Field Choice, Skill Specificity, and Labor Market Disruptions*

Valerie Smeets
Aarhus University & CEPR

Lin Tian
INSEAD & CEPR

Sharon Traiberman
NYU & NBER

July 7, 2026

Abstract

We study how labor markets adjust to disruptions through the field-of-study choices of new cohorts. We develop a dynamic general-equilibrium model in which forward-looking students choose a field at college entry and later switch occupations, and estimate it on Danish administrative data. Fields differ in the comparative advantage and switching costs they give workers, and ignoring field choice understates the elasticity of occupational switching by a factor of 6.7. We then study a trade war and an AI shock to skills. Field choice matters when a shock changes relative field values. The trade war leaves these values unchanged, so field choice barely responds. The AI shock moves them apart, and re-choosing fields nearly doubles the lifetime income gain, by 1.36 percentage points. Policies that enhance flexibility in field choice raise lifetime income by 4.46 percentage points. This value barely changes under the trade war but falls under AI.

Keywords: Skill Specificity, Human Capital Investment, Trade War, Automation, Labor Market Disruption

JEL: F14, F16, J24, I24

*Contact: vas@econ.au.dk, lin.tian@insead.edu, st1012@nyu.edu. We thank Pol Antras, Jackie Chan, Chris Conlon, Rafael Dix Carneiro, Christopher Flinn, Matt Notowidigdo, Esteban Rossi-Hansberg, Martin Rotemberg, Dan Trefler, Tim Van-Zandt, Jonathan Vogel, Dan Waldinger, Matt Wiswall, and numerous seminar and conference participants for helpful comments and suggestions. We thank Daniel Rock for providing the data on AI technological shock. Access to the Danish register data used in this study has been provided by ECONAU, Aarhus University.

1 Introduction

Trade disruptions and rapid technological change are major labor market shocks that reshape occupational structures and skill demands. Understanding how societies adapt to them raises questions central to economics and policy: How effectively can workers mitigate adverse impacts? Who gains or loses, and how can policy influence these outcomes? We study these questions by examining how new cohorts of workers adapt their skills, particularly through their college field-of-study choices, to evolving labor market conditions.

Existing literature in international trade and macroeconomics has documented how *incumbent* workers adjust along various margins in response to trade, automation, and migration shocks (e.g., Dix-Carneiro, 2014; Caliendo and Parro, 2015; Burstein et al., 2020). However, it has paid less attention to how *new cohorts* respond through forward-looking human capital investments.[1] This omission is especially notable given findings from the labor literature that college fields of study differ markedly in both earnings potential and adaptability to economic shocks (e.g., Altonji et al., 2012a). These differences not only shape individual career trajectories, but also influence the aggregate distribution of skills. Thus, modeling the joint determination of education and career decisions is necessary for capturing how economies respond to shocks.

To this end, we develop and estimate a dynamic model that jointly captures educational choices by new entrants and subsequent occupational decisions by incumbent workers. Agents enter our model at high school graduation—having already decided whether to pursue a college degree—and select a field of study based on expected returns and idiosyncratic preferences. They subsequently make occupational decisions in each period, subject to switching costs. We use our model to perform several counterfactuals that answer the questions posed above. First, we examine how field-of-study decisions influence the aggregate and distributional consequences of labor market shocks, focusing on two specific scenarios: a global trade war characterized by increased import costs and reduced export demand, and an AI shock that lowers the value of the skills some fields provide while raising the value of others. Second, we analyze how increased flexibility over students' education choices, e.g., relaxing the admission restrictions on field choice, impacts long-term welfare and the economy's

[1] Recent exceptions include Ferriere et al. (2021) and Adao et al. (2024). The former documents increased college enrollment in U.S. regions exposed to trade shocks, particularly among wealthier households, and model skill acquisition as a key adjustment margin. The latter studies how skill specificity affects the speed of labor market adjustments to technological shocks. Our work differs by explicitly modeling the horizontal dimension of skill acquisition—field-of-study choices—and integrating it into a general equilibrium framework to assess broader labor market adjustments. There is a longer tradition studying human capital investment and labor market outcomes (a recent example is Alon et al. (2024)), but this has abstracted from the horizontally differentiated nature of fields of study.

adjustment to these shocks.

We leverage rich Danish administrative data that links workers to their college applications, including students' ranked program choices. Denmark's centralized, algorithmic college admissions system provides a useful setting for identifying student preferences separately from selection effects. This allows us to model how field choices respond to changes in expected labor market outcomes and to perform realistic counterfactual analyses.

Before turning to our model, we use this dataset to document two motivating facts. First, we exploit quasi-experimental variation in entry to different fields of study to demonstrate that students who enter different fields of study sort into different occupations, indicating that educational choices shape career outcomes. Second, we show that student demand for specific fields is positively correlated with future returns to those fields, suggesting that students are forward-looking in their educational decisions.

Motivated by these findings, our dynamic model explicitly captures two interconnected decision stages: education and labor supply. At the education stage, students decide which field of study to pursue considering both expected labor market earnings and non-pecuniary preferences.[2] These choices are shaped by both observable characteristics—such as test scores, high school coursework, and gender—and unobserved latent preference types. Upon graduation, individuals draw occupational productivity from field-specific distributions conditioned on their observables. This productivity is multidimensional, creating differences in individuals' skills, particularly comparative advantages, even among students who share fields of study and observables.

The second stage of the model captures individuals' labor supply decisions. After drawing their labor market productivity, individuals choose an initial occupation and may switch occupations in each subsequent period, subject to switching costs à la Artuç et al. (2010). New entrants, even with the same educational background, may ultimately make different career decisions and respond differently to changes in the labor market for two reasons: differences in comparative advantage and idiosyncratic shocks to switching costs.

We estimate the model in two stages, corresponding to the education and labor supply blocks. For the education block, we leverage Denmark's centralized admissions system and students' complete ranked program choices to estimate latent preference heterogeneity. Adapting methods from Agarwal and Somaini (2018) and Fack et al. (2019), we recover a distribution of the latent types

[2]We do not model the extensive margin of entry into a Bachelor's program. While this is an important adjustment mechanism (Ferriere et al., 2021), our focus is on the horizontal decision over different fields.

that rationalizes observed field rankings. By linking worker registry data and education data, we exploit time variation in returns across fields to estimate students' elasticity of field-choice with respect to expected income.

