimate students' field preferences and key labor market parameters, while allowing for latent heterogeneity in workers' skills and student preferences. We find that heterogeneity in how responsive students are to shocks, with some students being substantially more responsive than previous estimates. However, we also find that students cluster into several preference types over fields of study—e.g., STEM—so that they are only elastic across a subset of fields. In ongoing work, we close the model and consider the general equilibrium consequences of globalization, as well as the effects of various labor market and education policies.

JEL: F14, F16, J24, I24

*Contact information: vas@econ.au.dk, lin.tian@insead.edu, st1012@nyu.edu.

1 Introduction

Economic shocks—for example, globalization, automation, or immigration—can lead to significant labor market disruptions. How do the impacts of these disruptions differ across workers? Answering this question requires understanding the differences in the tasks workers perform, and the extent to which their skills can be transferred to new domains. Crucially, workers' skills are shaped by the field of study they pursue while still students (Altonji et al., 2012a; Kirkeboen et al., 2016). Thus, understanding the distributional consequences of labor market shocks, and studying the effects of policy responses to these shocks, requires understanding how students' educational choices interact with the labor market.

In this paper, we develop a quantitative framework to study how students decide on their field of study, and subsequently how workers with different educational backgrounds navigate the labor market. In our model, students choose a field of study, understanding the labor market and forming expectation over the returns to each field. This expectation depends both on a forecast of future prices, and the expected ability on the labor market given a course of study. Workers then choose their occupation in every period, and can pay a cost to switch occupations in response to shocks. Crucially, workers differ in their comparative advantage across occupations. Their ability at different occupations is drawn at entry into the labor market, and the distribution is governed by their prior education decisions, test scores, and other covariates. Finally, our labor market and education model is embedded in a small open economy of labor demand where firms in different industries are exposed to import competition, and can choose whether to use offshored intermediates in production, and face varying export opportunities.

The context of our analysis is the Danish tertiary education system, which we discuss in detail in Section 2. We use administrative data from Denmark, where we observe a long panel of workers' career histories between 1998 and 2018. This is merged to a panel of tertiary education application data. In addition to the coverage and length of the data, the Danish education system has several attractive features ideal for our study. First, the data allows us to follow every student through multiple layers of the education system, and into their subsequent career. Second, similar to many countries[1], Denmark employs a centralized admission process to allocate students to programs. Specifically, students who opt to enter the college system apply for admission by submitting a list of priorities over institutions and programs (e.g., Business Economics at

[1] For example, Norway (Kirkeboen et al., 2016), Turkey (Balinski and Sonmez, 1999), and Chile (Hastings et al., 2013).

Aarhus University). The education ministry then allocates students according to a deferred acceptance algorithm, in which students are ranked by their application score and are assigned to places based on their preferred choices. In principle, the process is strategy-proof, provides us with truthful measures of individuals' ranking over fields. We also show that students tend to cluster on fields—for example, students whose first priority is a STEM field are highly likely to submit another STEM field as their second priority.

Using the dataset, Section 3 documents several stylized facts. We show that different fields of study prepare workers for a fairly specialized set of occupations, but with substantial heterogeneity in how specialized each field is. We then show that a large fraction of reallocation across occupations over time is explained by the sorting patterns of different cohorts. Specifically, for each occupation, we use the fixed effects decomposition proposed by Porzio et al. (2022). Depending on the occupation, cohort effects explain up to 90% of growth (or decline) between 1998 and 2018, highlighting the temporal variation in students' career choices. Finally, we use a Kitagawa-Blinder-Oaxaca decomposition to trace out the importance of field of study in how new cohorts re-sort across occupations. Altogether, this set of facts suggests that a model that seeks to understand how workers' respond to labor market shocks needs to incorporate both heterogeneity in workers' educational background as well as students' education choices.

Motivated by these facts, in Section 4 we develop a model of labor supply and educational choices. In the career choice block of our model, workers choose their occupation in each period à la Artuç et al. (2010). Workers make different career choices for two reasons: switching requires paying a partially-stochastic fixed cost, and workers differ in their comparative advantages across occupations. Specifically, at the beginning of their careers, workers draw a vector of productivities across different occupations. This vector is drawn from a distribution that depends on workers' innate talents, measured by test scores, their field of study, and other characteristics. Abilities differ within and across educational backgrounds. For example, social science majors, on average can have higher quantitative skills than humanities majors, while some social science majors can still have poorer quantitative skills and higher communication skills.

In the education block of our model, students choose which field of study to pursue. This block of the model builds on the large literature that models major choice (Kinsler and Pavan, 2015; Arcidiacono et al., 2020). Students do not know what their labor market type will be, but understand that it is mediated by their education choice. For example, they understand that they will draw from a distribution with higher quantitative skills if they study a STEM field. They are

allocated by reporting a ranking over fields to a centralized authority, who then allocates students using the deferred acceptance algorithm. Students are heterogeneous in three dimensions: they can vary in their non-pecuniary preferences across fields, they can vary in their preferences for income relative to non-pecuniary factors, and the value of a field of study can vary across students' observable characteristics, such as test scores.

In Section 4, we present our method for estimating the model. To estimate the labor supply block of the model we use a two-stage estimator, following the approach of Traiberman (2019), which builds on Arcidiacono and Miller (2011). In the first stage, we use a modified Expectation-Maximization (EM) algorithm to recover a subset of parameters, including the distribution of the comparative advantage. Workers' comparative advantage is unobserved to the econometrician, and treated as drawn from a set of finite latent types. Thus, we recover the probability that each worker is of each type. In the second stage, with the distribution of latent types known, we use the estimator of Artuç et al. (2010) to recover the remaining labor supply parameters.

We next estimate the parameters of the student's field choice problem. This exercise is complicated by two features of the model: first, students make decisions based on ability and preferences which are only partially observable to the econometrician; second, returns to different fields of study are not directly measurable in the data, as they depend on yet-unrealized future wages. To deal with the second problem, we develop a recursive representation of the returns to a field of study that depends on observed labor market variables and an unobserved continuation value. We exploit the fact that we observe multiple years of applications to nonparametrically estimate the unobserved continuation value in the final period, and use this as a control in earlier periods. To deal with the first problem, we build on the education literature that exploits the deferred acceptance algorithm (Luflade, 2019; Fack et al., 2019). Specifically, we posit that preferences are drawn from a mixed logit model (Berry et al., 1995), and use information on the students' ranking over many alternatives to identify latent heterogeneity over unobserved preferences (Beggs et al., 1981).

In Section 6 we present estimates from the labor market and field choice blocks of our model. On the labor market side, we show that controlling for workers' field of study, as well as their latent comparative advantage, is quantitatively important for estimating the elasticity of switching occupations, as well as occupational switching costs. For example, our model predicts that a 1% decline in the NPV of an occupation leads to an 11% increase in the exit from an occupation. Models omitting these mechanisms predict less than 1/5 of this effect, compensated with excessively

large estimates of occupational switching costs. We also find strong evidence that certain fields are very specialized in the set of occupations they prepare workers for, while others are more general (Leighton and Speer, 2020). Finally, we document both heterogeneity in comparative advantage across workers and, more novel, heterogeneity in the distribution of latent comparative advantage conditional on field of study.

Turning to the field choice block of the model, our results suggest an important role for heterogeneity across workers. First, we show that different types cluster into intuitive sets of fields. For example, one type clusters onto "STEM" fields, another type clusters into education and care fields, and so on. Second, we show heterogeneity in how responsive students are to changes in the returns to fields. For high-elasticity types, we show that a 1% increase in the NPV of a field increases entry into that field by .84%. This number is large relative to estimates in Wiswall and Zafar (2015). However, low-elasticity types are almost completely unresponsive to income changes. For all types, preferences are very strong—so even those who are relatively elastic are elastic only over a small set of fields. For example, we find that students who prefer STEM fields are relatively elastic *over STEM fields*, but have a near-zero probability of selecting fields outside of this set.

In Section 7, which is in progress, we will use our estimates to perform several counterfactual experiments. To do so, we first close our model by embedding the labor supply side into a small open economy model of labor demand. Firms in different industries face import competition and can choose to use offshored intermediates in production, generating heterogeneous responses to trade shocks across occupations. We adopt the exact hat algebra approach to estimate the labor demand parameters and solve the counterfactuals. The first counterfactual uses our model to quantify the labor market consequences of reductions in import and offshoring costs (hereafter, "Globalization"). We take full account of the distributional consequences across incumbents in different fields of study, as well as the dynamic implications of endogenous education decisions. In a planned second counterfactual, we consider the implications of an education policy that promotes entry into STEM fields.

Our project contributes primarily to the literature on trade and labor market dynamics. A bevy of influential reduced form work (Autor et al., 2013, 2014; Dix-Carneiro and Kovak, 2017; Hummels et al., 2014) has shown that trade shocks can have long-lasting effects on workers. While this literature has been very influential, it is incomplete when it comes to answering certain counterfactual questions—-especially those requiring thinking about how prices respond to policy changes

or alternative shocks. To answer these questions, there has been a large literature modeling labor market dynamics in an open economy. This project builds on this tradition. To date, authors have explored the importance of firm reallocation (Helpman et al., 2016), occupational and industrial reallocation (Artuç et al., 2010; Dix-Carneiro, 2014; Traiberman, 2019), and regional reallocation (Caliendo et al., 2019). These models have also explored questions beyond simple trade shocks, considering also trade deficits (Dix-Carneiro et al., 2023), migration (Burstein et al., 2020; Caliendo et al., 2021), and capital dynamics (Kleinman et al., 2023). Nevertheless, the distribution of human capital in these models is treated as a primitive, invariant to shocks and other changes in the economy. Indeed, in most papers, workers are treated as homogeneous conditional on their most recent sector, occupation, or region (and perhaps some observable characteristics). This is not in line with recent reduced form work in both developing countries, and advanced economies, that shows that human capital *does* respond to trade shocks (Atkin, 2016; Blanchard and Olney, 2017; Greenland and Lopresti, 2016). Our project leverages unique features of the Danish environment and richness of the Danish data—discussed in section 2—to integrate human capital decisions, labor, and trade.

Our paper also contributes to the large literature on college major choice and the labor market. Numerous papers have documented variation in the returns to different fields of study (Kirkeboen et al., 2016; Bleemer and Mehta, 2022; Andrews et al., 2022), and also explored the importance of different features of majors, such as specificity (Leighton and Speer, 2020).[2] Our paper is most closely related to the subset of the literature focused on estimating models of human capital accumulation in tertiary education—for example,Arcidiacono (2004) and Kinsler and Pavan (2015), who both estimate both a labor market process alongside a human capital decision. We also build on the seminal work of Keane and Wolpin (1997), who model labor demand in addition to a college decision. Our paper contributes to this literature in two ways. First, due to the breadth of the Danish data, we are able to estimate a very rich labor supply side of our model. This is important in order to capture the general equilibrium consequences of the globalization shocks we are interested in studying. Second, due to the centralized admissions system in Denmark, we can rely on recent advances in identifying education preferences, e.g., Luflade (2019), to discipline our model, featuring many different fields of study, and many different latent types.

[2] Altonji et al. (2016) and Altonji et al. (2012b) are excellent surveys, containing many more references, as well as discussions of structural modeling issues.

2 Data and Institutional Background

In this section, we describe the data sources and definitions used in the empirical analysis. We then provide a brief description of the Danish tertiary education system and show prominent features in applicants' applications and field choices.

2.1 Data Sources

Our study period is between 1998 and 2018. We combine several datasets for our analysis. The first is the *Integrated Database for Labour Market Research* (IDA)—a panel dataset with information on the universe of Danish workers over 15, both native and non-native. Briefly, this dataset includes information on employment status, earnings (in November of the year), ISCO-based occupation, age, sex at birth, and other demographic variables.

To IDA we merge the *Education Register* (UDDA), which contains information on highest completed level of education, year of graduation, as well as each individuals' field of study, categorized by the International Standard Classification of Education (ISCED). Finally, we use the data from the Coordinated Admissions System (KOT) run by the Ministry of Higher Education and Science. This dataset contains information on students' applications, including information on which programs they apply to and to which programs they are admitted. The Data Appendix contains a complete description of how we merged codes from KOT—which are based on the program and institution—to the UDDA codes.

Data on firms, which contains information on industry classifications, wages, and a list of products produced, come from two datasets: the *Firm Accounting Statistics Register* (FIRE) and the domestic production survey (ProdCom). The former allows for a firm level panel to be constructed. The latter, which is only available for industries classified as producing industries, contains information, for each firm, on product-level sales (both revenue and quantity) at the Combined Nomenclature (CN) 8 digit level. The CN is the EU's specialization of the HS 6, and we conduct all of our analysis aggregating to the HS 6 level. The last source of data from Denmark that we use is the *Danish Foreign Trade Statistics Register* (UHDI). These data are product-firm-origin/destination level customs data on all international transactions.[3] They contain data on value of imports and exports at the CN 8 level, and often also contain quantity and unit value information.

[3]For EU countries, there is a minimum reporting threshold.

We supplement this data with two publicly available datasets. First, we use the O*Net Database, which is maintained by the National Center for O*NET Development under sponsorship of the US Department of Labor. These data, which can be merged to the Danish occupation codes, are aggregated from survey responses pertaining to the importance of various tasks, skills, and types of knowledge to different occupations. Second, we use Comtrade for global imports and exports of goods at the HS 6 level.

2.2 Defining Fields and Occupations

Our definition of fields is based on the Danish variant of the ISCED-F classification of fields. The ISCED-F classification breaks fields into 27 broad categories. We aggregate several of the codes, and split apart a few codes (for example, separating Health to distinguish between Professional Bachelor's Programs and University Bachelor's Programs). The complete classification and our aggregation can be found in Table 1 of the Data Appendix. Our definition of level of education is also based on the ISCED system, and we divide students into three levels: no tertiary education, short-cycle or vocational education, and medium-cycle or higher education.

Occupations in Denmark are recorded according to the ISCO classification. There is a break in the recording of occupational codes beginning in 2009, as the ILO updated the codes. We use the concordance of (Humlum, 2021) to create a time-consistent series of occupations. For data quality issues, we focus on 2 digit occupational codes. After removing some small occupations, this yields 22 occupations. The Data Appendix contains a complete description of how we concord and aggregate occupations over time, how we deal with variation in the quality of occupational codes (e.g., some occupations are recorded according to an employee register, while others are imputed), and how we partially impute missing codes.

2.3 The Danish Education System

In our analysis we focus on two types of tertiary education in Denmark—both covered by the KOT data: medium-cycle, Professional Bachelor's degrees, and higher-cycle, University Bachelor's degrees.

Professional Bachelor's degrees require three years of schooling and focus on technical and readily-marketable skills. This class of programs can include vocational training, but is strictly broader and also includes business or education training. University Bachelor's programs are

closer to a standard American 4 year degree. In Denmark, a Master's program is very common. Therefore, in our analysis, we assume that students will complete a Master's and assume they enter the labor market only after having completed this program, some time in their mid 20s.

Entry in both of these sets of programs is based on high school GPA and exams. Students submit up to 8 choices for school and field. Different schools and fields of study have different GPA thresholds for entry. These thresholds are determined by demand from students and a supply that is determined by the Danish Ministry of Higher Education and Science. Admission into university by the GPA cutoff is referred to as the "Quota 1" system. There are alternative paths to education in Denmark. One may take extra exams or be admitted on certain alternative criteria. This is called the "Quota 2" system. Moreover, it is not uncommon for students to wait between high school and college if they do not receive their top option.

The Danish system is well understood by students, who are given ample information about the returns to different fields of study, and have information on historical GPA cutoffs for different fields and universities. It is also explained to them that they may truthfully reveal their preferences, as there is no penalty for doing so. Indeed, there is an online portal (formally a print publication) called the *UddannelsesGuiden* (Education Guide) that provides a great deal of information. For example, the listing for a Bachelor of Economics contains information on specific requirements for different institutions, the number of admits at every university for the last year, and information on possible career paths that includes information about income.[4] In our structural model, we treat students as well-informed both about their income prospects, as well as how the deferred acceptance algorithm works. However, it is not impossible—and indeed likely—that students still have bad information on the returns to fields. Unfortunately, lacking external measures of information and expectations we cannot explore this possibility. In our counterfactuals, we will assume that information frictions are still present in a way that makes our parameter estimates valid. We also assume that (at least at our level of aggregation of fields), strategic concerns about the finite list, as well as possible behavioral costs of non-admission, can be ignored.