For the labor supply block of the model, we use a two-step approach following Arcidiacono and Miller (2011) and Traiberman (2019), first employing the Expectation-Maximization (EM) algorithm to recover comparative advantages and occupational transition rates, and then regressing these transition rates on wage differentials to obtain the remaining parameters as in Artuç et al. (2010). Intuitively, the panel structure of the data allows us to control for changes in labor supply over time, while the variation in wages induced by demand identifies labor supply elasticities.

Our estimation results point to an important role for field choice in the labor market. First, there is substantial heterogeneity across fields in how responsive workers are to wage shocks. Labor market conditions also influence students' choices: a 1% increase in the net present value (NPV) of a field raises entry into that field by 0.92% (or 0.14 percentage points, pp henceforth). Moreover, we highlight the role of flexibly modeling student preferences by comparing our results to a model that ignores latent heterogeneity and does not exploit students' full ranked preferences. The predictions of this simpler approach deviate substantially from observed behavior. Perhaps most importantly, accounting for both field choices and the resulting skill heterogeneity dramatically alters how we measure labor supply responsiveness. Ignoring this margin, as is standard in the majority of models of labor market adjustment in trade and macroeconomics, understates the elasticity of occupational switching by a factor of 6.7. This has first-order implications: models that ignore field-of-study choices severely underestimate labor mobility in response to wage shifts and misrepresent the true nature of labor market adjustment.

We close the model with a multi-sector, small open economy labor demand side, and use it to study how the economy absorbs two disruptions, a global trade war and an AI shock to the value of skills. Field choice matters when a disruption changes the relative value of fields. The trade war falls unevenly across occupations. But workers switch occupations throughout their careers, so a degree's value reflects the best occupation its holder can reach, not the wage in any single one. The uneven wage changes therefore become, in general equilibrium, a nearly common decline in the value of every field. As a result, field choice barely responds: only 0.5% of students change field, and the long-run loss of about 16% of lifetime income is the same whether or not new cohorts re-choose. The AI shock works the other way. It changes the value of the skills different fields provide, raising some and lowering others. Occupational mobility cannot undo a shock to skills: a worker whose field is devalued is worth less in every occupation she might enter. In response, 3.3% of students change field, and re-choosing nearly doubles the long-run gain, by 1.36 pp, to 2.98% of lifetime income. The two shocks also differ in who bears them. The trade war is shared roughly equally across the income distribution, while AI falls on the highest earners, who hold the degrees it devalues, and so narrows the distribution. The two shocks also adjust at different speeds. The trade war is absorbed almost at once, as workers switch occupations. The AI shock builds up only as new cohorts enter, and its full effect takes a generation to realize.

Field choice can help only when students are free to choose, and that freedom is set by education policy. We study a reform removing Denmark's two admission restrictions, the capacity caps on oversubscribed programs and the high-school tracking that limits which fields a student may enter. The reform raises lifetime income by 4.46 pp in normal times. This value barely changes under the trade war but roughly halves under AI. The constraints had rationed the most selective fields, so the reform is worth most when those fields stay valuable. The trade war leaves their relative value almost unchanged. AI devalues the very fields the reform opens, and the reform's value falls with them.

Our paper is related to several strands of literature. First, we build on general-equilibrium models of how workers with heterogeneous abilities and horizontally differentiated skills make career decisions under mobility frictions. These have been used to study trade-induced reallocation (Artuç et al., 2010; Dix-Carneiro, 2014; Traiberman, 2019), labor market responses to immigration (Llull, 2018; Burstein et al., 2020; Monras, 2020; Khanna and Morales, 2017), technological advances (Dvorkin and Monge-Naranjo, 2019; Humlum, 2021; Adao et al., 2024), labor market discrimination (Hsieh et al., 2019), recessions (Grigsby, 2021), and labor market sorting and search patterns (Lise and Postel-Vinay, 2020). These models focus on how incumbents reallocate. A smaller literature studies how new cohorts, rather than incumbents, drive labor reallocation: Matsuyama (1992) shows that sectoral adjustment can run through demographic turnover; Adão et al. (2020) make this margin central to the speed of adjustment when the affected skills are specific; and Blanchard and Willmann (2011) model cohorts choosing how much education to acquire in anticipation of trade policy. In each, the cohort's margin is an abstract choice of sector or skill level. We locate it in a specific channel, the choice of college field of study, which we estimate and embed in general equilibrium. This lets us study both how incumbents reallocate and how the skill distribution shifts as new entrants choose fields, and how education policy interacts with the shock.

We also build on and contribute to the large literature on the joint determination of human capital and labor market returns. While much of the prior literature focuses on vertical education decisions (Blanchard and Olney, 2017; Heckman et al., 2018; Ferriere et al., 2021; Eckardt, 2024; Greenland and Lopresti, 2016), a growing body of work examines variation in returns across fields of study (Hastings et al., 2013; Gemici and Wiswall, 2014; Kirkeboen et al., 2016; Bleemer and Mehta, 2022; Andrews et al., 2022; Abramitzky et al., 2024) and the role of major specificity in shaping career trajectories (Kinsler and Pavan, 2015; Leighton and Speer, 2020).[3] Our paper is most closely related to studies that estimate models of human capital accumulation alongside labor market processes (Heckman et al., 1998; Arcidiacono, 2004; Beffy et al., 2012; Ransom, 2021) and builds on the seminal work by Keane and Wolpin (1997), which integrates labor demand into college decisions.[4] Aside from our focus on technological and trade shocks, our contribution lies in the development of a framework that includes large choice sets in both educational decisions as well as occupations in the labor market. We show that both incumbent worker transitions and new cohort decisions are key margins of adjustment, which interact as incumbent reallocation affects expected earnings for new entrants, while new entrants influence future labor supply.

Finally, our paper contributes to the expanding literature on the demand for fields of study (Robles and Krishna, 2012; Wiswall and Zafar, 2015; Agarwal and Somaini, 2018), by nesting field choice within a broader labor supply framework. Methodologically, we are close to the literature exploiting centralized assignment mechanisms as a source of identification for preferences over college majors (Luflade, 2019; Kapor et al., 2024; Fack et al., 2019; Larroucau and Rios, 2022). There is also a growing literature on how particular institutional arrangements affect students' decision making (Bordon and Fu, 2015; Bleemer and Mehta, 2024). Our approach brings these strands of the literature together, with a particular focus on occupational switching, and shocks that disrupt occupational structure.

The paper proceeds as follows. Section 2 describes the data and institutional background, and Section 3 presents the motivating facts. Sections 4 and 5 present the model, estimation, and results. Section 6 closes the model and runs the counterfactuals, and Section 7 concludes.