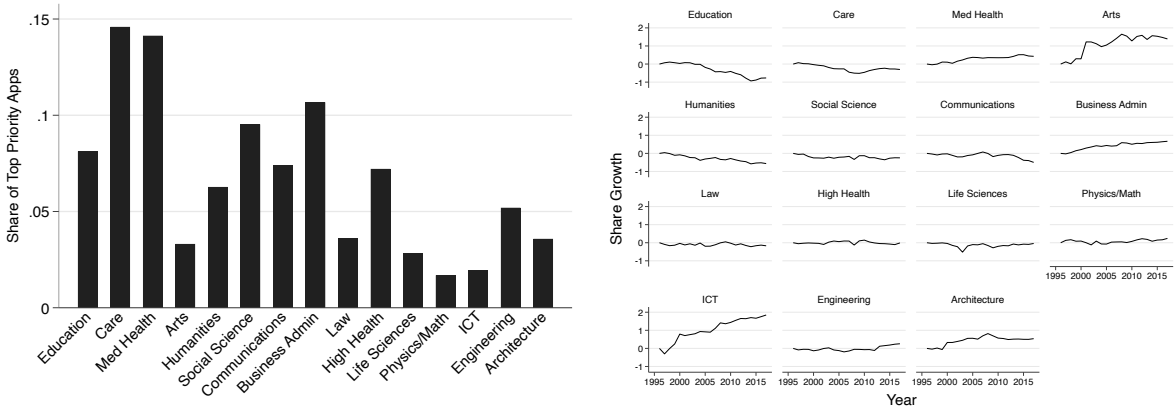
Table 1 displays the relationship between applicants' first-choice fields (represented in columns) and second-choice fields (represented in rows). Several notable patterns emerge from the data. First, there is clear diagonal dominance across all fields, indicating that majority of students prefer to select another program within the same field for their second choice. Second, in STEM-related

[4]See an example on economic major at https://www.ug.dk/uddannelser/bachelorogkandidatuddannelser/bacheloruddannelser/samfundsvidenskabeligebacheloruddannelser/erhvervsoekonomi/oekonomi. Past admission informationc can be found at https://www.ug.dk/KOT-tal

Figure 1: Summary Statistics on First Priority

(a) Unconditional Distribution

(b) Growth Over Time



fields such as life sciences, physics/mathematics, and engineering, we observe a substantial clustering of preferences, with a considerable proportion of students opting for their second or third choice within these interconnected fields. Conversely, in Education and Health, student preferences are markedly more narrowly focused, with a lower likelihood of selecting a second-choice field outside of their initial field of interest. This pattern suggests a strong inclination towards specialization in these disciplines.

Figure 1 displays the unconditional distribution of first priorities for students across fields of study, as well as the growth in fields over time. The most popular fields are in Care and Medium Health, followed closely by Business, Social Sciences and Education. The least popular fields are those in the STEM fields, as well as Arts. The unconditional distribution misses a few noteworthy trends in the data. In terms of shares, the education field is declining in popularity while the medium health is gaining. ICT, which begins on a very small base, has seen explosive growth, as has Business. The Humanities have shrunk in popularity. However, the distribution of applicants also changes over time. The number of students applying to tertiary education through the Coordinated Admissions system has grown 61% between 1996 and 2017. 96% of this growth has occurred since 2009, and is driven by more programs falling under the Coordinated Admissions System, rather than population growth in Denmark, which was only 9% between 1997 and 2017.

	First Priority														
Second Priority	Education	Care	Med Health	Arts	Humanities	Social Science	Communications	Business Admin	Law	High Health	Life Sciences	Physics/Math	ICT	Engineering	Architecture
Education	75.92			4.54	4.19										
Care	8.38	88.43	5.03												
Med Health	3.51	4.02	85.57							7.56	4.80				
Arts				46.98											7.54
Humanities				13.27	51.48	13.89	16.81		5.72						
Social Science				9.40	15.40	49.09	11.60	9.46	16.07	4.10		7.62	4.69		
Communications				9.98	15.48	7.46	42.61	10.59	5.84				3.79		
Business Admin						9.32	14.11	66.77	22.30			4.62	6.98	4.13	
Law						4.74			39.98						
High Health										67.36	11.37	3.57			
Life Sciences										9.25	50.37	15.56	4.60	7.53	
Physics/Math											10.61	39.84	4.85	6.39	
ICT												4.17	46.64	9.18	
Engineering											11.37	13.35	19.15	53.37	14.73
Architecture				5.85										10.11	55.75

Table 1: $P(Priority_{r+1}|Priority_r)$

Notes: The columns display students' first-choice fields and the rows display students' second-choice fields. The definition of fields is based on the Danish variant of the ISCED classification fields and is discussed in Section 2.2. Cells are censored below 3.5% for readability.

3 Stylized Facts

In this section, we document three stylized facts on the relationship of fields of study and career choices. The subsequent analysis focuses on students and workers in the highest level of education. Results for workers with the second highest level of tertiary education are current underway.

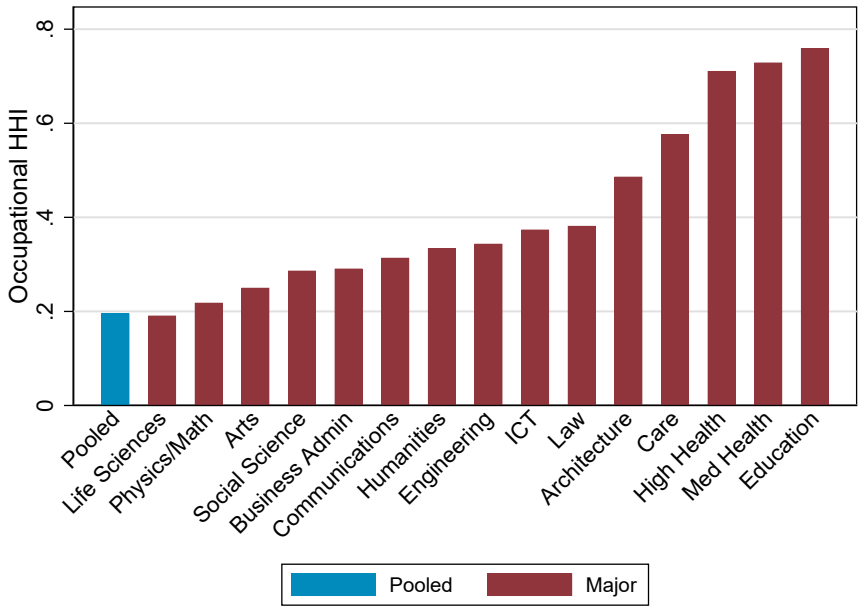
Fact 1: Different fields of study are associated with specialization in different occupations.

We first demonstrate that fields of study prepare students for a narrow band of occupations, but that there is substantial heterogeneity in this specificity. To measure how specialized a field is, we calculate the Herfindahl-Hirschman Index (HHI) across occupations for each field:

$$HHI_m = \sum_{o \in \mathcal{O}} s_{o|m}^2,$$

where $s_{o|m}$ is the share of workers from education field m working in occupation o. We calculate shares pooling across the whole economy. Figure 2 contains a bar graph for each field of study, as well as the HHI calculated by pooling all workers (in blue).

Figure 2: Field of Study-Occupation HHIs



Notes: The red bars represent the HHI for different fields of study. The definition of fields is based on the Danish variant of the ISCED classification fields and is discussed in Section 2.2. The blue bar represents the HHI calculated by pooling all workers.

The graph reveals several patterns in the data. First, many fields of study are very specialized. Some of these are not so surprising—for example health, education, and legal fields. Others more so, for example social science. Second, there is a great deal of heterogeneity in this specialization measure. For example, in contrast to social science, the HHI in technology and humanities is only slightly larger than the pooled HHI, suggesting little to no specialization relative to a random worker.

While the specialization pattern is driven by both labor demand and supply decisions, the results still suggest that both high specialization and heterogeneity may have important implications for labor market dynamics. For example, some workers—those in fields like humanities or business—may easily flow across occupations in response to shocks. Other workers—such as who studied the natural and social sciences—may be more rigid. Our next fact more explicitly looks at reallocation, and brings in the role of new entrants.

Fact 2: Occupational reallocation is driven by new cohorts.

Our second fact is that a substantial fraction of occupational reallocation occurs through the entry of new cohorts. Intuitively, changes in occupational shares can occur in two ways: incumbents can reallocate themselves (either by switching occupations, or by workers in certain occupations exiting and entering the labor force), or through new cohorts engaging in a different set of occupations.

In order to quantify these channels we adopt the approach of Porzio et al. (2022), who studied reallocation of workers in and out of agriculture in emerging economies. Specifically, let $s_{o|c,t}$ be the share of workers in cohort c working in occupation o at time t. We then project this share on cohort and time fixed effects for each occupation:

$$logs_{o|c,t} = \varphi_{oc} + \psi_{ot} + \varepsilon_{oct}. \tag{1}$$

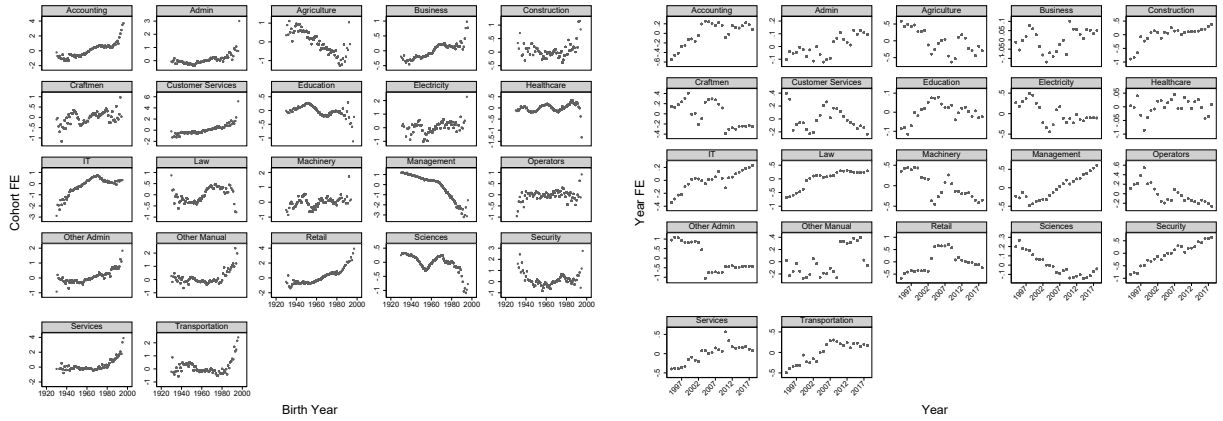
To interpret this equation, consider the case that cohort effects are irrelevant: workers are freely mobile (or at least equally mobile) across generations. In this case, the conditional share of workers in each occupation would be similar, so that $s_{o|c,t} \approx s_{o|c',t} \forall c, c'$. The only reason occupational shares would change over time would be due to shocks to labor demand that reallocated workers or common shocks to labor supply, both of which would be reflected in the year fixed effects.

Figure 3 plots both cohort and year fixed effects for each occupation. The first panel plots

Figure 3: Reallocation Decomposition

(a) Cohort Fixed Effects

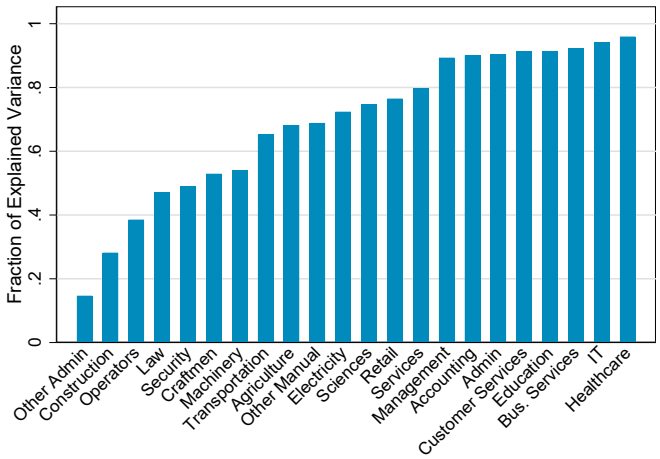
(b) Year Fixed Effects



Notes: The figure plots the cohort and year fixed effects from (1). The null hypothesis is a flat line bunched at zero.

cohort effects, with birth year of the cohort on the x-axis, and the fixed effect—interpreted as the deviation from the long-run economy-average share of that occupation—on the y-axis. If reallocation did not occur through changes in new cohorts' decisions, then cohort fixed effects would be flat and bunched at zero. Instead, we see for most occupations, the cohort effects are not flat. Indeed, the patterns are often not even monotonic, implying that in some occupations younger cohorts have actually re-entered into previously shrinking occupations.

Figure 4: Variance Explained by Cohort Effects



We can also ask what fraction of explained reallocation can be attributed to cohort fixed effects: $\sigma^2_{\varphi_{oc}} / \sigma^2_{\widehat{s_{o|c,t}}}$. We plot this, for each occupation, in Figure 4. This measure becomes as high as 90%. Nevertheless, neither time effects nor cohort effects are dominant in explaining realloca-

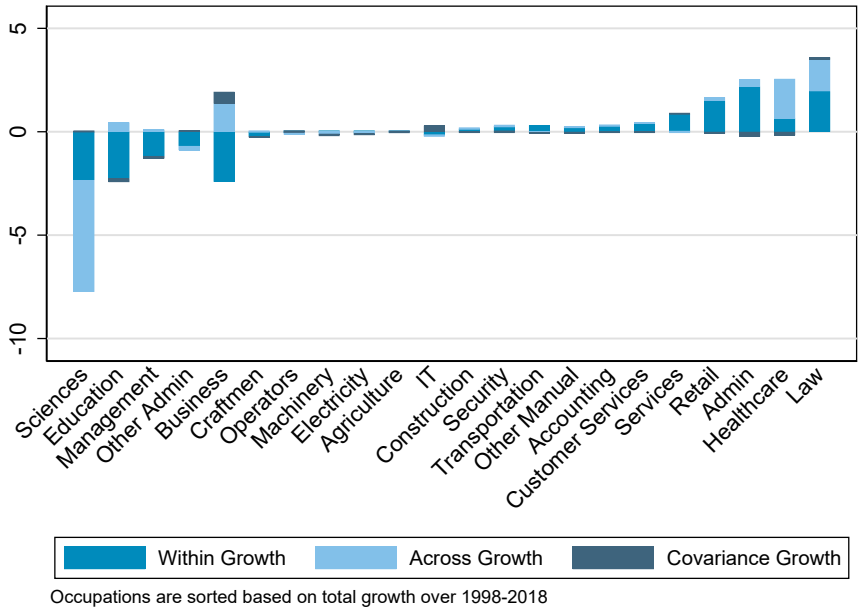
tion, motivating our focus on jointly modeling incumbents' occupational choices and new cohorts' education decisions.

Fact 3: New cohorts reallocate through their choice of field of study.

Finally, in our third fact, we document the extent to which reallocation across cohorts occurs through changes in the composition of fields of study. To do this, we do a Kitagawa-Oaxaca-Blinder decomposition of changes in initial occupations for new cohorts into within- and across-field of study. Specifically we consider the following decomposition:

$$\Delta s_{ot} = \underbrace{\sum_m \Delta s_{o|m} \times s_{m,t-1}}_{\text{Within}} + \underbrace{\sum_m \Delta s_m \times s_{o|m,t-1}}_{\text{Across}} + \underbrace{\sum_m \Delta s_m \times \Delta s_{o|m}}_{\text{Growth Covariance}}. \tag{2}$$

Figure 5: Decomposing the New Cohort Entry Margin



Notes: The decomposition of occupational reallocation follows (2). Occupations are sorted based on total growth over 1998 to 2018.

Figure 5 plots this decomposition for each occupation. The light blue bars, which capture across-field reallocation, are often dominant, suggesting that the majority of growth or decline in an occupation can be explained by changes in field of study, rather than changes in what workers with a certain field of study do. All together our facts paint a clear picture of how the occupational structure of the economy changes: incumbents and new cohorts both play a large role in shaping

reallocation. Moreover, new cohorts do this not by shifting their initial occupation in response to shocks, but shifting their field of study over time, giving them to new skills and thus access to new opportunities. With this in mind, we now turn to our model of labor supply and its estimation.

4 A Model of Labor Supply and Field Choice

We describe students' education choices and labor supply decisions in this section, with labor demand left to Section 7. Our labor supply model combines three main ingredients: first, students make forward-looking choices over which which field to pursue; second, workers with different educational backgrounds make forward-looking occupational choices; third, we incorporate latent heterogeneity in preferences over fields, as well as unobserved comparative advantage across occupations, in order to capture rich substitution patterns in response to shocks.

4.1 Environment

We consider a dynamic, small, and open economy. Time, indexed by t, is discrete. There is a measure L of workers, differentiated by their field, m, and a comparative advantage type k. Workers exit the labor force at a rate δ and are replaced by a new cohort.

In this economy, workers inelastically supply one unit of labor to an occupation. We denote the set of occupations by $o \in \mathcal{O}$. Before entering the economy, the young devote their entire time to education, choosing a field-level pair, giving rise to intergenerational changes in comparative advantage across occupations. We denote the set of education choices by $m \in \mathcal{F}$.

Figure 6 provides a schematic illustration of our labor supply model. The model starts with an *initial state* of high school graduates who enter with their test score and a type l, which reflect students' differences in talents and preferences. Students start their lives in the *field choice block*, in which they pick their field of study based on anticipated returns to different fields and a preference shock. Students graduate into the *career choice block* with their chosen field of study and comparative advantages across different occupations. New graduates make a forward-looking decision on their initial occupation. In each period afterwards, incumbent workers can choose to switch occupations, subject to switching costs. In the subsequent sections, we describe the model backwards, starting from the incumbents, moving onto the new cohorts, and finally the students.

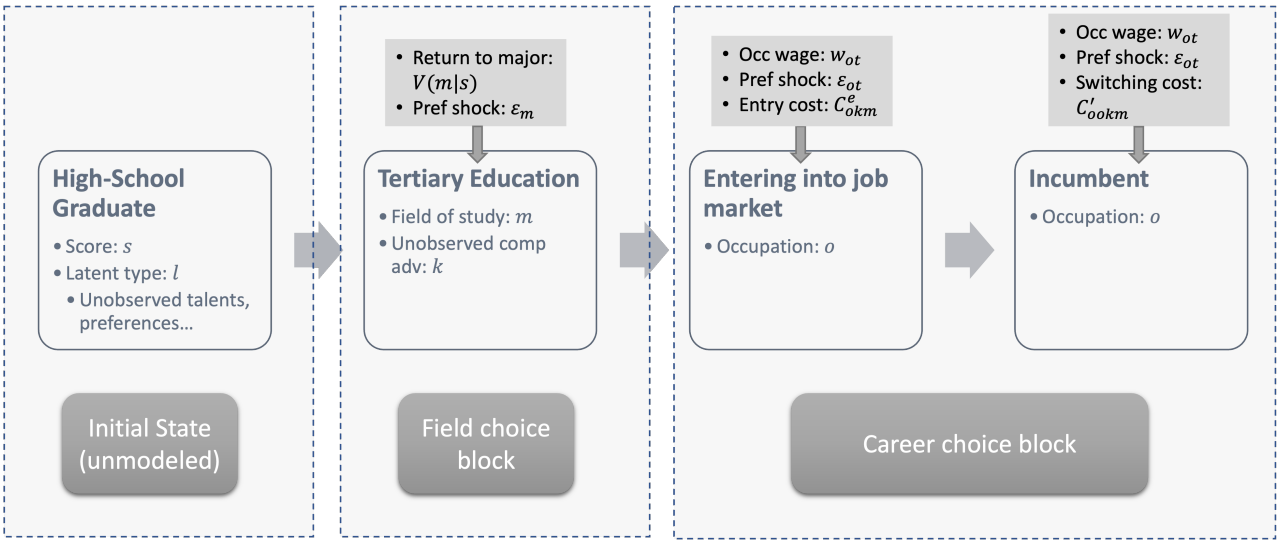


Figure 6: Illustration of the Labor Supply Model

4.1.1 Incumbent Workers

Incumbent workers who spent the last period in occupation o choose an occupation o' at the *beginning* of period t. Workers of type (k, m) supply human capital $H_o(k, m)$ to occupation o. In order to switch occupations they face moving costs $C_{oo'km}$ and receive a stochastic occupation preference shock, ε_o. The shocks are distributed Gumbel with mean 0 and scale parameter ν.

A worker's Bellman equation in period t is given by,

$$v(k, m, o, \varepsilon) = \max_{o''} w_{o''t} H_{o''}(k, m) - C_{oo''km} - \varepsilon_{o''t} + \beta E_t V(k, m, o''), \tag{3}$$

where $V \equiv \int v(\cdot, \varepsilon) dG(\varepsilon)$, is the value function v integrated over shocks, and β is the discount rate, $\tilde{\beta}$, multiplied by the probability of exit, δ.

We now describe how we parameterize H. We project each occupation onto Z tasks. We denote by $\zeta_{z,o}$ the loading on task z in occupation o, where a larger value of ζ means that the task is more important for output in that occupation. With this notation, each occupation is attached to a vector in the positive orthant of $\mathbb{R}^Z$, $\zeta_o \in \mathbb{R}^Z_{++}$. Each type of worker has a different ability at each task, $\alpha_{k,z}$. Worker's log output in an occupation is the dot product of α and ζ. We additionally allow for a field effect, α_{om} ,that scales up and down the human capital supply of workers in each occupation, and a type effect α_k that reflects the absolute advantage of type-k workers within a

field. Putting this together we have that,

$$\log H_o(k,m) = \alpha_{om} + \alpha_k + \sum_{z=1}^{Z} \alpha_{k,z}\zeta_{z,o}. \tag{4}$$

A few comments are in order H. First, workers have comparative advantage across different occupations through differences in their type k. As an example, suppose there were two occupations, two tasks, and two types. Suppose that for the first occupation, $\zeta = (1,0)$, and for the other occupation, $\zeta = (0,1)$. Similarly, suppose that $\alpha_1 = (1,0)$ and $\alpha_2 = (0,1)$. Then, type 1 workers would only produce output in occupation 1, while type 2 workers would only produce output in occupation 2. Nevertheless, given this horizontal differentiation there is no clear way to argue that one type has any absolute advantage.

Second, through the α_{om} terms, workers in different fields of study can have different comparative advantages, even conditional on type k. For example, workers in STEM backgrounds may have large values of α_{om} in technical occupations, such as engineering. Later, we discuss how during the education phase, the conditional distribution of types, k, is a function of m.

The above value function yields the following solution for V and a policy function on states that aggregates to the following transition matrix:

$$V_t(k,m,o) = \nu \log\left(\sum_{o'} \exp\left(\frac{-C_{oo'km} + w_{o't}H_{o'}(k,m) + \beta E_t V_{t+1}(k,m,o')}{\nu}\right)\right) \tag{5}$$

$$\lambda_t(o'|o,k,m) = \frac{\exp\left(\frac{-C_{oo'km} + w_{o't}H_{o'}(k,m) + \beta E_t V_{t+1}(k,m,o')}{\nu}\right)}{\sum_{o''}\exp\left(\frac{-C_{oo''km} + w_{o''t}H_{o''}(k,m) + \beta E_t V_{t+1}(k,m,o'')}{\nu}\right)} \tag{6}$$

Equations (5) and (6) play a crucial role in how we identify the structural parameters of the model in the next section.

4.1.2 New Cohorts

Like incumbents, new cohorts enter with a (k,m) pair. They make a multinomial choice over initial occupation, according to the utility function,

$$v_t^e(k,m) = \max_o -C_{o,km}^e - \epsilon_o + w_{ot}H_o(k,m) + \beta E_t V_{t+1}(k,m,o) \tag{7}$$

We assume ϵ is Gumbel with the same scale parameter as for incumbent workers. This yields the following choice probability for new cohorts:

$$\lambda_t^e(o|k,m) = \frac{\exp\left(\frac{-C^e_{o,km}+w_{ot}H_o(k,m)+\beta E_t V_{t+1}(k,m,o)}{\nu}\right)}{\sum_{o'}\exp\left(\frac{-C^e_{o',km}+w_{o't}H_{o'}(k,m)+\beta E_t V_{t+1}(k,m,o')}{\nu}\right)}. \tag{8}$$

Moreover, one can integrate to solve for the inclusive value of being type (k,m) at a time t:

$$V_t^e(k,m) = \nu \log \sum_{o'} \exp\left(\frac{-C^e_{o,km} + w_{o't}H_{o'}(k,m) + \beta E_t V_{t+1}(k,m,o')}{\nu}\right).$$

4.2 Field Choice

We now describe how students make a choice over field. Students enter their field choice decision with a test score, s, and an *ex ante* latent type, $l \in 1,...,L$. We think of a students' type as reflecting initial differences in both comparative advantage (e.g., a proclivity for science), and preferences. If students' preferences are very heterogeneous and strongly tilted towards particular fields, then students may not respond much to changing labor market conditions; by the same token, if the preferences across fields are not very strong, students may be reallocate quickly in response to anticipated shocks. In addition to their test score and type, students receive an idiosyncratic Gumbel-distributed preference shock with mean 0 and variance $\pi^2/6$ (i.e., we normalize the scale parameter to 1).

We model a field, m, as a *technology* that transforms a student into a type (k,m) worker. This technology is a conditional density over types k given test score s, which we denote by $\pi(k|m,s)$. Given this conditional distribution, we can calculate the value of field m for a student with scores s as:

$$V_t(m|s) = \sum_k V^e_{t+5}(k,m)\pi(k|m,s),$$

where recall $V^e_{t+5}(k,m)$ is the lifetime payoff to the student upon graduation with a labor market type k and a field m.

Our assumption on the education technology, $\pi(k|m,s)$, says that while s can inform k, and while l can inform one's preferences (as we will see below), *conditional* on k and m, one's initial conditions no longer matter for the labor market. We stress that this assumption does not rule out either observed or unobserved human capital differentiation in the labor market—it simply

assumes that k, once revealed, is a sufficient statistic for s as far as wages are concerned. This simplification maintains the model's tractability.

Given this technology, student i of type l with test score s has the following utility for each field:

$$U_{im} = \vartheta_{lm} + \theta_l V_t(m|s) + v \log(N_m) + \varepsilon^m_{im},$$

where ε^m is the field choice shock. The term ϑ_{lm} reflects the mean value of field m to type l students. This term reflects myriad factors such as preferences, or the perceived difficulty of finishing a field. An important parameter is θ_l, which governs how responsive workers are to labor market conditions. Finally, we include a field-specific control, $\log(N_m)$ to account for the number of programs within each field—for example "Math/Physics" can include math programs, physics programs, statistics programs, and so forth. This term serves as a proxy for the "love for variety" concept, capturing the diversity of choices available in each field. Upon receiving their shocks, students fill out a ranked list of fields of length R. Students are assigned to a field via the deferred acceptance algorithm. We assume that students reveal their preferences over fields truthfully.[5]

For ease of notation, let $\Theta_{lmt}(s)$ be the value of field m to type l at time t as a function of test score s. Moreover, let m_r be how an individual ranks their fields, so that m_1 means ranked first, and so on. Then, under the Gumbel assumption, one can show that the probability of observing a list $(m_1, m_2, ...,)$ is given by,

$$P(m_1, m_2, ..., m_R) = \prod_{r=1}^{R} \frac{\exp(\Theta_{lm_r t}(s))}{\sum_{m \in \mathcal{F} \setminus m_{-r}} \exp(\Theta_{lmt}(s))}.$$

This expression is a concatenated probability of the usual logit choice probabilities, with the choice set adjusted at each stage.

Despite the seeming complexity of our demand system, it is actually quite tractable and nearly encompasses standard models as a special case. First, if there is no unobservable heterogeneity or occupational choice component in the labor market, then this model collapses to one where workers estimate the net present value of earnings by taking the NPV of cross-sectional average earnings in each field. Second, if $\beta = 0$, then this collapses to a model where workers only care about their initial earnings on the job market.

[5]Given that the list is truncated list, students may be incentivized to include a "safety" choice. At our level of aggregation, few students are ever actually bound on the list, so we ignore this constraint.

5 Estimation

We present our estimation strategy in the same order as the model: first we estimate the career choice block of the model, then turn to the education block of the model.

5.1 Labor Supply of Incumbent Workers

Our estimation of the labor supply parameters for incumbent workers largely follows Traiberman (2019), building on Arcidiacono and Miller (2011) and Artuç et al. (2010). We implement a two stage estimator: first using an EM algorithm to estimate the parameters of human capital supply, the distribution of unobserved heterogeneity, and the transition probabilities across occupations; second, with the estimated unobserved heterogeneity in hand, we estimate the remaining structural parameters, C and ν, through an inversion of the choice probabilities in equation (6) as in Hotz and Miller (1993).

Turning to the first stage, we assume that log observed wages of individual i at t are the model's wages plus a Gaussian noise term.[6] Let $f(w_{it}|k,o,m,t;\alpha)$ be the Gaussian PDF of log wages conditional on type, field, occupation, and year, with parameters α. Let q_k be the probability of being type k. In this case, the likelihood contribution of person i is given by,

$$\tilde{\mathcal{L}}_i = \sum_{k=1}^{K} q_k \times \lambda^e(o_1, m|k) f(w_{i1}|k,o,m,1;\alpha) \times \left[\prod_{t=2}^{T} f(w_{it}|k,o,m,t;\alpha) \times \lambda_t(o_t|o_{t-1},m,k)\right].$$

The full likelihood, given that q_k is unobserved, is the product over $\tilde{\mathcal{L}}_i$. The above function is not amenable to easy maximization since the product of sums destroys the log-linearity typically exploited in maximum likelihood estimation.

A simple solution to this issue is the Expectation-Maximization (EM) algorithm. To proceed, we define the following function, which is the expectation of the complete likelihood if k were observed, which we continue to call the log likelihood in an abuse of terminology:

$$\mathcal{L} = \sum_{i=1}^{N}\sum_{k=1}^{K} q_{ik}\left[\log q_k + \sum_{t=1}^{T}\log f(w_{it}|k,o_{it},m_i,t;\alpha) + \sum_{t=2}^{T}\log\lambda_t(o_{it}|o_{i,t-1},m_i,k) + \log\lambda^e(o_{i1},m_i|k)\right],$$

where q_{ik} is the probability that individual i is type k.[7] This function is the starting point for

[6] In Traiberman (2019), this noise term is treated as ex post iid heterogeneity. This can be included in the model, as it shifts the value of an occupation. Nevertheless, since this ex post shock has no effect on selection, we ignore it in our present setting.

[7] The reason this is the expectation is that if we *observed* k then q_{ik} would be replaced with an indicator for i's type.

the EM algorithm. However, this remains a difficult problem due to the presence of λ_t, the choice probabilities. Solving for the choice probabilities requires solving the full structural model—which can be computationally expensive and also requires specifying the process on aggregate shocks. To circumvent this issue we introduce an approximation to $\mathcal{L}$. In particular, following Arcidiacono and Miller (2011), we approximate λ non-parameterically:

$$\hat{\mathcal{L}} = \sum_{i=1}^{N} \sum_{k=1}^{K} \left[\sum_{t=1}^{T} \log f(w_{it}|k, o_{it}, m_i, t; \alpha) + \log \hat{\lambda}_t(o_{it}|o_{i,t-1}, m_i, k) \right] q_{ik}, \tag{9}$$

where the hats atop the λ terms imply that they are not outputs of the model, but estimated directly from the data.