[3] Altonji et al. (2012b) and Altonji et al. (2016) provide extensive surveys, including discussions of structural modeling issues.

[4] While our paper focuses on higher education, a substantial literature highlights the importance of early human capital accumulation (e.g., Caucutt and Lochner, 2020).

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 education system and show prominent features in students' applications and field choices.

2.1 Data Sources

We employ several administrative datasets covering the Danish economy from 1995 to 2019. 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. IDA includes information on employment status, earnings, International Standard Classification of Occupation (ISCO)-based occupation, age, gender and other demographic variables. Throughout the text, "earnings" refers to the earnings concept in this register—which are total annual earnings in workers' main job. Workers are assigned an employer and an occupation based on their employment status in November of each year.

Our main source of information regarding individuals' education comes from the *Education Register* (UDDA). The data include the highest level of degree earned, the field of study if applicable, and the year of graduation. The education register also includes a mapping from all education programs to the International Standard Classification of Education (ISCED) which allows us to classify education programs into fields of study. We supplement the education register with information on student applications to college programs from the *Coordinated Admissions System* (KOT), which is administered by the Danish Ministry of Higher Education and Science. For each student, we observe the set of programs to which they applied, whether they received admission in a program, the system under which they received admission, their high-school track and GPA. We hand-link the education programs in KOT to UDDA and calculate admission cutoffs based on all applications data.[5]

In addition to these datasets, we bring in standard firm level data from the *Firm Accounting Statistics Register* (FIRE). This allows us to identify workers' industries. Finally, we use the *Danish Foreign Trade Statistics Register* (UHDI). These data are product-firm-origin/destination level customs data on all international transactions. They contain data on value of imports and exports

[5]We also need to determine the effective GPA cutoff threshold for each program with a binding testing cutoff, as this is not available (digitally at least). Supplementary Appendix SA.2.1 contains a complete description of how we concorded KOT to UDDA and calculated admission cutoffs.

at the CN8 level, and often also contain quantity and unit value information.

2.2 Defining Fields and Occupations

We define the level of education based on the ISCED system, and divide students into three levels: high school graduates, short-cycle education,[6] and college and above. The last group is the primary focus of our analysis. Since the KOT data only include information on students applying to college programs, we focus our attention on this group in the estimation, treating the decision to enter the KOT system (i.e., whether to pursue a college degree) as exogenous.[7] Our definition of fields is based on the Danish variant of the ISCED classification of fields, i.e., ISCED-F. The ISCED-F classification breaks fields into 27 broad categories. We aggregate several codes—especially in the natural sciences—due to their small sizes. We also break apart the health code into programs at the Professional Bachelor's level and the University Bachelor's level. Table A.1 contains our list of educational fields and their ISCED-F equivalent codes.

Occupations in Denmark are recorded according to the ISCO, developed by the International Labor Organization (ILO). Supplementary Appendix SA.3 contains a complete description of how we aggregate and concord occupations over time, as well as how we deal with a host of quality and measurement issues. Ultimately, we arrive at 29 time-consistent occupation codes. Table A.2 contains our list of occupations and their ISCO-08 equivalent codes. Finally, Supplementary Appendix SA.4 discusses how we use the ONET Database to construct occupational tasks.

2.3 The Danish Education System

The Higher Education System in Denmark

Our analysis focuses on college education which includes two main types of programs: Professional Bachelor's degrees and University Bachelor's degrees. Both of these programs are classified as Level 6 according to the ISCED, and are commonly pooled in studying skilled workers in Denmark (Hummels et al., 2014). Professional Bachelor's degrees, also referred to as Academy Bachelor's degrees, typically require three years of study and are designed to provide students with practical, job-oriented skills tailored to meet labor market demands. These programs cover fields such as

[6]Short-cycle programs, corresponding to ISCED Level 5, are practical programs designed to prepare students for specific occupations, similar to vocational programs. In Denmark, these include, for example, police academies, some health care programs, maritime programs, and some vocational tracks.

[7]There are a handful of programs below college administered through the KOT. However, this is a very small fraction of total programs. They tend to be highly specialized programs—such as military and maritime programs that are more outside the traditional scope of tertiary education.

business, education, engineering, and various forms of vocational training, emphasizing technical expertise and immediate employability. University Bachelor's degrees, on the other hand, follow a more academically focused path, comparable to a standard American four-year degree but completed in three years.[8] These programs are oriented toward theoretical knowledge and research, preparing students for further academic pursuits or careers requiring in-depth expertise in their chosen fields. An increasingly large share of Danish students pursue a degree from a Professional Bachelor's or University Bachelor's (henceforth "college"). Figure A.1 plots the share of 30 year olds at each level of education. At the start of our sample, in 1995, around 18% of the sample has some form of college, and this more than doubles to 41% by the end of the sample.[9]

In Denmark, as in much of Europe, the higher education system emphasizes early specialization, differing from the U.S. model of majors and minors. Instead, students apply directly to specific programs, such as engineering or business. Switching fields can be complicated, and may involve a completely new application.

High school tracks play a critical role in shaping students' options for further education. Danish upper secondary education consists of several tracks, each influencing the types of programs students can pursue beyond high school (see Supplementary Appendix SA.1 for more information on college education and high school tracks in Denmark). While the system has become more fluid over time, with options for students to take supplementary courses to broaden their eligibility, these tracks still determine much of the access to higher education. To this end, in our estimation, we use upper secondary background, including specialization when available, to construct students' choice sets for college education. We also use high school tracks as a control when estimating students' returns and preferences over different fields of study. In our counterfactual analysis, we assess the economic implications of removing high school tracking from the college admissions process.

The Application System

Entry into college programs occurs through a centralized admissions process. Students can apply to up to eight programs (each application being a combination of field and institution). Each program has its own high-school GPA threshold for entry, determined by student demand and program supply, and regulated by the Danish Ministry of Higher Education and Science. Some programs

[8] While the Bachelor's is awarded after three years, well over 90% of students pursue a Master's program. In our analysis, we treat this as one continuous 5 year program.

[9] Our use of 30 year olds may include students who dropped out of college and students who are still studying. Nevertheless, the estimates line up reasonably well with the reported national statistics from Denmark, available at https://www.statbank.dk/HFUDD11.

also have specific prerequisites, such as sufficient mathematics preparation.

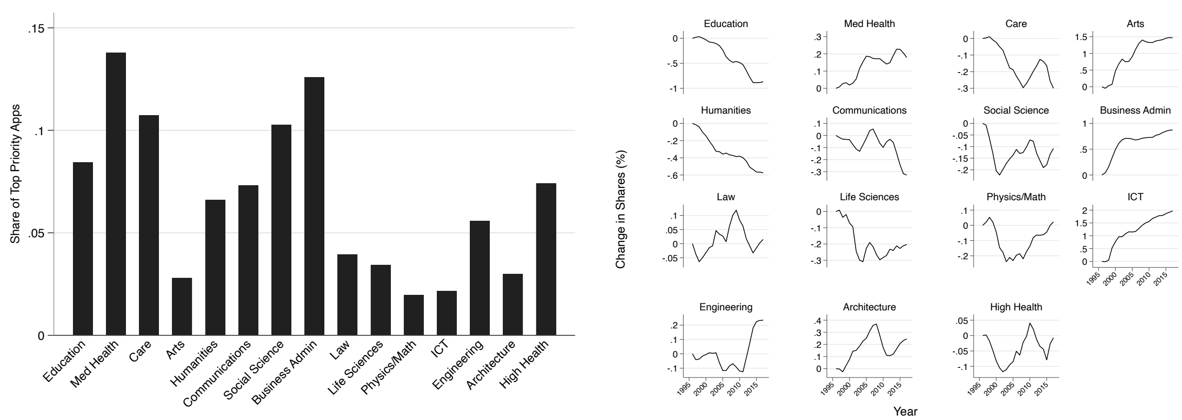
Students in Denmark are provided with significant resources to help them make informed education choices and improve their information about alternative educational paths.[10] While information gaps no doubt remain, significant efforts are made to provide guidance and resources.[11]

The Danish admissions system uses a Deferred Acceptance mechanism, ensuring that if a student is ineligible for their first choice, it does not affect their chances of being considered for subsequent choices. Thus, in principle, students ought to list their uncensored preferences for programs. In reality, both because the number of choices is small and for psychological reasons, students may not list programs they have an exceedingly low probability of accessing. We discuss how we construct choice sets and our assumptions on students' rankings more in Section 4.3.

Figure 1
Summary Statistics on First Priority

(a) Unconditional Distribution

(b) Growth Over Time



Note: Panel (a) plots the share of first-choice fields (first priority) reported by high school graduates who apply to college education through the KOT system, regardless of final admission status. The sample is pooled across years, and only fields with at least 100 applications over the course of the entire sample period are considered. Panel (b) plots the change in shares over time relative to the base year, 1995. Time series plots are plotted using a 3 year MA smoother.

Figure 1 displays the unconditional distribution of first priorities for students across college fields of study, as well as the growth in fields over time. The most popular fields are in Medium Health (e.g., nursing and health technicians), followed closely by Business Admin, Care, Social Sciences and Education. STEM fields, when including High Health (e.g., doctors and dentists), account for

[10] For example, the *UddannelsesGuiden* (Education Guide) website (https://www.ug.dk/) provides detailed information on program-specific prerequisites, historical GPA cutoffs, admissions numbers, labor market outcomes, demand trends across fields of study, and subject matter expectations. Study guidance centers (*Studievalg*) further support students through counseling services.

[11] Such information gaps have been explored in Wiswall and Zafar (2015) and Kapor et al. (2020). Lacking direct information on students' choice sets, or similar experimental variation, we must assume that students have rational expectations when we estimate our model.

roughly one quarter of students, while Arts and Humanities account for around one tenth. The most notable changes in preferences involve ICT and Education. ICT, which begins on a very small base, has seen explosive growth, as have Business Admin and Arts to a lesser extent, whereas Education has experienced significant decline.

Table 1 provides insight into students' preferences by examining the relationship between first-choice fields (columns) and second-choice fields (rows). Two patterns stand out. First, the data show strong diagonal dominance across fields, indicating that most students select a second-choice program within the same field as their first choice.[12] Second, preferences vary significantly by discipline. In STEM fields such as Life Sciences, Physics/Math, and Engineering, there is notable clustering, with a large share of students selecting second or third choices within these closely related fields. In contrast, preferences in Education and Health appear more narrowly focused, with students showing a lower likelihood of considering programs outside their initial field of interest.[13] This suggests a strong tendency toward specialization in these disciplines.

3 Stylized Facts

In this section, we present two stylized facts about the relationship between fields of study and the labor market. These facts highlight how fields shape individuals' occupational choices and how students adjust their field choices in response to labor market returns—key relationships we seek to capture in the model.

Fact 1: Fields of study lead to different occupational choices

We first use quasi-experimental variation to show how field of study influences occupational choices at labor market entry. Specifically, we implement a fuzzy regression discontinuity (RD) design exploiting admission cutoffs into different fields.[14] Under Denmark's centralized admission system,

[12] In our subsequent analysis, we do not distinguish between students choosing a similar program within the same institution (e.g., Physics vs. Mathematics at Aarhus) or applying to the same program at different institutions (e.g., Economics at Copenhagen vs. Aarhus). We do not find a clear pattern that suggests students engage in one kind of behavior or another: 37% of the time students apply across programs within the same institution, and the other 63% of the time they apply for similar programs across institutions.

[13] Some qualification is warranted here, as students are partially constrained in their choice sets by their high school background. However, in our structural estimation, we will show that controlling for high school tracking alone, without considering latent preference heterogeneity, is insufficient to generate patterns in line with the ranked choice data.

[14] Since the majority of admissions takes place through the score-based assignment mechanism, but some students who are not admitted may still enroll through an alternative admissions route based on supplementary exams or other criteria, admission does not deterministically dictate enrollment. Our design is therefore properly characterized as a

Table 1

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

Second Priority	First Priority: Education	Care	Med Health	Arts	Humanities	Social Sciences	Communication	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 Sciences				9.40	15.40	49.09	11.60	9.46	16.07	4.10		7.62	4.69		
Communication				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

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.

which uses the deferred acceptance algorithm, students who narrowly miss or exceed field-specific GPA thresholds provide a natural source of variation. Following Kirkeboen et al. (2016), we compare individuals just above and below these cutoffs to estimate the impact of being admitted to a field m versus not being admitted.