Another issue in maximizing the likelihood function is the problem of initial conditions. In particular, agents vary in their field and initial occupation—neither of which is independent of k. We also have information on a subset of workers' test scores, s_i, which our model allows to inform one's unobserved type. And so, instead of estimating q_k and $\lambda(o_{i0}, m|k)$, we follow Wooldridge (2005) to deal with initial conditions. Specifically, we estimate $\lambda(o_{i0}|m_i, k_i, t_{0i})$—where t_{0i} is a cohort fixed effect—and then estimate separately $q(k|m, t_0, s)$, where t_0 is again cohort fixed effects and s is test scores. We model this parametrically with a logit link function:[8]

$$q(k|m, t_0, s) = \frac{\exp(\delta_{kmt_0} + \beta_k s)}{\sum_{k'} \exp(\delta_{k'mt_0} + \beta_{k'} s)}.$$

We include our full implementation of the EM algorithm in appendix B, but briefly describe the idea. Given an initial guess of q_k and α, one can proceed in the following loop:

1. Update the conditional probabilities for each type:

$$q'_{ik} = \frac{\left[\sum_{t=1}^{T} \log f(w_{it}|k, o_{it}, m_i, t; \alpha) + \log \hat{\lambda}_t(o_{it}|o_{i,t-1}, m_i, k) \right] q_k}{\sum_{k'=1}^{K} \left[\sum_{t=1}^{T} \log f(w_{it}|k', o_{it}, m_i, t; \alpha) + \log \hat{\lambda}_t(o_{it}|o_{i,t-1}, m_i, k') \right] q_{k'}},$$

   which is Bayes' rule.

2. Update model parameters. We update q_k by taking the mean of q_{ik} across individuals. Wage parameters can be estimated using a weighted linear regression of $\log w_{it}$ on occupation-

In the expression above, this indicator is replaced by its expectation, q_{ik}.

[8] If s were continuous we could use a fully non-parametric bin estimator. Indeed, for cohorts where we do not observe test scores, this reduces to a bin estimator. For cohorts with test scores, we also include a dummy whenever the score is missing (which we treat as an exogenous event).

year fixed effects, occupation-field fixed effects, and tasks, where the weights are q_{ik}. $\hat{\lambda}$ can be updated through a weighted bin estimator of transition rates:

$$\hat{\lambda}(o'|o,k,m,t) = \frac{\sum_{i\in(o',o,k,m,t)} q_{ik}}{\sum_{i'\in(o,k,m)} q_{i'k}}.$$

This process is guaranteed to monotonically increase towards the maximizer of the likelihood function.

In the second stage of the estimation procedure, we use the Hotz-Miller inversion to solve for the switching cost parameters and the scale of the Gumbel distribution. To see how this is done, notice that after estimation of the first stage, $w_{ot}H_o(k,m)$ is observed for every possible (o,k,m), as are the switching probabilities, $\lambda_t(o'|o,k,m)$.[9] In Appendix B.3, we show how to use (6) to arrive at the following regression equation:

$$\log\left[\frac{\lambda_t(o'|o,k,m)}{\lambda_t(o|o,k,m)}\right] + \beta\log\left[\frac{\lambda_{t+1}(o'|o',k,m)}{\lambda_{t+1}(o'|o,k,m)}\right] = -\frac{C_{oo'km}(1-\beta)}{\nu} + \frac{1}{\nu}\left(w_{o't}H_o(k,m) - w_{ot}H_o(k,m)\right) + \xi_{oo'kmt+1}, \quad (10)$$

where $\xi_{oo'kmt+1}$ is a time $t+1$ expectational error. Compared to equation (6), the current and next period's continuation values are differenced out, which addresses the remaining selection issue. Under rational expectations, the error is uncorrelated with time t variables, so that equation (10) can be estimated by ordinary least squares.

5.2 Labor Supply of New Cohorts

This is a standard discrete choice problem. We normalize, without loss of generality, the entry cost into the occupation 1 ("Management") to zero. In Appendix B.3 we derive the following estimating equation:

$$C^e_{o,km} = \left(w_{ot}H(o,k,m) - w_{1t}H(1,k,m)\right) + \beta C_{1,o,km} - \nu \times \left\{\log\frac{\lambda^e_t(o|k,m)}{\lambda^e_t(1|k,m)} + \beta\log\frac{\lambda_{t+1}(o|o,k,m)}{\lambda_{t+1}(o|1,k,m)}\right\} \quad (11)$$

Afte this step, all parameters of the labor supply side of the model are estimated. However, we stress that at this stage the value function V remains unobserved unless we were to solve the model. To make this clear, and ignoring expectational error, we rewrite the entry value for an

[9] Even though these are generated regressors, we will avoid the cumbersome use of hat notation. Instead, it should be understood that these are estimated.

occupation in the following discounted form:

$$V_t^e(o,k,m) = -C^e_{o,km} + \sum_{\tau=t}^{\mathcal{T}} \beta^{\tau-t} \left[w_{ot} H(o,k,m) - v \log \lambda_t(o|o,m,k) \right] + \beta^{\mathcal{T}+1} V_{\mathcal{T}+1}(o,k,m).$$

Since $V_{\mathcal{T}+1}(o,k,m)$ is unobserved, we cannot recover $V_t^e(o,k,m)$ without solving the model. Nevertheless, the net present value of entering an occupation is an input into the field choice problem. We now turn to how we estimate the field choice block of the model, and describe a recursive formulation of the utility function that allows us to avoid needing to observe V_t^e.

5.3 Field Choice

To describe how we estimate field choice, we first describe how we set up a recursive representation of field-choice preferences, and then describe the likelihood and how we maximize this function. First, under the assumption that one's latent preference type, l, is independent of the conditional probability of one's labor market type, k, conditional on observable characteristics and field choice, we can estimate the education production function through the regression:

$$q_{ik} = \beta_m + f(s; \beta^s_m) + \epsilon_{ikm},$$

where q_{ik} is the probability that individual i is type k, m is their field, and s is their test score, and f is a function of test scores. We bin test scores into quintiles for the function f.[10] With this function in hand, and the choice probabilities conditional on type known, we can construct the payoff to a field m at time t for type l individuals as,

$$U_{tlm}(s) = \vartheta_{lm} + v \log(N_{mT}) + \theta_l \left[\sum_k \pi(k|s,m) V^e_{t+5}(k,m) \right]. \tag{12}$$

As explained above, this is not observable for the last period in the sample—which here corresponds to 5 periods before the last period in the labor market sample. To resolve this issue, we first estimate the composite term $\tilde{U}_{tlm}(s) \equiv (U_{Tlm}(s) - v \log(N_{mT}))$ in the final period. Then, in Appendix B.3, we derive the following representation for the value of entering the labor market

[10] In the previous section we had estimated a similar function for initial conditions. Here we bin students into quintiles to economize on computation time.

as a (k, m) type worker at t, which holds true for any choice of initial occupation o:

$$\begin{aligned} V_t^e(k,m) = &-(1-\beta)C_{okm}^e + w_{ot}H(o,k,m) + \beta V_{t+1}^e(k,m) - \nu \log \lambda_t^e(o,k,m) \\ &- \beta\nu \log \frac{\lambda_{t+1}(o|o,k,m)}{\lambda_{t+1}^e(o|k,m)}. \end{aligned} \tag{13}$$

Finally, we can plug (13) into (12) and rearrange to derive the following:

$$U_{tlm}(s) = (1-\beta)\vartheta_{lm} + \upsilon(\log(N_{mt}) - \beta \log N_{m,t-1}) + \theta_l \sum_{k=1}^{K} \pi(k|m,s) V_{t+5}^F(m,s) + \beta U_{t+1lm}(s), \tag{14}$$

where,

$$V_{t+5}^F(m,s) = -(1-\beta)C_{okm}^e + w_{o,t+5}H(1,k,m) - \nu \log \lambda_{t+5}^e(o|k,m) - \beta\nu \log \left(\frac{\lambda_{t+6}(o|o,k,m)}{\lambda_{t+6}^e(o|k,m)} \right),$$

is an entry value term that *only* depends on observable data. We can iterate forward on this equation to arrive at our final formulation of U:

$$U_{\bar{T}-\tau,lm}(s) = \vartheta_{lm}(1-\beta^{\bar{T}-\tau}) + \upsilon \log(N_{m,T-\tau}) + \theta_l \underbrace{\left[\sum_{\tau'=\tau}^{\bar{T}-1} \beta^{\tau'-\tau} V_{\bar{T}-\tau'+5}^F(m,s) \right]}_{XV_{\tau,\bar{T}}^F(m,s)} + \beta^{\bar{T}-\tau} \tilde{U}_{\bar{T},lm}(s). \tag{15}$$

To interpret this equation, notice, for example, that if $\tau = \bar{T}$ this collapses to $U_{\bar{T}-\tau,lm}(s) = U_{\bar{T}-\tau,lm}(s)$. While, if $\tau = \bar{T} - 1$ this reduces to the one step forward iteration in equation (14). For ease of notation, we further denote the term in brackets as $XV_{\tau,\bar{T}}^F(m,s)$.

With (15) in hand, we turn to deriving the likelihood function. We start by deriving the likelihood conditional on l, and then move to the estimation of latent types. To construct the likelihood conditional on type, let i index individuals, let $\mathcal{C}_i$ be their choice set (which may change over time, or be a function of their test score s_i), and let $(m_1, m_2, ..., m_R)$ be their ranked choice list. Suppose first that we observe individual i in the final period of the education data, $T-6$. For their *first* ranked choice, conditional on being of type l and having score in decile s_i, the likelihood contribution will be:

$$L_{i,T-6,1} = \frac{N_{m_{r_i}}^{\upsilon} \exp\left(U_{m_r l_i T-6}(s_i)\right)}{\sum_{m' \in \mathcal{C}_i} N_{m'}^{\upsilon} \exp\left(\tilde{U}_{m' l_i T-6}(s_i)\right)}.$$

After a choice has been made, the number of options needs to be adjusted. The notation can now be cumbersome, so let $N_{i,t,r,m}$ be the number of options person i at time t has at position r in field

m given their past choices. For example, if there were 40 nursing programs, and an individual filled out a nursing program as their first priority, then there are 39 programs still available for the second priority. With this notation, the likelihood contribution for the full ranked list is given by,

$$L_{i,T-6}(m_1, m_2, ..., m_R, s_i) = \prod_{r=1}^{R} \frac{N^v_{i,T-6,r,m_{r_i}} \exp\left(\tilde{U}_{m_r l_i T-6}(s_i)\right)}{\sum_{m' \in \mathcal{C}_i} N^v_{i,T-6,,r,m'} \exp\left(\tilde{U}_{m' l_i T-6}(s_i)\right)}.$$

While the utility term, $\tilde{U}$, appears in the remainder of the likelihood, for computational convenience, we will only update this term using last period data. Ultimately, we will use the EM algorithm once again, since l is unobserved. To this end, letting q_{il} be the probability that i is of type l, we can write the log likelihood contribution from the final period as,

$$LL_{T-6} = \sum_l \sum_i \sum_r q_{il} \left(v \log N_{i,r,m_{r_i}} + \tilde{U}_{m_r l_i T-6}(s_i) - \log \left(\sum_{m' \in \mathcal{C}_i} N^v_{i,r,m'} \exp\left(\tilde{U}_{m' l_i T-6}(s_i)\right) \right) \right).$$

Given a value of v—which we discuss below—one can maximize this expression with respect to $\tilde{U}_{\bar{T},lm}(s)$ by iterating on the first order condition, as in Berry et al. (1995). Details, including the exact expression from the FOCs, can be found in Appendix B.3.

With a current estimate of $\tilde{U}_{\bar{T},lm}(s)$, we turn to discussing how to estimate the remaining choice parameters conditional on knowing type l. As before, we can construct the individual likelihood contribution for individual i at time t if they are at type l using the multinomial logit probabilities. Taking the logarithm and summing across all periods and individuals leads to the following objective:

$$\begin{aligned} LL_t = \sum_l \sum_i \sum_r q_{itl} \Bigg(& v \log N_{irm_{r_i}} + \vartheta_{lm}(1 - \beta^{T-6-t}) + \theta_l X V^F_{t,T-6}(m, s) + \beta^{T-6-t} U_{T-6,lm}(s) - \\ & \log \left(\sum_{m' \in \mathcal{C}_i} N^v_{irm'} \exp\left(\vartheta_{lm}(1 - \beta^{T-6-t}) + \theta_l X V^F_{t,T-6}(m, s) + \beta^{T-6-t} U_{T-6,lm}(s) \right) \right) \Bigg) \end{aligned} \quad . \quad (16)$$

We maximize (16) by iterating on the first order conditions. Derivations of these conditions, and a discussion of the moments they map to in the data are in Appendix B.3.

Finally we turn to estimation when l is unknown. In this case, we use the EM algorithm estimate the distribution of types. To update the conditional probability that individual i of type

l, conditional on a guess of parameters, we use Bayes' rule:

$$q_{itl}^{(g+1)} = \frac{\exp(LL_{itl}^{(g)})q_l^{(g)}}{\sum_{l'}\exp(LL_{itl'}^{(g)})q_{l'}^{(g)}}.$$

We update the model parameters given q_{il} by iterating on the likelihood's first order conditions, as described above. We repeat this process, increasing the likelihood in each step, until the parameters converge within some tolerance.Notice that the ranked lists are *crucial* for identification. If we only observed a single choice for each individual, the distribution of types, l, would not be identified.

6 Estimation Results

We now discuss our results. In the current draft, we focus on workers and students at the highest level of education (professional and university bachelor's). In ongoing work, we are estimating the parameters for the middle level of tertiary education, as well as the labor supply side of the model for those with no tertiary education.

6.1 Labor Supply

Table 2 displays our results for the estimation of the labor supply parameters in our model.[11] The first row is the value of $1/\nu$ we obtain, scaled so that average wages are equal to 1.[12] The second row is the average value of Cnu across all types and occupations. This second row is a simple average over all occupations and is *not* waited by actual switches, nor adjusted for moving shocks. And so some care needs to be taken in interpreting these parameters—a point we return to below. The columns are the results of three different estimation strategies.

In the first column, we estimate the model of Artuç et al. (2010) across occupations—we do not include any heterogeneity in comparative advantage, nor do we include heterogeneity by field. We estimate the model using OLS, with a complete set of occupation-pair FE to capture moving

[11] At this time, we are only estimating the model for workers with University Bachelor's degree.

[12] We divide raw income by mean income. After further sample selection (detailed in the Data Appendix as well as code), which mostly amounts to our current focus on college graduates, and age restrictions, the mean wage is 1.26. Our estimated value of $\hat{\nu}$ in the main specification is 1.12, so that we report 1.26/1.12 as our estimate of the re-scaled value of $1/\nu$. Nevertheless, for consistency in the model, we do not use the re-scaled for ν in our counterfactual analysis—this is just for interpretation as a-semi elasticity.

costs.[13] In the second column, we extend this model to one with workers differentiated by field choice. In the final column, we present the result of the full estimation routine.

Table 2: Labor Supply Parameters

Parameter	(1)	(2)	(3)
$\bar{w}/\nu$	0.30	0.38	1.21
	(.122)	(.041)	(-)
$C/(\bar{w} \times \nu)$	5.77	4.60	4.01
Field Controls		✓	✓
Latent Types			✓

Notes: The first row presents the estimates for $1/\nu$ in (5), scaled by the mean wage. The second row presents the estimates for C/ν in (5). The first column reports the estimates using Artuç et al. (2010), which does not consider heterogeneity by field or heterogeneity in unobserved comparative advantage. The second column introduces worker heterogeneity by field. The third column includes both.

Turning to the first row, we see that in a model that ignores all worker heterogeneity, the value of the inverse semi-elasticity is quite low at 0.24. To further interpret this number, we can calculate the switching elasticity with respect to a temporary shock to wages in occupation o:

$$\frac{d \log(1 - \pi_{oo})}{d \log w_o} = -\frac{w_o}{\nu} \pi_{oo}.$$

The mean staying-rate in the economy is approximately 85%. Hence, the first column can be interpreted as saying that a 1% decrease in the returns to an occupation will raise the exit rate from that occupation by .26%—a very small number given a base of 15%. Permanent shocks will magnify the response by approximately $1/(1-\beta)$, a factor of 12.7.[14] Controlling for field increases the elasticity, reduces the estimated switching costs, and increases the precision of the estimates.

The full model, which incorporates observable and latent heterogeneity, yields a semi-elasticity of 1.21—more than a three-fold increase over the models ignoring latent heterogeneity. In elasticity terms, a temporary 1% decrease in wages in an occupation will, *ceteris paribus*, lead to a 1.02% increase in the switching rate—or about a .15 percentage point increase in the exit rate for an occupation. For a 1% permanent decline in wages, the effect on switching out is approximately 11% or 1.7 percentage points.

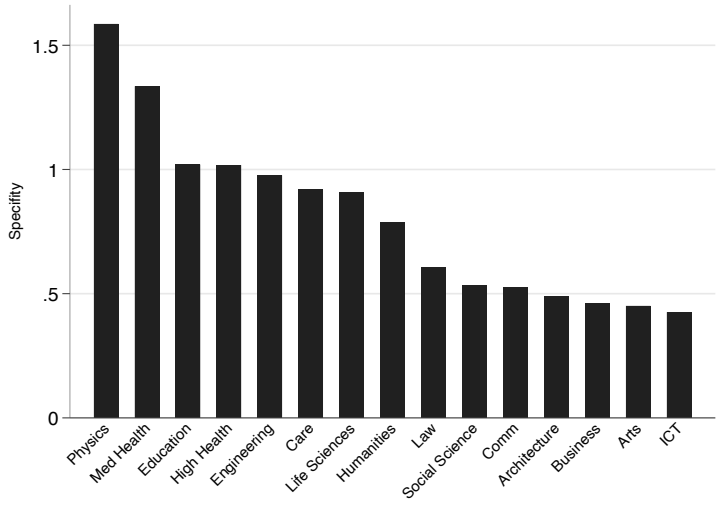
Turning to the second row of the table, in reporting costs we scale by the mean wage and

[13] In ACM, there is a different timing assumption on when agents make decisions. If we use an IV estimator inspired by their assumptions, we find broadly similar results in terms of the ordering across models. However, the IV estimator is substantially noisier.

[14] This approximation is not exact due to the presence of the option value term in the inclusive value of an occupation. More precisely, in a stationary environment, $V_o = \frac{w_o - \nu \log \pi_{oo}}{1-\beta}$. If $\pi_{oo} \approx .85$ then the option value term is relatively small if ν is small.

the parameter ν. This allows the switching costs to be interpreted in units of annual wage. The costs are larger in models that do not take account of worker heterogeneity. This result is similar to Dix-Carneiro (2014) and Traiberman (2019), who point out that ignoring heterogeneity equates the earnings of workers on the margin of switching with inframarginal workers. This pushes down ν, and consequently raises the measure of costs.

Figure 7: Field Specificity



In the interest of space, we leave the discussion of latent types and comparative advantage to Appendix A.1. However, we briefly comment on differences in comparative advantage by field, which is novel to our study. To construct a simple measure of the specificity of a field of study—our notion of the importance of comparative advantage—we compare the maximum value of α_{om} to the minimum across occupations:

$$\text{Specificity}_m \triangleq \exp\left(\max_o \{\alpha_{om}\} - \min_o \{\alpha_{om}\}\right).$$

Figure 7 plots this measure across fields of study. The most "specific" fields of study are physics, health fields, and education. This suggests that there are strong returns for these fields to sorting into very specific occupations. To interpret these numbers, all else equal, physics majors earn 1.5 times as much in their best-matched occupation relative to their worst. Studying the arts, ICT, or business is quite different, with a much flatter gradient. For example, for ICT majors, there is a sizable, but smaller, 45% gain to being in the best-matched occupation relative to the worst.

6.2 Field Choice

We allow for $L = 4$ latent types in our field choice estimation. We assume that both ϑ_{lm} and θ_l are type-specific, but υ does not vary with type or field. We assume that students' choice sets are common, and include all programs in a given year. Table 3 contains our estimates of θ and υ, as well as the share of each type l in the population, q_l.

Table 3: Field Choice Parameters

Parameter	Type 1	Type 2	Type 3	Type 4
θ	0.021	0.074	0.027	0.073
υ	0.082	0.082	0.082	0.082
q_l	0.418	0.108	0.268	0.206

Notes: The table reports the results of field choice estimation by latent type l. The first and second rows report the estimates of parameters in (12). The third row reports the share of each type in the population.

The estimates of θ seem to suggest that types 1 and 3 are relatively inelastic, while types 2 and 4 are more elastic. While we show below that type 2 is something of a special case—essentially completely inelastic on one field—the results for other types suggests a good deal of heterogeneity in their responsiveness to shocks. To better interpret the coefficients, we note that the elasticity of choosing m with respect to a shock to the value of that field is given by,

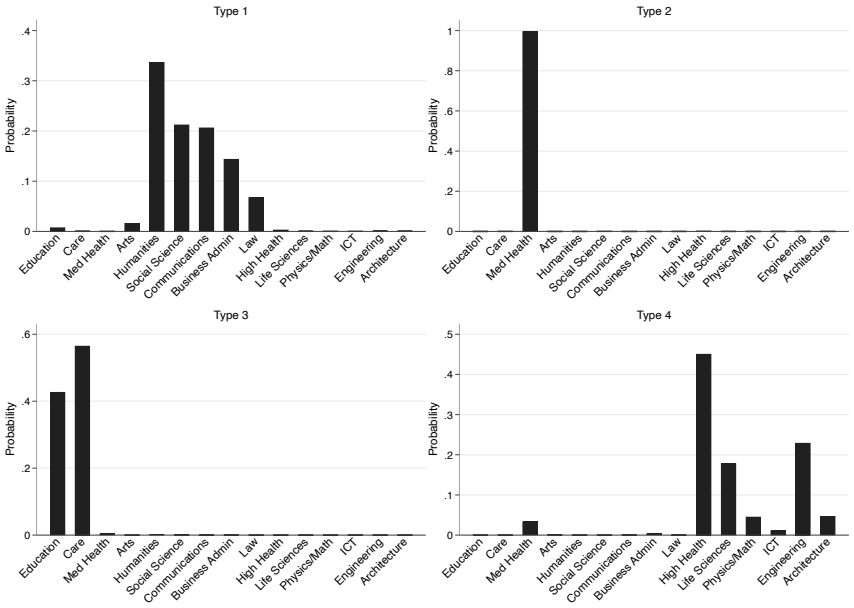
$$\frac{\partial \log \pi_m}{\partial \log V_m} = \theta \times (1 - \pi_m) \times V_m.$$

While we never directly estimate V_m, we can approximate the magnitude using $V \approx -C_o^e + \frac{w_o - \nu \log \pi_{oo}}{1-\beta}$, which would be exact in a stationary economy. Using the average values of w and π_{oo}, this number is on the order of 13. There are 15 fields so that $\pi_m \approx 1/15$. Plugging these numbers in, the elasticity for the least elastic type is .24, while for the most elastic type it is .84. On the one hand, this suggests a rather high responsiveness to shocks compared to other estimates in the literature (Wiswall and Zafar, 2014). However, it still suggests a relatively small response to shocks—a 10% increase in the returns to a field would only increase entry into that field by about 8%, which would amount to half of a percentage point off a base of $1/M$.

In addition to estimates of the responsiveness of students to returns to fields, we also estimate heterogeneity in preferences for fields. We define the non-pecuniary component of preferences to be

$$\vartheta_{lm} + \upsilon \log N_m,$$

Figure 8: Preferences Over Field of Study



Notes: The figure plots the estimated preferences over field of study by latent types. The preferences are computed by summing the estimates of the non-pecuniary components in the utility function (12), i.e., $\vartheta_{lm} + v \log N_m$.

and use the time average of $\log N_m$ over years below. To understand the importance and heterogeneity of these factors, we plot the probability that each type of student would choose field m if the returns across all fields were equalized. Figure 8 contains our results.

The four types of students cluster on very distinct sets of fields, similar to the clustering observed in the empirical joint distribution of first and second choices displayed in Table 1. Type 1 individuals cluster in the humanities and social sciences, while Type 4 individuals cluster in traditionally "STEM" fields such as life sciences, medicine, physics, and math. Types 2 and 3 cluster into medium health (e.g., nursing) degrees, as well as education and care degrees. As noted above, Type 2 individuals are ostensibly "elastic" given their high level of θ. However, they actually inelastically demand medium health programs, and so θ is largely meaningless for this type, and unlikely to be identified.

The clustering patterns have important implications for our counterfactuals—while students respond to income changes to some degree, they still appear to largely cluster in a narrow set of fields, suggesting that even large shocks may induce only small changes in the distribution of comparative advantage across types. We now turn to exploring this quantitatively in our counterfactuals.

7 Counterfactuals

This section first outlines the remaining ingredients of the model, and concludes with two counterfactual experiments.

7.1 Closing the Model

We briefly describe the labor demand side of the model. More details are available in Appendix C.1. In what follows, $s \in \mathcal{S}$ index industries, and D and F refer to Denmark and Foreign, respectively.

There is a final non-traded good from which agents consumed. The final good aggregate industrial outputs, as well as domestic and foreign varieties in a nested-CES fashion. In the top nest, the final good aggregator, Y_t, is given by

$$Y_t = \left(\sum_s \mu_s^{\frac{1}{\sigma}} Y_{st}^{\frac{\sigma-1}{\sigma}} \right)^{\frac{\sigma}{\sigma-1}}, \tag{17}$$

where Y_{st} are industrial outputs and μ_s is an exogenous demand shifter for industry s. Industrial outputs are themselves non-traded aggregates of domestic and foreign varieties given by:

$$Y_{st} = \left(\left(Y_{st}^D\right)^{\frac{\eta-1}{\eta}} + \left(Y_{st}^F\right)^{\frac{\eta-1}{\eta}} \right)^{\frac{\eta}{\eta-1}} \tag{18}$$

where Y_{st}^D is local absorption of domestic (i.e., Danish) output and Y_{st}^F are imported varieties for an aggregate foreign output, with P_{st}^D and P_{st}^F being their appropriately defined price indices. The elasticity of substitution on industrial outputs, σ, and on domestic versus foreign varieties, η, obey the standard regularity condition $\eta > \sigma > 0$.

In addition to consumer demand at home, there is demand for domestic output abroad. We assume that export demand in industry s is exogenously determined by the following demand curve:

$$X_{st} = A_s^F \left(P_{st}^D\right)^{-\eta}, \tag{19}$$

where A_s^F is a demand shifter.

Domestic production in each industry s is carried out by a large number of perfectly competitive production units. Each unit produces domestic output Q_{st} by aggregating production

and non-production services. Accordingly, we categorize occupations into production and non-production groups, denoted by $\mathcal{O}^P$ and $\mathcal{O}^N$, respectively. Production takes place according to the following CES function:

$$Q_{st} = \left[\left(A_s^N\right)^{\frac{1}{\alpha}} \left(L_{st}^N\right)^{\frac{\alpha-1}{\alpha}} + \left(\left(A_s^P\right)^{\frac{1}{\rho}} \left(L_{st}^P\right)^{\frac{\rho-1}{\rho}} + \left(A_s^M\right)^{\frac{1}{\rho}} \left(Q_{st}^M\right)^{\frac{\rho-1}{\rho}} \right)^{\frac{\rho}{\rho-1}\frac{\alpha-1}{\alpha}} \right]^{\frac{\alpha}{\alpha-1}}, \tag{20}$$

where L_{st}^N is the aggregate output by non-production occupation services, L_{st}^P is the aggregate output by non-production occupation services, Q_{st}^M is the aggregated imported material inputs, A_s^N, A_s^P, A_s^M and are the productivities of non-production, production services and imported inputs, respectively, $\alpha > 0$ is the elasticity of substitution between L_{st}^N and the the production aggregate within each industry, and $\rho > 0$ is the elasticity of substitution between L_{st}^N and Q_{st}^M within each industry.

We assume that production and non-production inputs are themselves Cobb-Douglas aggregates over occupations, for $f = \{N, P\}$

$$L_{st}^f = \prod_{o \in \mathcal{O}^f} \left(\frac{L_{sot}^f}{\beta_{so}} \right)^{\beta_{so}}, \tag{21}$$

where $\sum_{o \in \mathcal{O}^f} \beta_{so} = 1$, and H_{sot} is the efficiency units (human capital) in each occupation within each industry. Since workers are indifferent across sectors within an occupation, total human capital within an occupation is given by $L_{ot} = \sum_s L_{sot}$.

Finally, consistent with the labor supply model, workers spend their entire income in the period it is earned. The aggregate expenditure is therefore equivalent to the aggregate income:

$$E_t = \sum_o L_{ot} W_{ot}.$$

We provide details on labor demand derivation in Appendix C.1. However, to illustrate the implications of the model for the impact of offshoring price shocks on labor demand, we aggregate across sectors and totally log differentiate. Denoted by $z = d \log Z$, we obtain the following

expressions for labor demand:

$$\ell_{ot}^{N} = \sum_{s} \pi_{sot} \times \left[\alpha p_{st}^{D} - w_{ot} + S_{st}^{x} x_{st} + (1 - S_{st}^{x}) y_{st}^{D} - \alpha w_{st}^{N} \right]; \tag{22}$$

$$\ell_{ot}^{P} = \sum_{i} \pi_{sot} \times \left[\alpha p_{st}^{D} - w_{ot} + S_{st}^{x} x_{st} + (1 - S_{st}^{x}) y_{st}^{D} + (\rho - \alpha) p_{st}^{P} - \rho w_{st}^{P} \right]. \tag{23}$$

Offshoring shock affects demand of nonproduction workers through the industrial output demand channel. For production workers, however, in addition to the demand channel, offshoring shock also affects demand of production workers through the substitution between offshored inputs and production labor ($\alpha - \rho > 0$).

7.1.1 Demand Side Calibration

Table 4 summarizes the full set of parameters to be calibrated in order to close the model. All production coefficients are estimated as the time-averaged expenditure shares. We detail the construction of price indices in the appendix.

Table 4: Labor Demand Parameters and Variables

Parameters	Meaning	Value/Data Source
β_{so}^{P}, β_{so}^{N}	Labor inputs coefficients	IDA
P_{st}^{M}	Offshored input prices	UHDI, Broda and Weinstein (2006)
P_{st}^{F}	Imported good prices	UHDI, Broda and Weinstein (2006)
w_{ot}	Occupation wage	IDA
ρ, α	EOS in production	Doraszelski and Jaumandreu (2018)
σ	EOS in final good	Atalay (2017)
η	EOS btwn domestic and foreign	Simonovska and Waugh (2014)

Total wage bills in occupation-industry pairs identify the labor input coefficients. Danish customs data provides provide and quantity level information on goods sold to Denmark by each trading partner, which can be fed into the procedure of Broda and Weinstein (2006) to calculate the price indices for offshored inputs and imported varieties.[15] We set the trade elasticity as 4, following Simonovska and Waugh (2014). We obtain raw wage data from the Integrated Database for Labour Market Research (IDA) at Statistics Denmark. Occupation wages are computed by first running a mincerian regression and then divide the residuals by the computed human capital using (4).