Relying on the fuzzy RD implementation of Calonico et al. (2017), we use students' GPA at the time of application as the running variable. To quantify field-specific occupational sorting, we define each field's top occupation as the one where graduates of that field sort most heavily relative to the population:

$$occ_m^{Top} = \arg\ max_o(s_{o|m} - s_o), \tag{1}$$

where $s_{o|m}$ is the share of graduates from field m in occupation o, and s_o is the overall population share in occupation o. For example, the top occupation for those studying Life Sciences is "Work in the Natural Sciences" (ISCO code 21). We estimate the increase in the probability of entering a field's top occupation. We calculate this by pooling across all years, but we run our regressions at the program-year level.

Table 2
Effect of Field of Study on Occupational Entry

	Dep Var: Prob. of Entering Top Occ.		
	(1)	(2)	(3)
		Second Stage	
β	0.313**	0.297**	0.264*
	(0.129)	(0.144)	(0.157)
		First Stage	
β	0.160***	0.193***	0.214***
	(0.026)	(0.032)	(0.037)
N	33920	22160	17776

Notes: Regression discontinuity estimation of the effect of entry into field of study on entry into top occupation as defined in equation (1). First column pools all individuals with at least two field priorities, second conditions on entry to first or second priority, third removes education, law, and medicine. Running variable is GPA at end of high school, covariates include year FE, high school track FE, and gender. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. Robust statistics calculated according to `rdrobust` implementation of Calonico et al. (2017).

Table 2 reports the estimated effects. Column (1) includes all individuals with at least two field priorities. Column (2) restricts to those who were accepted into either their first or second choice, excluding those assigned further down the list. Column (3) excludes highly specialized fields, i.e., education, law, and medicine, to demonstrate that our results are not driven by these fields. Across specifications, we control for year fixed effects, gender, and high school track. The first row

fuzzy regression discontinuity design (Imbens and Lemieux, 2008).

presents the coefficient of interest, while the second row is the first stage coefficient of the GPA on the probability of entry to a field. Admission to a field raises the likelihood of entering one's top occupation by 26–31 pp. The estimates remain significant at the 10% level even when we exclude the most specialized fields.

Fact 2: Field of study decisions are correlated with future earnings growth

We now show that student demand for a field rises with expected future earnings. Establishing this relationship is empirically challenging. First, many degree programs are capacity constrained, so changes in enrollment composition can reflect either demand responses or supply constraints. Second, graduating cohorts influence equilibrium wages on entry.

To address these issues, we leverage a key advantage of our data: we observe students' first-choice program preferences, not just their final enrollment. Specifically, we measure *excess demand*:

$$\text{ExcessDemand}_{mt} \equiv \log(n_{mt}) - \log(a_{mt}),$$

where n_{mt} is the number of students listing program m as their top priority and a_{mt} is the number of students admitted to program m in year t. This metric is zero when supply is unconstrained and positive when latent demand exceeds capacity.

Our regressions below focus on how this variable responds to income. Specifically, we regress a program's excess demand in year t on the mean labor income of its former students—contemporaneously and at five- and ten-year horizons—with program and year fixed effects and a control for the program's admission cutoff. Suppose incomes in a program rise. If the increase reflects declining student interest (e.g., due to lower non-pecuniary amenities), we would expect excess demand to remain flat or fall. Conversely, if income growth stems from labor demand outpacing supply capacity, we should observe rising excess demand. While this reduced-form test focuses on wage levels, students presumably care about the full earnings path, occupational mobility, and amenities, which are features we model structurally later.

Table 3 presents our results. Across all specifications, and controlling for program-specific admission thresholds, excess demand is positively associated with both current and future earnings.[15] These findings suggest that students are responsive to income expectations when making educational

[15] In Tables A.3 and A.4 we provide several robustness checks: we use the log of income instead of levels (albeit our model is in levels). We also use shares of applicants, controlling for capacity, instead of excess demand. Finally, we also include future shares as a control—this is in line with our model, where future choices control for future expectations of returns.

Table 3
Excess Demand for Field of Study and Income

	Dep. Var: Excess Demand			
	(1)	(2)	(3)	(4)
Income_t	0.220*			0.259*
	(0.126)			(0.137)
Income_{t+5}		0.321**		0.289**
		(0.133)		(0.130)
Income_{t+10}			0.614***	
			(0.177)	
N	330	270	195	270
Within R^2	0.060	0.106	0.165	0.119

Note: An observation is a program in a given year. The dependent variable is the program's *excess demand*, $\log n_{mt} - \log a_{mt}$: the log number of students listing a program m as their first priority (n_{mt}) less the log number admitted to the program (a_{mt}) in year t. The regressors are the mean labor income earned by the program's former students, measured contemporaneously (Income_t) and five and ten years later. All specifications include program and year fixed effects and control for the program's mean GPA admission cutoff. Robust standard errors in parentheses: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.1$.

choices. We next develop a structural model that maps these relationships to primitives to study how educational choices mediate the economy's adjustment to shocks.

4 A Model of Labor Supply and Field Choice

In this section we describe students' education and labor supply decisions, with labor demand left to Section 6. Our labor supply model combines three main ingredients: first, students make forward-looking choices over which field to pursue; second, workers with different educational backgrounds make forward-looking occupational choices; third, we incorporate latent heterogeneities in preferences over fields and in comparative advantages across occupations.

4.1 Environment

We consider a dynamic, small, and open economy operating in discrete time, indexed by t. The total measure of individuals in the labor market is constant at $\bar{L}$. Each period, a fraction $\delta\bar{L}$ of incumbent workers exits the labor force due to retirement or other exogenous reasons and is replaced by an equally sized new cohort of graduates.

Individuals enter the model upon completing high school and prior to entry into college. At this point, they are characterized by observable characteristics, including test scores, high school track (e.g., Mathematics), gender, and their latent preference type $l \in 1, \ldots, L$. The preference type l captures heterogeneity in individuals' unobserved preferences over fields of study, as detailed later. We collect students' observable characteristics into the vector $\mathbf{X}$.

In this *field choice block*, individuals select a field of study $m \in \mathcal{M}$ based on anticipated returns and preferences, both of which may depend on their initial characteristics. Upon graduation, individuals possess a qualification in field m and an unobserved labor market type $k \in 1, \ldots, K$. Together, the field of study m and labor market type k determine each worker's absolute and comparative advantages across occupations.

Each new cohort of graduates enters the *labor supply block*, supplying one unit of labor inelastically in each period. Graduates make a forward-looking decision on their initial occupation. In subsequent periods, incumbent workers may switch occupations, incurring associated switching costs.

Since the model is solved backwards, we describe the model accordingly: beginning with the occupational decisions of incumbent workers, then the occupational entry decisions of new entrants, and finally the field of study choices by incoming college students.