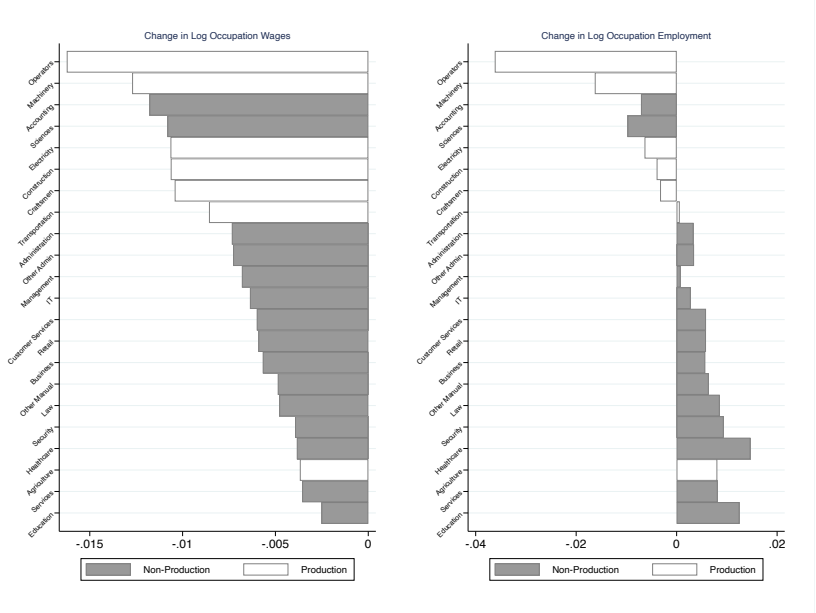
[15] We follow the approach in Feenstra (1994) to correct for changes in product varieties across time.

7.2 Counterfactual Analysis

Using data for 2018 as the initial period, the counterfactual experiments hold worker preferences and technology shocks fixed and compares workers across two simulated economies: one subject to hypothetical shocks, and the other not. Our current emphasis is on analyzing comparisons of steady states. Meanwhile, the investigation of transition dynamics is in progress.

In the first experiment, we study the potential labor market impact resulting from a globalization shock. The scenario is motivated by the prospect of further intensification of foreign competition in goods markets and production processes. To illustrate this, we operationalize this change by applying a reduction in the costs of imported goods and offshored inputs by 20% and 30%, respectively. The reduction in costs for imported goods is expected to affect workers across all sectors, whereas the decrease in costs for offshored inputs will differentially affect occupations that are more substitutable with these inputs. The labor market adjustment to the shock takes the form of changes in wages, a resorting of workers across occupations within a field, and shifts in students' choices of major, reflecting the changing demand for specific skills and occupations.

Figure 9: Globalization shock: Change across occupations in wage (left) and in employment (right)

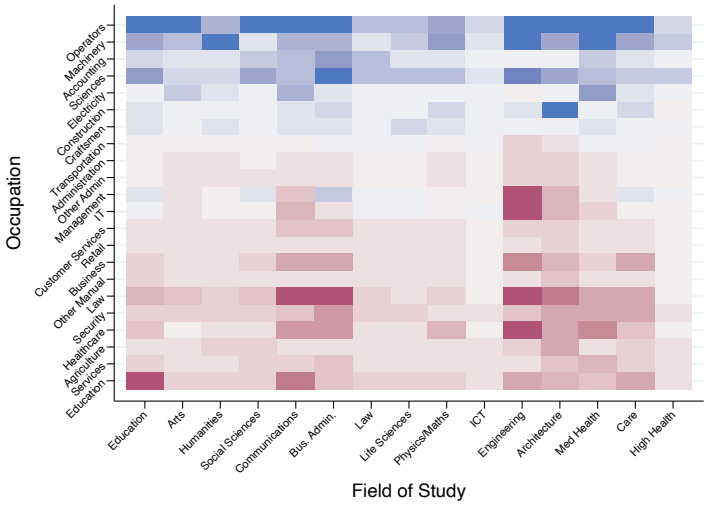


We begin by examining the impact of the globalization shock on changes in occupation wages and employment. Figure 9 plots, on the x-axis, the log change in average occupation wages in the left panel and the log change in occupation employment in the right panel. Occupations are sorted on the y-axis according to the magnitude of wage changes. In response to increased for-

eign competition, average wages fall across all occupations, ranging from a 0.25% decrease in Education occupation to a 1.6% decrease among Operators. In general, production occupations experience greater wage reductions due to their higher substitutability with offshored inputs. The right panel of Figure 9 shows that changes to occupation wages largely drive the shifts in occupation employments. Specifically, the occupation that experiences the greatest negative growth, at -3.6%, is Operator, whereas the occupations witnessing the largest positive growth are Healthcare, 1.5%, and Education, at 1.3%.

These outcomes, which are the focus of much previous literature, capture differences across occupations in labor market impacts. They do not reveal variation within occupations, particularly how exposure and resilience to shocks can differ based on fields of study. Figure 10 illustrates the variation in how different fields respond to the globalization shock by changing their occupation distribution. The vertical axis, which lists occupations, is organized according to the changes in occupation wages as displayed in Figure 9. Across all fields, there is a shift away from occupations experiencing negative wage growth (shown in blue) and towards those with positive wage growth (shown in red). Nonetheless, the patterns of occupational reallocation significantly vary across fields. For instance, fields such as High health, ICT, and Law exhibit uniform reallocation patterns across occupations, while fields like Education, Communications, Business Administration, and Engineering display much more varied reallocation behaviors.

Figure 10: Globalization shock: Change across fields in occupation sorting



Notes: Occupations are sorted based on change in occupation wage, as in Figure 9.

The distinct processes of occupation reallocation across various fields also impact the real re-

turns for these fields, which are displayed in Figure 11. Somewhat surprisingly, the fields experiencing the lowest growth in returns include both those traditionally associated with production occupations, such as ICT and Engineering, and those not typically specialized in these areas, like Business Administration and Education. This result underscores the importance of considering both demand-side shocks and supply-side responses to fully understand the distributional consequences of labor market disruptions.

Figure 11: Globalization shock: Change across fields in returns

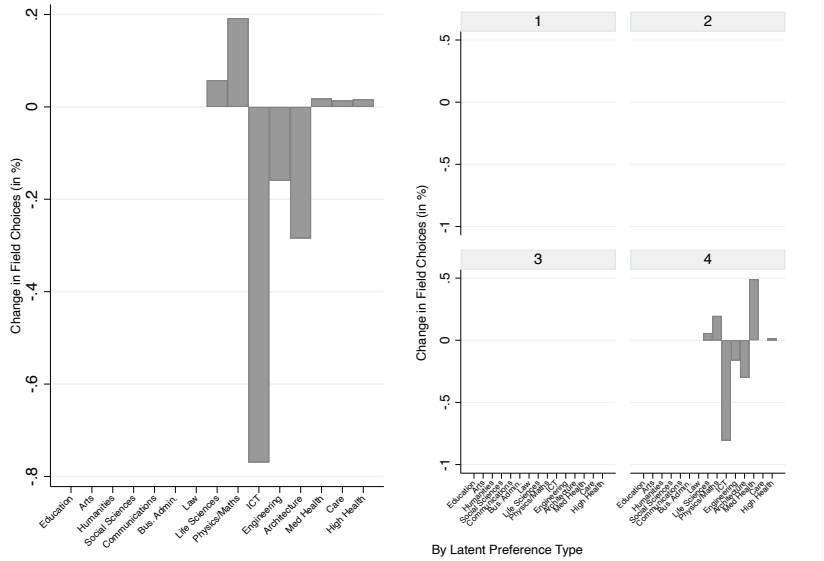


Our analysis further explores the changes in students' field choices in response to the labor demand shocks. The left panel of Figure 12 plots these shifts, indicating that, relative to changes in the returns, the shifts in field choices are very small and primarily observed within STEM-related fields. The right panel breaks down the overall trend across different latent preference groups. Notably, no reallocation occurs in Types 1, 2, and 3 , which are driven by their strong preferences in non-STEM fields. Conversely, the entirety of the reallocation is driven by Type 4, an elastic type with strong preferences for STEM fields.

Overall, we find that in the short run, majors account for XX% of the variation in the gains and losses from trade across workers. The importance of majors continues in the long-run, because we find that students do not switch between STEM and non-STEM majors. This latter result hinges on our modeling of latent heterogeneity in preferences—a model ignoring these preferences would imply too much reallocation and lower dispersion in the gains from trade across workers.

In a second planned counterfactual, we will consider implications of an education policy that promotes entry into STEM field.

Figure 12: Globalization Shock: Change in field choices



8 Conclusion

TBD...

References

Altonji, J.G., P. Arcidiacono, and A. Maurel, "Chapter 7 - The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects," in Eric A. Hanushek, Stephen Machin, and Ludger Woessmann, eds., *Handbook of the Economics of Education*, Vol. 5, Elsevier, 2016, pp. 305–396.

Altonji, Joseph G., Erica Blom, and Costas Meghir, "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, 2012, *4* (1), 185–223.

_ , _ , and _ , "Heterogeneity in Human Capital Investments: High School Curriculum, College Major, and Careers," *Annual Review of Economics*, 2012, *4* (1), 185–223.

Andrews, Rodney J, Scott A Imberman, Michael F Lovenheim, and Kevin M Stange, "The returns to college major choice: Average and distributional effects, career trajectories, and earnings variability," Technical Report, National Bureau of Economic Research 2022.

Arcidiacono, Peter, "Ability sorting and the returns to college major," *Journal of econometrics*, 2004, *121* (1-2), 343–375.

_ and Robert A. Miller, "CCP Estimation of Dynamic Models with Unobserved Heterogeneity," *Econometrica*, November 2011, *79* (6), 1823–1868.

_ , V. Joseph Hotz, Arnaud Maurel, and Teresa Romano, "Ex Ante Returns and Occupational Choice," *Journal of Political Economy*, 2020, *128* (12), 4475–4522.

Artuç, Erhan, Shubham Chaudhuri, and John McLaren, "Trade Shocks and Labor Adjustment: A Structural Empirical Approach," *American Economic Review*, June 2010, *100* (3), 1008–45.

Atalay, Enghin, "How Important Are Sectoral Shocks?," *American Economic Journal: Macroeconomics*, October 2017, *9* (4), 254–80.

Atkin, David, "Endogenous skill acquisition and export manufacturing in Mexico," *American Economic Review*, 2016, *106* (8), 2046–2085.

Autor, David H., David Dorn, and Gordon H. Hanson, "The China Syndrome: Local Labor Market Effects of Import Competition in the United States," *American Economic Review*, October 2013, *103* (6), 2121–68.

_ , _ , _ , and Jae Song, "Trade Adjustment: Worker-Level Evidence," *The Quarterly Journal of Economics*, 2014, *129* (4), 1799–1860.

Balinski, Michel and Tayfun Sonmez, "A Tale of Two Mechanisms: Student Placement," *Journal of Economic Theory*, 1999, *84* (1), 73–94.

Beggs, S, S Cardell, and J Hausman, "Assessing the potential demand for electric cars," *Journal of Econometrics*, 1981, *17* (1), 1–19.

Berry, Steven, James Levinsohn, and Ariel Pakes, "Automobile Prices in Market Equilibrium," *Econometrica*, 1995, *63* (4), 841–90.

Blanchard, Emily and William Olney, "Globalization and human capital investment: Export composition drives educational attainment," *Journal of International Economics*, 2017, *106* (C), 165–183.

Bleemer, Zachary and Aashish Mehta, "Will Studying Economics Make You Rich? A Regression Discontinuity Analysis of the Returns to College Major," *American Economic Journal: Applied Economics*, April 2022, *14* (2), 1–22.

Broda, Christian and David E. Weinstein, "Globalization and the Gains from Variety," *The Quarterly Journal of Economics*, May 2006, *121* (2), 541–585.

Burstein, Ariel, Gordon Hanson, Lin Tian, and Jonathan Vogel, "Tradability and the Labor-Market Impact of Immigration: Theory and Evidence From the United States," *Econometrica*, 2020, *88* (3), 1071–1112.

Caliendo, Lorenzo, Luca David Opromolla, Fernando Parro, and Alessandro Sforza, "Goods and factor market integration: A quantitative assessment of the EU enlargement," *Journal of Political Economy*, 2021, *129* (12), 3491–3545.

_ , Maximiliano Dvorkin, and Fernando Parro, "Trade and Labor Market Dynamics: General Equilibrium Analysis of the China Trade Shock," *Econometrica*, 2019, *87* (3), 741–835.

Dix-Carneiro, Rafael, "Trade Liberalization and Labor Market Dynamics," *Econometrica*, 2014, *82* (3), 825–885.

_ and Brian K. Kovak, "Trade Liberalization and Regional Dynamics," *American Economic Review*, October 2017, *107* (10), 2908–46.

_ , João Paulo Pessoa, Ricardo Reyes-Heroles, and Sharon Traiberman, "Globalization, trade imbalances, and labor market adjustment," *The Quarterly Journal of Economics*, 2023, *138* (2), 1109–1171.

Doraszelski, Ulrich and Jordi Jaumandreu, "Measuring the Bias of Technological Change," *Journal of Political Economy*, 2018, *126* (3), 1027–1084.

Fack, Gabrielle, Julien Grenet, and Yinghua He, "Beyond Truth-Telling: Preference Estimation with Centralized School Choice and College Admissions," *American Economic Review*, April 2019, *109* (4), 1486–1529.

Feenstra, Robert C., "New Product Varieties and the Measurement of International Prices," *The American Economic Review*, 1994, *84* (1), 157–177.

Greenland, Andrew and John Lopresti, "Import exposure and human capital adjustment: Evidence from the US," *Journal of International economics*, 2016, *100*, 50–60.

Hastings, Justine S, Christopher A Neilson, and Seth D Zimmerman, "Are Some Degrees Worth More than Others? Evidence from college admission cutoffs in Chile," Working Paper 19241, National Bureau of Economic Research July 2013.

Helpman, Elhanan, Oleg Itskhoki, Marc-Andreas Muendler, and Stephen J. Redding, "Trade and Inequality: From Theory to Estimation," *The Review of Economic Studies*, 06 2016, *84* (1), 357–405.

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

Humlum, Anders, "Robot Adoption and Labor Mafrket Dynamics," Working Paper, University of Chicago Booth School of Business 2021.

Hummels, David, Rasmus Jorgensen, Jakob Munch, and Chong Xiang, "The Wage Effects of Offshoring: Evidence from Danish Matched Worker-Firm Data," *American Economic Review*, June 2014, *104* (6), 1597–1629.

Keane, Michael and Kenneth Wolpin, "The Career Decisions of Young Men," *Journal of Political Economy*, 1997, *105* (3), 473–522.

Kinsler, Josh and Ronni Pavan, "The Specificity of General Human Capital: Evidence from College Major Choice," *Journal of Labor Economics*, 2015, *33* (4), 933–972.

Kirkeboen, Lars J, Edwin Leuven, and Magne Mogstad, "Field of study, earnings, and self-selection," *The Quarterly Journal of Economics*, 2016, *131* (3), 1057–1111.

Kleinman, Benny, Ernest Liu, and Stephen J Redding, "Dynamic spatial general equilibrium," *Econometrica*, 2023, *91* (2), 385–424.

Leighton, Margaret and Jamin D. Speer, "Labor market returns to college major specificity," *European Economic Review*, 2020, *128*, 103489.

Luflade, Margaux, "The Value of Information in Centralized School Choice Systems," *Mimeo*, 2019.

Porzio, Tommaso, Federico Rossi, and Gabriella Santangelo, "The Human Side of Structural Transformation," *American Economic Review*, August 2022, *112* (8), 2774–2814.

Simonovska, Ina and Michael E Waugh, "The elasticity of trade: Estimates and evidence," *Journal of International Economics*, 2014, *92* (1), 34–50.

Traiberman, Sharon, "Occupations and Import Competition: Evidence from Denmark," *American Economic Review*, December 2019, *109* (12), 4260–4301.

Wiswall, Matthew and Basit Zafar, "Determinants of College Major Choice: Identification using an Information Experiment," *The Review of Economic Studies*, 12 2014, *82* (2), 791–824.

_ **and** _ , "Determinants of college major choice: Identification using an information experiment," *The Review of Economic Studies*, 2015, *82* (2), 791–824.

Wooldridge, Jeffrey M, "Simple solutions to the initial conditions problem in dynamic, nonlinear panel data models with unobserved heterogeneity," *Journal of applied econometrics*, 2005, *20* (1), 39–54.

A Additional Results

In this appendix, we discuss several additional results regarding the labor market estimation. We also present tables of all estimated parameters.