4.2 Labor Markets

4.2.1 Incumbent Workers

At the *beginning* of period t, incumbent workers who spent the previous period in occupation o choose an occupation o'. Workers are characterized by (k, m) and supply human capital $H_o(k, m)$ to occupation o. Switching occupations entails a moving cost $C_{oo'km}$, which depends on the worker's type and field, and the origin-destination pair. Workers also receive a stochastic occupation preference shock ε_{ot} in each period.

To parameterize human capital, we project each occupation onto Z tasks. Each occupation is represented by a vector $\zeta_o = (\zeta_{1,o}, \cdots, \zeta_{Z,o}) \in \mathbb{R}^Z_{++}$, the positive orthant of $\mathbb{R}^Z$, where $\zeta_{z,o}$ denotes the loading on task z in occupation o. A higher value of $\zeta_{z,o}$ indicates that task z is more important for output in that occupation. For example, $\zeta_{z,o}$ can be mathematical tasks, which will be more important in professions such as Engineering or Accounting, and less important in Law.

Each worker type has task-specific abilities $\alpha_{k,z}$, and their log output in an occupation is the dot product of α and ζ. We additionally include a field effect, α_{om}, which scales the human capital supply of workers in each occupation based on their field of study, and a type effect α_k, which captures the absolute advantage of type-k workers.[16] Combining, we have:

$$\log H_o(k,m) = \alpha_{om} + \alpha_k + \underbrace{\sum_{z=1}^{Z} \alpha_{k,z}\zeta_{z,o}}_{\equiv \alpha_{ok}}, \tag{2}$$

where α_{ok} reflects a worker's task-based comparative advantage.

A few comments are in order. First, the α_{om} terms allow workers in different fields of study to have distinct comparative advantages. For example, workers with an education background in STEM may have large α_{om} values in technical occupations, such as engineering. Second, even within a field, workers can have distinct comparative advantages across occupations due to differences in their labor market type k.[17] Third, the conditional distribution of k depends on the chosen field of study m, a relationship we discuss in detail below.

As a final note, our model abstracts from experience-based or occupation-specific human capital accumulation in order to focus on how educational fields shape skill development and initial labor market sorting. While this simplifies the dynamics of incumbent behavior, it leaves out an important channel through which workers adapt and evolve over time. Modeling these tenure effects could allow richer insights into mobility and occupational transitions, as emphasized in Adao et al. (2024) and Humlum et al. (n.d.). Our framework is flexible enough to accommodate such extensions (e.g., Traiberman, 2019). However, while we view this as a promising direction for future work, we abstract from this dimension in order to focus on field choice while maintaining a relatively parsimonious model.

Returns to an occupation are determined by both workers' human capital and competitively determined skill prices w_{ot}. The worker's Bellman equation in period t is given by,

$$v_t(k,m,o,\varepsilon) = \max_{o'} w_{o't} H_{o'}(k,m) - C_{oo'km} + \varepsilon_{o't} + \beta E_t V_{t+1}(k,m,o'), \tag{3}$$

where $V \equiv \int v(\cdot,\varepsilon)dG(\varepsilon)$ is the value function v integrated over shocks, and β is the effective discount factor, which is the product of the time discount rate, $\tilde{\beta}$, and the probability of survival, $1-\delta$.

[16] Technically we use one index for all productivity differences across workers—absolute and comparative advantages. This is without loss of generality. For example, one could have workers with the same comparative advantage but different levels at each task simply by having two types with parameters shifted up by a constant for one type. Moreover, since α_{om} is flexible, no additional α_{mz} term could be identified. The purpose of the tasks is to lower the dimensionality of the latent variables in estimation.

[17] As an example, consider a simple economy with two occupations, two tasks, and two types. Suppose the task vector for the first occupation is $\zeta_1 = (1,0)$ and for the second occupation $\zeta_2 = (0,1)$. Similarly, let worker abilities be $\alpha_1 = (1,0)$ and $\alpha_2 = (0,1)$. In this case, type 1 workers would produce output only in occupation 1, while type 2 workers would only produce output in occupation 2. Despite this horizontal differentiation, neither type has an absolute advantage.

We assume preference shocks, ε_o, are conditionally independent of other variables and follow a Gumbel distribution with mean zero and scale parameter ν. Under these assumptions, the value function V simplifies to:

$$V_t(k,m,o) = \nu \log\left(\sum_{o'} \exp\left(\frac{-C_{oo'km} + w_{o't}H_{o'}(k,m) + \beta E_t V_{t+1}(k,m,o')}{\nu}\right)\right). \tag{4}$$

Moreover, the corresponding policy function, which determines the transition probabilities across occupations, is:

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

Equation (5) can be used to interpret ν as governing the elasticity of occupational switching with respect to wages, as opposed to idiosyncratic shocks. At the extremes, as $\nu \to 0$, the transition probabilities collapse to a deterministic outcome where all workers choose the occupation with the highest expected payoff. Conversely, as $\nu \to \infty$, workers allocate themselves uniformly across occupations, becoming indifferent to differences in deterministic payoffs.

4.2.2 New Entrants

Like incumbents, new entrants are characterized by a (k,m) pair. Upon entering the labor market, they choose an initial occupation, according to the utility function,

$$v_t^e(k,m) = \max_o -C^e_{o,km} + \varepsilon_o + w_{ot}H_o(k,m) + \beta E_t V_{t+1}(k,m,o), \tag{6}$$

where $C^e_{o,km}$ denotes entry costs specific to occupation o for type (k,m) and ε_o is a stochastic preference shock at the initial entry, assumed to follow the same Gumbel distribution as for incumbents. The resulting choice probabilities for new entrants are given by:

$$\lambda_t^e(o|k,m) = \frac{\exp\left(\frac{-C^e_{o,km} + w_{ot}H_o(k,m) + \beta E_t V_{t+1}(k,m,o)}{\nu}\right)}{\sum_{o'} \exp\left(\frac{-C^e_{o',km} + w_{o't}H_{o'}(k,m) + \beta E_t V_{t+1}(k,m,o')}{\nu}\right)}. \tag{7}$$

The expected value of entering the labor market with type (k,m) is given by:

$$V_t^e(k,m) = \nu \log \sum_o \exp\left(\frac{-C^e_{o,km} + w_{ot}H_o(k,m) + \beta E_t V_{t+1}(k,m,o)}{\nu}\right). \tag{8}$$

4.3 Field Choice

We begin by describing the returns to a field before outlining how students construct their ranked choice lists. Students enter the field choice block with a latent preference type $l \in 1, \ldots, L$ and a vector of observable characteristics $\mathbf{X}$, which may include test scores, high school background, and demographic factors. We mention here that throughout the paper, we treat the decision to enter college as exogenous. That is, we model field choices only for students who enter the college system, and take the distribution of fields among non-college students as given.