A.1 Labor Market Estimates Discussion

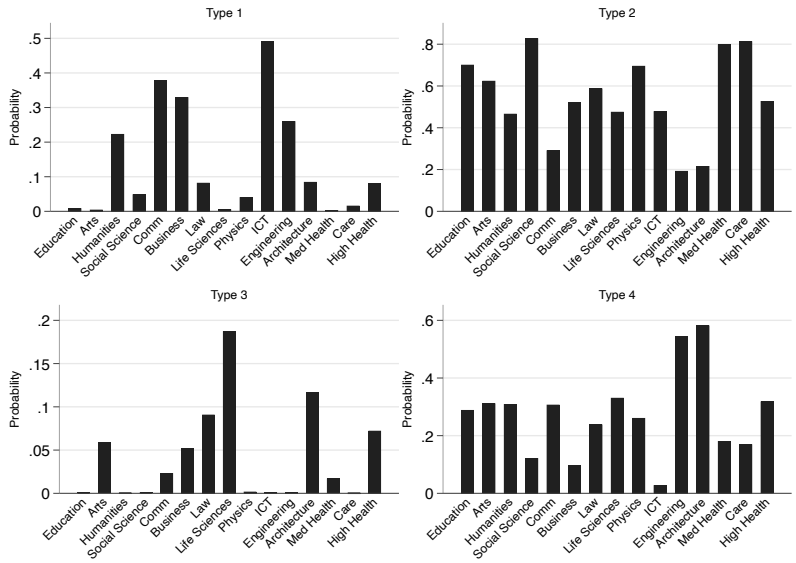
In this section, we briefly discuss the patterns of latent heterogeneity in the labor market. Broadly, latent types govern both absolute advantage through the α_k term, and comparative advantage through the $\alpha_{k,z}$ terms in (4). Turning to absolute advantage, Table 5 displays the value of α_k for each type k, while Figure 13 displays the conditional probability each major is of type k, $P(k|m)$.

Table 5: Absolute Advantage ($\exp(\alpha_k)$)

Type 1	Type 2	Type 3	Type 4
1.00	0.66	1.58	0.11

There are substantial differences in absolute advantage across types. Type 3 workers, holding everything else constant, earn *fifteen* times type 4 workers. While, this does not account for comparative advantage, nor the absolute advantage of majors,i it reflects the substantial dispersion in earnings, even within occupation and field of study.

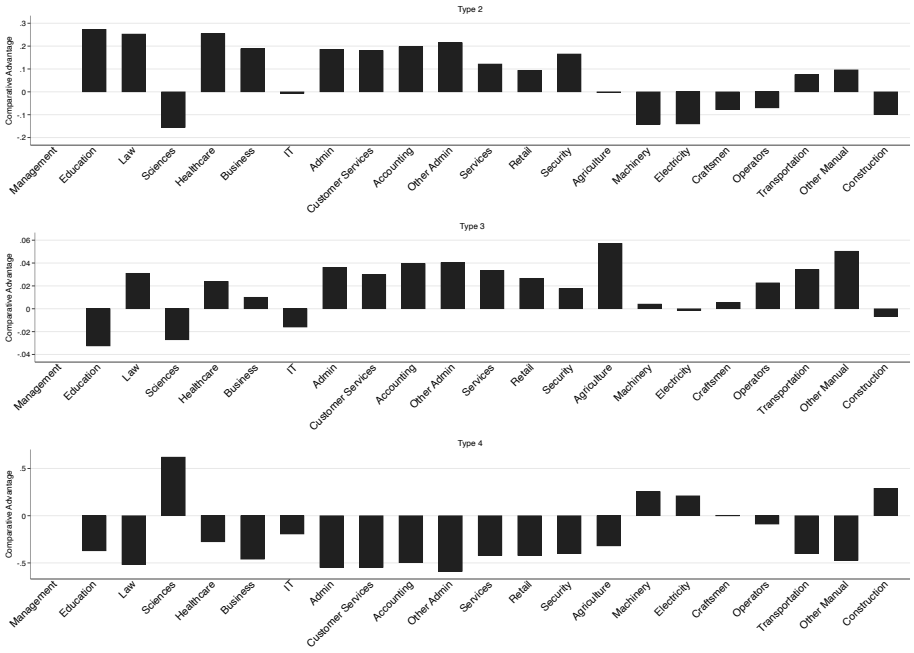
Figure 13: Type Probabilities By Major



Turning to the type probabilities, the most common type is Type 2, and with few exceptions,

the probability that a worker is type 2 given any major is well over 40%. The second most common type is Type 4. For both of these types, again with a few exceptions, the probabilities are somewhat flat across fields of study. Type 1 and Type 3—both higher-earnings types—are substantially more concentrated on a few fields of study. For example, $P(\text{Type } 1|m = \text{ICT})$ is nearly 50%, but is below 10% (and often below 5%) for 2/3 of fields. For Type 3, disparities are particularly pronounced.

Figure 14: Comparative Advantage by Type



To discuss comparative advantage, rather than focus on individual values of the $\alpha_{k,z}$ terms, we focus on the composite term,

$$\alpha_{ko} \triangleq \sum_{z=1}^{Z} \alpha_{k,z}\zeta_{z,o}.$$

To describe comparative advantage, we plot this relative to type 1 in occupation 1 ("Management"):

$$CA_{ko} = \frac{\exp\left(\alpha_{k,o} - \alpha_{1,o}\right)}{\exp\left(\alpha_{k,Man} - \alpha_{1,Man}\right)} - 1. \tag{24}$$

Figure 14 plots equation (24) for each type k and occupation o.

To interpret the magnitudes, our comparative advantage measure is how much more (or less) a worker earns in occupation o relative to management than a type 1 worker. So, for example, looking at Education of Type 2 workers, the value of .28 suggests that Type 2 workers earn 28% more as teachers relative to managers than Type 1 workers. This is the same order of magnitude

as the absolute advantage measure, suggesting comparative advantage is an important source of earnings dispersion and sorting.

With this in mind, while α_{ko} is a very high dimensional object, a few clear patterns emerge. Type 2 workers have comparative advantage in occupations that are more socially-oriented (e.g., Education, Law, Business), and less technically oriented (e.g., Sciences, IT, and manufacturing-oriented occupations). Type 3 workers display no real comparative advantage differences relative to type 1—notice that the scale of the y-axis is an order of magnitude smaller than for the other types. Finally, Type 4 workers tend to have large comparative advantage in the sciences, and more technical occupations such as machinery. This lines up well with the fact that $P(4|m)$ is largest for engineering and architecture.The overall impression from Figure 14 is that different types are varied in their skills at different occupations, and that these magnitudes are quantitatively meaningful.

B Estimation Algorithms

B.1 Generating Data for Monte Carlo

This section assumes that one has already defined the model and solved for the parameters of the simulation. Subroutines for this procedure are defined elsewhere. Here we outline the loop to generate simulated data. We will index each simulated individual by i and calendar time by t.

Step 1. Simulate skill prices, w_{ot} and solve for the value functions and policy functions in the labor market.

Step 2. Given a cohort of size N_c and T periods, pre-allocate the following data matrices:

- DLM: an $N_c(T+1)T/2 \times 4$ (this will be excessive given death and graduation time) matrix where each row is (i, t, o_{it}, w_{it})
- DF: $N_cT \times 2$ vector that contains vector of i's field and test score
- DK: $N_cT \times 1$ vector that contains vector of i's unobserved labor market type
- DL: $N_cT \times 1$ vector that contains vector of i's unobserved ex ante type
- DR: $N_cT \times R$ matrix that contains each N

Step 3. For each i simulate a career path:

Step 3.a. Generate a test score s_i for i

Step 3.b. Generate a type for i from the conditional CDF of the type distribution given s_i:

$$l_i = 1 + \sum_l \mathbf{1}(e > Q_l(s_i)),$$

where e is a uniform random variable.

Step 3.c. For $r = 1, ..., R$ calculate the ranked list of choices for individual i:

- First calculate the conditional CDF of choices for field:

$$\tilde{Q}^F_{ir}(m) = \frac{\sum_{m'=1\setminus m_{-r}}^{m} \exp\left(\vartheta_{mt} + \vartheta_{lm} + \theta_l \sum_k q(k|s,m) V_{t+5}(k,m)\right)}{\sum_{m'=1\setminus m_{-r}}^{M} \exp\left(\vartheta_{mt} + \vartheta_{lm} + \theta_l \sum_k q(k|s,m) V_{t+5}(k,m)\right)}$$

- Second generate a draw from this distribution:

$$m_{r_i} = 1 + \sum_{m \in m_{-r}} \mathbf{1}(e > \tilde{Q}^F_{ir}(m)),$$

for a uniform random variable.

Step 3.d. Assign the agent the highest ranked m in their choice set conditional on s. If no choice meets the criterion remove them from the sample.

Step 3.e. Determine the agents type:

$$k_i = 1 + \sum_{m \in m_{-r}} \mathbf{1}(e > \tilde{Q}_i^K(m, s)),$$

where Q_k is type CDF for agent i with field m and score s, and e is a uniform random variable.

Step 3.f. Assign an initial occupation beginning in t+5 based on the initial occupation policy function, λ_{t+5}^e:

$$o_{it} = 1 + \sum_{o=1}^{|\mathcal{O}|} \mathbf{1}(e > \sum_{o'=1}^{o} \lambda_t^e(o, m_i, k_i)),$$

where e is a uniform random variable.

Step 3.g. Starting the agent at $t + 5$, simulate a labor market history

i. Determine if the agent exits by testing a uniform random variable against δ. If retire, continue to next i. Else...

ii. Draw next occupation and wage:

$$\begin{aligned} o_{it} =& 1 + \sum_{o'=1}^{|\mathcal{O}|} \mathbf{1}(e > \sum_{o''=1}^{o''} \lambda_t(o' | o_{it-1}, k, m)) \\ w_{it} =& w_{o_{it}t} \times H(o_{it}, k, m) \end{aligned}$$

B.2 EM Algorithm Implementation Details

Inputs into the EM Algorithm are the following:

File Name	Description	Size
iDex	Key for *worker* for each i and t in panel	$NT \times 1$
tDex	Key for *calendar year* for each i and t in panel	$NT \times 1$
ptDex	Worker's observation within panel for each i and t	$NT \times 1$
field	Each worker's field	$N \times 1$
occupation	Each worker's occupation at time t	$NT \times 1$
logW	Each worker's log wage at time t	$NT \times 1$
zeta	Occupational characteristics	$\lvert\mathcal{O}\rvert \times Z$

These are the files read in by the Algorithm, however they are an incomplete list of what is needed for the Algorithm to run.

Step 1. Initialize the EM algorithm.

- Set types by k-means clustering on wages, ignoring occupations (as one possibility)
- Assign initial probabilities by,

$$q_{ik} = \begin{cases} 0.95 & \text{if clustered into } k \\ 0.05 & \text{if clustered into } k \end{cases}$$

- Construct the following sparse matrices *once*:
  - **XOT**: $NT \times |\mathcal{O}|$ matrix of occ-year dummies
  - **XOM**: $NT \times (M|\mathcal{O}| - M)$ matrix of occ-field dummies
  - For each k construct a matrix **XKZ** that is $NT \times (K-1) \times Z$ with all zeros for $k = 1$ (this is a normalization) and for $k \geq 2$:

$$\mathbf{XKZ}(k)(:, Z * (k-2) + Z) = \zeta_{o_{it},z}$$

  - Finally, construct the design matrix for each type:

$$\mathbf{XTEMP}(k) = [\mathbf{XOT}, \mathbf{XOM}, \mathbf{XKZ}(k)]$$

Note: In matlab this can be stored as a cell or multi-dimensional in array. In other

languages, a multidimensional array. The notation here is Matlab's cell. This matrix does *not* change with iterations—only the weight matrix is updated.

- Estimate initial parameters for wages.
  (a) Preallocate the covariance matrix, **XtX**, to a $|\mathcal{O}| \times T + (M-1) \times |\mathcal{O}| + (K-1) \times Z$ square matrix of zeros.
  (b) Preallocate the covaraince matrix between Y and X, **XtY**, to a $|\mathcal{O}| \times T + (M-1) \times |\mathcal{O}| + (K-1) \times Z$ length vector.
  (c) For $k = 1, ..., K$...
    i. Generate the weight matrix,

$$W = diag(q_{ik})$$

    ii. Update the RHS covariance matrix:

$$\mathbf{XtX} = \mathbf{XtX} + \mathbf{XTEMP}(k)'W\mathbf{XTEMP}(k)/(NT)$$

    iii. Update the LHS covariance matrix

$$\mathbf{XtY} = \mathbf{XtY} + \mathbf{XTEMP}(k)'W[\log w]/(NT)$$

  (d) Estimate the Mincer regression parameters,

$$\beta^{Mincer} = (XtX)^{-1}XtY$$

  (e) Estimate residual variance
    – For each k,

$$r_{itk} = \log w_{it} - XTEMP(k)_{it} \times \beta^{Mincer}$$

    – Estimate variance:

$$\hat{\sigma}^2 = \frac{1}{NT}\sum_k \sum_i \sum_t r_{itk}^2 q_{ik}$$

- Estimate transitions:

$$\hat{\lambda}(o'|o,t,m,k) = \frac{\sum_{i:o_{it-1}=o \cap o_{it}=o' \cap i \in m} q_{ik}}{\sum_{i:o_{it-1}=o \cap i \in m} q_{ik}}$$

- Estimate initial conditions:

$$\hat{\lambda}^0(o|t,m,k) = \frac{\sum_{i:o_{it}=o \cap i \in m} q_{ik}\mathbf{1}(p_i^1 = t)}{\sum_{i:i\in m} q_{ik}\mathbf{1}(p_i^1 = t)},$$

where p_i^1 is the first year the worker appears in the data.

Step 2. Initial value of EM objective, $J^{(0)} = 0$. **Begin E Step:**

Step 3. For iteration g, update the EM objective, J^g. For each k calculate the log likelihood, LL_k

$$LL_k = \sum_{i=1}^{N} q_{ik} \left[\sum_{t=1}^{T} \left(-r_{itk}^2/(2\sigma^2) - \log(\sigma^2)/2 + \log \hat{\lambda}(o_{it}|o_{it-1}, m_i, k) \right) + \log(\hat{\lambda}^0(o_{io}|t_0, m_i, k) \right]$$

And define,

$$LL = \sum_k LL_k$$

Finally add cdf term:

$$J = \sum_k LL_k + \sum q_k \log(q_k),$$

where $q_k = \frac{1}{N}\sum_i q_{ik}$.

Step 4. Update weights,

$$q_{ik} = \frac{\exp(LL_{ik})q_k}{\sum_{k'} \exp(LL_{ik'})q_{k'}}$$

Begin M Step:

Step 5. Update unconditional weights,

$$q_k = \frac{1}{N}\sum_i q_{ik}$$

Step 6. Update Mincer Parameters

(a) For $k = 1, ..., K$...

i. Generate the weight matrix,

$$W = diag(q_{ik})$$

ii. Update the RHS covariance matrix:

$$\mathbf{XtX} = \mathbf{XtX} + \mathbf{XTEMP}(k)'W\mathbf{XTEMP}(k)/(NT)$$

iii. Update the LHS covariance matrix

$$\mathbf{XtY} = \mathbf{XtY} + \mathbf{XTEMP}(k)'W[\log w]/(NT)$$

(b) Estimate the Mincer regression parameters,

$$\beta^{Mincer} = (XtX)^{-1}XtY$$

(c) Update residual variance

- For each k,

$$r_{itk} = \log w_{it} - XTEMP(k)_{it} \times \beta^{Mincer}$$

- Estimate variance:

$$\hat{\sigma}^2 = \frac{1}{NT}\sum_k \sum_i \sum_t r_{itk}^2 q_{ik}$$

Step 7. Update transitions:

$$\hat{\lambda}(o'|o,t,m,k) = \frac{\sum_{i:o_{it-1}=o \cap o_{it}=o' \cap i \in m} q_{ik}}{\sum_{i:o_{it-1}=o \cap i \in m} q_{ik}}$$

Step 8. Update initial conditions:

$$\hat{\lambda}^0(o|t,m,k) = \frac{\sum_{i:o_{it}=o \cap i \in m} q_{ik}\mathbf{1}(p_i^1 = t)}{\sum_{i:i \in m} q_{ik}\mathbf{1}(p_i^1 = t)},$$

Step 9. Step 8: If $|J^g - J^{g-1}| < tol$, where g is iteration, stop. Else, go to Step 2.

Outputs to be recovered from β^{Mincer}:

File Name	Description	Size
$\hat{\mu}_{ot}$	Skill Prices	$\|\mathcal{O}\| \times T$
$\hat{\alpha}_{om}$	Occ-Field FE	$\|\mathcal{O}\| \times M$
$\hat{\alpha}_{kz}$	Type-Task Productivity	$K \times Z$

Additional outputs to save:

File Name	Description	Size
$\hat{\lambda}_{oo}$	Transition matrix estimate	$\|\mathcal{O}\| \times \|\mathcal{O}\| \times T \times M \times K$
$\hat{Q}_K$	Matrix of type weights	$NT \times K$

B.3 Stage 2 of Labor Supply

To derive (10),

$$\log\left[\frac{\lambda_t(o'|o,k,m)}{\lambda_t(o|o,k,m)}\right] = -\frac{1}{\nu}\Big(-C_{oo'km} + w_{o't}H_{o'}(k,m) - w_{ot}H_o(k,m) + \beta\big(V_{t+1}(k,m,o') - V_{t+1}(k,m,o)\big)\Big),$$

and

$$\begin{aligned}\log\left[\frac{\lambda_{t+1}(o'|o',k,m)}{\lambda_{t+1}(o'|o,k,m)}\right] =& -\frac{1}{\nu}\Bigg(C_{oo'km} + w_{o't+1}H_{o'}(k,m) - w_{o't+1}H_{o'}(k,m)\\ &+\beta\big(V_{t+2}(k,m,o') - V_{t+2}(k,m,o')\big) + \big(V_{t+1}(k,m,o') - V_{t+1}(k,m,o)\big)\Bigg)\\ =& -\frac{C_{oo'km} + \big(V_{t+1}(k,m,o) - V_{t+1}(k,m,o')\big)}{\nu}\end{aligned}$$

Combining, we get,

$$\begin{aligned}\log\left[\frac{\lambda_t(o'|o,k,m)}{\lambda_t(o|o,k,m)}\right] + \beta\log\left[\frac{\lambda_{t+1}(o'|o't,k,m)}{\lambda_{t+1}(o'|o,k,m)}\right] =& -\frac{C_{oo'km}(1-\beta)}{\nu} +\\ &\frac{1}{\nu}\left(w_{o't}H_o(k,m) - w_{ot}H_o(k,m)\right) + \xi_{oo'kmt+1}\end{aligned},$$

In principle, to ensure positivity of the costs we apply an exponential transform. Specifically we estimate the following regression via unweighted NLLS:

$$\begin{aligned}\log\left(\frac{\hat{\lambda}_{oo'tmk}}{\hat{\lambda}_{ootmk}}\right) + \beta\log\left(\frac{\hat{\lambda}_{o'o't+1,mk}}{\hat{\lambda}_{oo't+1,mk}}\right) =& \exp\left(\gamma_{km} + \sum_z \gamma_z|\zeta_{o'z} - \zeta_{oz}|\right)\times(1-\beta)\tilde{\nu} -\\ &\tilde{\nu}\left(e^{\hat{\mu}_{o't}}\hat{H}(o',m,k) - e^{\hat{\mu}_{ot}}\hat{H}(o,m,k)\right)\end{aligned},$$

where $\tilde{\nu} = \nu^{-1}$ and $\hat{H}(o,m,k) = \exp(\alpha_{om} + \sum_z \alpha_{kz}\zeta_{zo})$.

To derive (11),

$$\log\left[\frac{\lambda^e_t(o|k,m)}{\lambda^e_t(1|k,m)}\right] = -\frac{1}{\nu}\left(C^e_{okm} + w_{ot}H_o(k,m) - w_{1t}H_1(k,m) + \beta\Big(V_{t+1}(k,m,o) - V_{t+1}(k,m,1)\Big)\right)$$

Combining with $\log\left[\frac{\lambda_{t+1}(o'|o',k,m)}{\lambda_{t+1}(o'|o,k,m)}\right]$ and after some algebra, we obtain (11).

To derive (13)

$$
\begin{aligned}
V_t^e(k,m) =& \nu \log\left(\sum_o \exp\left(\frac{-C^e_{okm} + w_{ot}H(o,k,m) + \beta V_{t+1}(o,k,m)}{\nu}\right)\right) \\
=& -C^e_{okm} + w_{ot}H(o,k,m) + \beta V_{t+1}(o,k,m) - \nu \log \lambda_t^e(o|k,m) \\
=& -C^e_{okm} + w_{ot}H(o,k,m) + \beta\left\{w_{o,t+1}H(o,k,m) + \beta V_{t+2}(o,k,m)\right\} - \nu \log \lambda_t^e(o|k,m) - \\
& \beta\nu \log \lambda_{t+1}(o|o,k,m) \\
=& -(1-\beta)C^e_{okm} + w_{ot}H(o,k,m) + \\
& \beta\left\{-C^e_{okm} + w_{o,t+1}H(o,k,m) + \beta V_{t+2}(o,k,m) - \nu \log \lambda^e_{t+1}(o|k,m)\right\} - \\
& \nu \log \lambda_t^e(o|k,m) - \beta\nu \log \frac{\lambda_{t+1}(o|o,k,m)}{\lambda^e_{t+1}(o|k,m)} \\
=& -(1-\beta)C^e_{okm} + w_{ot}H(o,k,m) + \beta V^e_{t+1}(k,m) - \nu \log \lambda_t^e(o,k,m) - \beta\nu \log \frac{\lambda_{t+1}(o|o,k,m)}{\lambda^e_{t+1}(o|k,m)}.
\end{aligned}
$$

B.4 Estimating Entry Costs

C Details on Counterfactuals

C.1 Labor Demand Model Details

Given wages and prices of imported inputs, the profit maximizing choice of input quantity and the zero profit condition yield demand for occupation o in industry s:

$$L_{sot}^P = \frac{\beta_{so}^P}{W_{ot}} \times A_s^P \times Q_{st} \times \left(P_{st}^D\right)^{\alpha} \times \left(P_{st}^P\right)^{\rho-\alpha} \times \left(W_{st}^P\right)^{-\rho}, \ \ \forall o \in \mathcal{O}^P; \tag{25}$$

$$L_{sot}^N = \frac{\beta_{so}^N}{W_{ot}} \times A_s^N \times Q_{st} \times \left(P_{st}^D\right)^{\alpha} \times \left(W_{st}^N\right)^{-\alpha}, \ \ \forall o \in \mathcal{O}^N; \tag{26}$$

and demand for imported goods is

$$Q_{st}^M = A_s^M \times Q_{st} \times \left(P_{st}^D\right)^{\alpha} \times \left(P_{st}^P\right)^{\rho-\alpha} \times \left(P_{st}^M\right)^{-\rho}, \tag{27}$$

where

$$Q_{st} = A_s^F \left(P_{st}^D\right)^{-\eta} + \mu_s E_t \times \left(\frac{P_{st}}{P_t}\right)^{1-\sigma} \times \frac{\left(P_{st}^D\right)^{-\eta}}{\left(P_{st}^D\right)^{1-\eta} + \left(P_{st}^F\right)^{1-\eta}}; \tag{28}$$

$$W_{st}^f = \prod_{o \in \mathcal{O}^f} w_{ot}^{\beta_{so}}, \quad f = \{P, N\}; \tag{29}$$

$$P_{st}^P = \left[A_s^P \left(W_{st}^P\right)^{1-\rho} + A_s^M \left(P_{st}^M\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}; \tag{30}$$

$$P_{st}^D = \left[A_s^N \left(W_{st}^N\right)^{1-\alpha} + \left(P_{st}^P\right)^{1-\alpha}\right]^{\frac{1}{1-\alpha}} \tag{31}$$

$$P_{st} = \left[\left(P_{st}^D\right)^{1-\eta} + \left(P_{st}^F\right)^{1-\eta}\right]^{\frac{1}{1-\eta}} \tag{32}$$

$$P_t = \left[\sum_s \mu_s \left(P_{st}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{33}$$

$$E_t = \sum_o w_{ot} \left(\sum_s L_{sot}\right) \tag{34}$$

C.2 Solution Methods

C.3 Algorithm: Steady-state equilibrium following shock

Consider shocks $\{P_{st}^{F,0}\} \to \{P_{st}^{F,1}\}$, $\{P_{st}^{M,0}\} \to \{P_{st}^{M,1}\}$. We use 0 superscripts to denote the initial steady state, and 1 superscripts to denote the final steady state.

Start from the initial steady state: $\{w_o^0\}$. We assume q_l and $\pi(k|m,s) = \frac{\exp(\delta_{km}+\beta_k s)}{\sum_{k'}\exp(\delta_{k'm}+\beta_{k'}s)}$ are constant. Denote relative changes in variable x by $\hat{x} = \frac{x^1}{x^0}$.

Step 1. Guess $\{w_o^1\}$

Step 2. Compute Bellman Equations by iterating on the following until convergence:

$$V^1(k,m,o) = \nu \log\left(\sum_{o'}\exp\left(\frac{-C_{oo'km} + w_{o'}^1 H_{o'}(k,m) + \beta V^1(k,m,o')}{\nu}\right)\right)$$

Step 3. Using $V^1(k,m,o)$, solve for the inclusive value of being type (k,m):

$$V^{e,1}(k,m) = \nu \log\sum_{o'}\exp\left(\frac{-C_{o,km}^e + w_{o'}^1 H_{o'}(k,m) + \beta V^1(k,m,o')}{\nu}\right)$$

Step 4. Compute the NPV for each field m

$$V^1(m|s) = \sum_k V^{e,1}(k,m)\pi(k|m,s).$$

Step 5. Compute $\{L_{km}^1\}$ by following the steps in the Monte Carlo simulation in B.1.

Step 5.a. For each individual i, generate a test score s_i for i

Step 5.b. Generate a type for i from the conditional CDF of the type distribution given s_i:

$$l_i = 1 + \sum_l \mathbf{1}(e > Q_l(s_i)),$$

where e is a uniform random variable.

Step 5.c. For $r = 1, ..., R$ calculate the ranked list of choices for individual i:

- Calculate the conditional CDF of choices for field:

$$\tilde{Q}_{ir}^F(m) = \frac{\sum_{m'=1\setminus m_{-r}}^{m}\exp\left(\vartheta_{mt} + \vartheta_{lm} + \theta_l V^1(m|s)\right)}{\sum_{m'=1\setminus m_{-r}}^{M}\exp\left(\vartheta_{mt} + \vartheta_{lm} + \theta_l V^1(m|s)\right)}$$

- Generate a draw from this distribution:

$$m_{r_i} = 1 + \sum_{m\in m_{-r}} \mathbf{1}(e > \tilde{Q}_{ir}^F(m)),$$

for a uniform random variable.

Step 5.d. Assign the agent the highest ranked m in their choice set conditional on s. If no choice meets the criterion remove them from the sample.

Step 5.e. Determine the agents type:

$$k_i = 1 + \sum_{m \in m_{-r}} \mathbf{1}(e > \tilde{Q}_i^K(m, s)),$$

where Q_k is type CDF for agent i with field m and score s, and e is a uniform random variable.

Step 5.f. Calculate the share of individuals belong to type (k, m), scale up to $\bar{L}$.

Step 6. Using $V^1(k, m, o)$ solve for λ_{km}^1 and $\lambda^{e,1}(o|k, m)$

$$\lambda^1(o'|o, k, m) = \frac{\exp\left(\frac{-C_{oo'km} + w_{o't} H_{o'}(k,m) + \beta V^1(k,m,o')}{\nu}\right)}{\sum_{o''} \exp\left(\frac{-C_{oo''km} + w_{o''t} H_{o''}(k,m) + \beta V^1(k,m,o'')}{\nu}\right)}$$

$$\lambda^{e,1}(o|k, m) = \frac{\exp\left(\frac{-C_{o,km} + w_{ot} H_o(k,m) + \beta V^1(k,m,o)}{\nu}\right)}{\sum_{o'} \exp\left(\frac{-C_{o,km} + w_{o't} H_{o'}(k,m) + \beta V^1(k,m,o')}{\nu}\right)}$$

Step 7. Compute $\hat{H}_o^{supply}$

Step 7.a. Construct $\{\Lambda_{km}^1\}$:

$$\Lambda_{km}^1 = \left[\begin{array}{c|c} 0 & \lambda_{km}^{e,1} \\ \hline \delta & (1-\delta) \times \lambda_{km}^1 \end{array}\right].$$

Step 7.b. Solve the system of linear equations,

$$(\Lambda_{km}^{1'} - I) v_{km}^1 = 0$$

Note: one can solve this by solving for the unit eigenvector and eigenvalue of $\Lambda_{km}^{1'}$ or through iteration on the system, $v_{km}^{(p+1),1} = \Lambda_{km}^{1'} v_{km}^{(p),1}$.

Step 7.c. Scale $v_{km}^1 = \begin{pmatrix} L_{km1}^1 \\ \dots \\ L_{kmO}^1 \end{pmatrix}$ so that $\sum_o L_{kmo}^1 = L_{km}^1$.

Step 7.d. Compute changes in human capital supply across occupations

$$\hat{H}_o^{supply} = \sum_{k,m} L_{kmo}^1 / H_o^0$$

Step 8. Compute changes in prices

$$\hat{W}_s^f = \prod_{o \in \mathcal{O}^f} \hat{W}_o^{\beta_{so}}; \quad f = P, N$$

$$\hat{P}_s^P = \left(S_{st}^P \left(\hat{W}_s^P \right)^{1-\rho} + (1 - S_{st}^P) \left(\hat{P}_s^M \right)^{1-\rho} \right)^{\frac{1}{1-\rho}}$$

$$\hat{P}_s^D = \left[S_{st}^N \left(\hat{W}_s^N \right)^{1-\alpha} + (1 - S_{st}^N) \left(\hat{P}_s^P \right)^{\frac{1-\alpha}{1-\rho}} \right]^{\frac{1}{1-\alpha}}$$

$$\hat{P}_s = \left[(1 - S_{st}^I) \left(\hat{P}_s^D \right)^{1-\eta} + S_{st}^I \left(\hat{P}_s^F \right)^{1-\eta} \right]^{\frac{1}{1-\eta}}$$

$$\hat{P} = \left(\sum_s S_{st} \left(\hat{P}_s \right)^{1-\sigma} \right)^{\frac{1}{1-\sigma}}$$

Step 9. Calculate aggregate expenditure (and income)

$$\hat{E} = \sum_o S_{ot} \hat{H}_o \hat{W}_o$$

Step 10. Calculate changes in domestic output

$$\hat{Q}_s = S_{st}^x \left(\hat{P}_s^D \right)^{-\eta} + (1 - S_{st}^x) \hat{E} \hat{P}_s^{\eta - \sigma} \hat{P}^{\sigma - 1} \left(\hat{P}_s^D \right)^{-\eta}$$

Step 11. Calculate a new guess for wages, by enforcing labor market clearing and output market clearing. Choose $\hat{W}_o$ such that

$$\begin{aligned} \hat{H}_o &= \sum_s \frac{L_{sot}}{H_{ot}} \hat{L}_{so} \\ &= \sum_s \frac{L_{sot}}{H_{ot}} \frac{1}{\hat{W}_o} \hat{Q}_s \left(\hat{P}_s^D \right)^{\alpha} \left(\hat{W}_s^N \right)^{-\alpha}, \end{aligned}$$

for all $o \in \mathcal{O}^N$; and

$$\begin{aligned}\hat{H}_o &= \sum_s \frac{L_{sot}}{H_{ot}} \hat{L}_{so} \\ &= \sum_s \frac{L_{sot}}{H_{ot}} \frac{1}{\hat{W}_o} \hat{Q}_s \left(\hat{P}_s^D\right)^{\alpha} \left(\hat{P}_s^P\right)^{\rho-\alpha} \left(\hat{W}_s^P\right)^{-\rho},\end{aligned}$$

for all $o \in \mathcal{O}^P$.