Returns to Field

In deciding on a field of study, students consider the expected labor market returns associated with each option, denoted $V_t(m|\mathbf{X})$. These returns depend on two key factors. First, there is the common productivity effect, given by α_{om} in (2). This term reflects the average productivity associated with each field in different occupations. For example, degrees in Physics/Math may offer relatively higher returns in technical occupations than the return in those occupations to Communication degree holders. Second, each field of study induces a conditional distribution over labor market types $k \in K$, denoted $\pi(k|m, \mathbf{X})$.[18] This stochastic term allows for heterogeneity in outcomes among students within the same field, as students may possess varying talents that influence their career trajectories. Given this conditional distribution, the expected labor market return for a student with observables $\mathbf{X}$ choosing field m is:

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

where $E_t V^e_{t+5}(k, m)$ is the expected lifetime payoff to the student upon graduation with a labor market type k and a field m, given by (8).

Consistent with workers, we assume that students form rational expectations. This implies that students (i) observe the current aggregate state, and therefore know current returns across fields, and (ii) have knowledge of the stochastic process governing the economy—both aggregate and idiosyncratic states—allowing them to optimally forecast future returns. While this may not fully reflect how students behave in reality (Hastings et al., 2016; Conlon and Patel, 2022), rational expectations remain a useful benchmark for understanding how the education choices interact

[18] An alternative interpretation is that the college education production function is given by $\pi(k|m, \mathbf{X})$, where field choice m and observables $\mathbf{X}$ serve as inputs. The output is a distribution of labor market types k, which, as described in Section 4.2, is linked to worker productivities and occupational switching costs.

with labor market shocks when expectations are fully aligned with economic fundamentals. Our framework is actually flexible enough to handle alternative belief formations, such as myopic or biased expectations. However, our empirical approach is also practically motivated: we lack data on students' information sets or subjective forecasting rules. Given this limitation, the rational expectations assumption provides a coherent and tractable way to model forward-looking behavior.

In addition to the expected labor market returns, students' field choices reflect non-pecuniary factors that vary by latent preference type l. Each type-l student has a time-invariant utility for a field denoted by ϑ_{lm}, which captures intrinsic preferences unrelated to earnings. The relative importance that students assign to future earnings varies across types too and is governed by the parameter θ_l. Students may also prefer fields that offer a larger number of programs. This is captured by a love-of-variety term $\upsilon \log(N_m)$, where N_m is the number of programs available in field m. For example, the "Physics/Math" field may encompass multiple mathematics and physics programs across institutions, which students may find attractive.[19] Finally, students receive a conditionally independent Gumbel(0,1) preference shock $\varepsilon_m^{\mathcal{M}}$ over fields, introducing idiosyncratic variation in their choices.

Combining these components, the utility that students of preference type l with observables $\mathbf{X}$ assign to field m is:

$$\tilde{U}_{lmt}(\mathbf{X}) = \vartheta_{lm} + \theta_l V_t(m|\mathbf{X}) + \upsilon \log(N_m) + \varepsilon_m^{\mathcal{M}}. \tag{9}$$

Henceforth, we use $\tilde{U}$ to denote the utility inclusive of the shock, $\varepsilon^{\mathcal{M}}$, and U to refer to the deterministic component of utility excluding the shock, i.e., $U_{lmt}(\mathbf{X}) = \vartheta_{lm} + \theta_l V_t(m|\mathbf{X}) + \upsilon \log(N_m)$.

Ranked List Determination

Upon receiving their preference shocks $\varepsilon_m^{\mathcal{M}}$, students construct a ranked list of field choices of length R. Field assignments are determined using the deferred acceptance algorithm, and we assume that students truthfully reveal their preferences over their choice sets. Students' choice sets, $\mathcal{C} \subset \mathcal{M}$, may depend on their observable characteristics. Some restrictions on choice sets are institutional: students without the proper prerequisites may not apply to certain programs. However, students may also avoid listing programs that are unobtainable. We follow Fack et al. (2019) and Kapor et al. (2024) in bounding choice sets using test scores. We assume that students have sharp expectations

[19] As the focus of our paper is not on institutional quality or on within-field variation, we treat programs within a field of study as symmetric. Field-level differences in quality or popularity are partially reflected in the admission cut-offs described below.

about what cutoffs will be, and only list programs for which they know they will be below the admission cutoff. However, these expectations may still be noisy, and so we assume that students apply within a "realistic" set of alternatives determined by "score bounds". In particular, students only apply to programs for which their score is below the equilibrium cutoff plus some buffer points.

For ease of notation, let m_r denote the field ranked in position r, with m_1 being the top-ranked field, and so on. Let $\mathbf{m}_{-r} = \{m_1, ..., m_{r-1}\}$ be the set of programs that have already been ranked before the r^{th} choice (with the understanding that $\mathbf{m}_{-1} = \varnothing$). Under the Gumbel assumption, the probability of observing a ranked list $(m_1, m_2, ..., m_R)$ is given by,

$$P(m_1, m_2, ..., m_R; \mathbf{X}, \mathcal{C}) = \prod_{r=1}^{R} \frac{\exp(U_{lm_rt}(\mathbf{X}))}{\sum_{m \in \mathcal{C} \setminus \mathbf{m}_{-r}} \exp(U_{lmt}(\mathbf{X}))}.$$

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

Our demand system is tractable and encompasses many models as special cases. First, in the absence of unobservable heterogeneity or occupational choice component in the labor market, the model reduces to one where workers assess the net present value of earnings based on cross-sectional average earnings in each field. Second, when $\beta = 0$, the model simplifies to one where workers focus solely on their initial earnings in the job market.

5 Estimation

In this section, we first describe our estimation strategy for the labor supply and field choice blocks, and then present the corresponding estimation results.

5.1 Estimation Strategies

5.1.1 Labor Market Parameters

Incumbent Workers Our estimation approach builds on Traiberman (2019), extending it to allow for heterogeneity across college fields of study. The estimation proceeds in two stages. In the first stage, we estimate wage parameters, occupation transition probabilities, and the distribution of worker types using an expectation-maximization (EM) algorithm. In the second stage, we recover the switching costs and scale parameter of the Gumbel distribution governing occupation choice behavior.

We assume that observed log wages are a noisy signal of model-implied earnings. Let $f(w_{it}|k, o, m, t; \alpha)$ denote the likelihood of observing wage w_{it} for an individual of type k, in occupation o, field m, and year t, with human capital parameters α. The probability that an individual is type k is q_k. The likelihood for individual i is:

$$\tilde{\mathcal{L}}_i = \sum_{k=1}^{K} q_k \cdot \lambda^e(o_1, m|k) \cdot f(w_{i1}|\cdot) \cdot \prod_{t=2}^{T} f(w_{it}|\cdot) \cdot \lambda_t(o_t|o_{t-1}, m, k), \tag{10}$$

where $\lambda^e(o_1, m|k)$ is the initial occupation distribution for a type-k worker in field m, and $\lambda_t(o_t|o_{t-1}, m, k)$ is the transition probability to occupation o_t from o_{t-1}. These transition probabilities reflect both worker preferences and the model's dynamic incentives.

Maximizing this likelihood directly is challenging because (i) the worker types k are unobserved, and (ii) the transition probabilities λ depend on dynamic value functions. We address the first issue using the EM algorithm, which iteratively updates the posterior distribution over types, i.e., q_k, and the associated parameters. To handle the second issue, we follow Arcidiacono and Miller (2011) by treating the transition probabilities λ as auxiliary parameters to be estimated directly in this stage. This simplifies computation without compromising consistency. Further implementation details, including the full log-likelihood expression optimized in the EM algorithm, treatment of initial conditions, and bootstrap procedures for inference, are provided in Appendix B.1.

In the second stage, we estimate occupation switching costs using a Hotz-Miller inversion (Hotz and Miller, 1993), similar to Artuç et al. (2010). After the first stage, we observe estimated wages and switching probabilities for each type-field pair (k, m). These sufficient statistics allow us to estimate the cost of moving from one occupation to another:

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

where the left-hand side reflects the relative choice probability of switching versus staying, and the right-hand side captures net utility gains, adjusted for switching costs and value continuation. The error term ξ represents expectation shocks and is orthogonal to time-t variables under rational expectations, allowing estimation by OLS.

New Cohorts For new entrants, the only additional parameters to estimate are entry costs C^e. We apply a similar inversion strategy, normalizing the cost of entering occupation 1 ("Management") to zero. As shown in Appendix B.1.2, this leads to the following estimating equation:

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

where $C_{1,o,km}$ is the switching cost from Management to o, estimated in the previous step. We estimate C^e by taking the time average of this expression.

5.1.2 Worker-Type Production Function

In order to recover the worker-type production function, $\pi(k|m, \mathbf{X})$, we project the previously recovered q_{ik}, the probability that individual i is type k, onto student observables. Precisely, we implement a nonparametric binning approach, discretizing continuous test scores into quintiles and forming bins based on the combination of score quintile (s_q), high school track (h), and gender (g). The production function is then estimated as the empirical frequency of type k within each bin:

$$\pi(k|m, \mathbf{X}) = \pi(k|m, s_q, h, g) = \frac{\sum_{i \in (m,s_q,h,g)} q_{ik}}{n_{m,s_q,h,g}},$$

where $n_{m,s_q,h,g}$ is the number of individuals in bin (m, s_q, h, g). Estimating $\pi(\cdot)$ cell by cell rather than pooling across the score distribution means that a given field maps into a different distribution over types k at each score quintile. The expected labor-market return therefore admits interactions between observables and field of study.

5.1.3 Field Choice Parameters

We now turn to the estimation of field-choice parameters. Similar to the previous section, we provide intuition while deferring technical details to Appendix B.2. The estimation proceeds by leveraging variation in observed field rankings to recover utility parameters and the distribution of latent preference types.

We begin by describing how to estimate utility parameters assuming each individual's latent preference type ℓ is known. Recall the payoff to a field m at time t for type l individuals is,

$$U_{tlm}(\mathbf{X}) = \vartheta_{lm} + \theta_l \left[\sum_k \pi(k|m, \mathbf{X}) E_t V^e_{t+5}(k, m) \right] + \upsilon \log(N_{mt}), \tag{12}$$

It is useful to define:

$$\bar{U}_{tlm}(\mathbf{X}) \equiv U_{tlm}(\mathbf{X}) - \upsilon \log(N_{mt}) = \vartheta_{lm} + \theta_l \left[\sum_k \pi(k|m, \mathbf{X}) E_t V^e_{t+5}(k, m) \right]. \tag{13}$$

The term involving V^e—the expected value of entering the labor market with type k and field m—is not directly observed, because it is differenced out in the labor market estimation.[20] To resolve this

[20]The key challenge of estimating V^e directly is that we do not observe the final continuation value of V. Even differences across fields, $V^e(k, m) - V^e(k, m')$ cannot be directly observed either, as there is no simple representation

issue, we treat the utility in the last period, i.e., $\bar{U}_{Tlm}(\mathbf{X})$ as a parameter to be estimated, where T refers to the last period for which we observe field choices. In Appendix B.2.1, we show that for any $t = T - \tau$, (13) can be iterated forward and rewritten with the following representation:

$$\bar{U}_{T-\tau,lm}(\mathbf{X}) = \vartheta_{lm}(1-\beta^{T-\tau}) + \theta_l \underbrace{\left[\sum_{\tau'=\tau}^{T-1} \beta^{\tau'-\tau} V^F_{T-\tau'+5}(m,\mathbf{X})\right]}_{XV^F_{\tau,T}(m,\mathbf{X})} + \beta^{T-\tau}\bar{U}_{Tlm}(\mathbf{X}), \tag{14}$$

where V^F is an entry value term that *only* depends on observable data.[21] For ease of notation, we collect these observable terms into a single term and denote it as $XV^F_{\tau,T}(m,\mathbf{X}) \equiv \sum_{\tau'=\tau}^{T-1} \beta^{\tau'-\tau} V^F_{T-\tau'+5}(m,\mathbf{X})$. After plugging in outcomes for expectations this approach allows us to express $\bar{U}_{t\ell m}(s)$ entirely in terms of observable data, terminal utility parameters, and expectation error. By iterating backward from period T, we recover all necessary values of $\bar{U}_{t\ell m}(\mathbf{X})$. Notice that if $\tau = T$, this collapses to the final period utility. Otherwise, this is a discounted sum of observed payoffs, using the discounted final period utility, i.e., $\beta^{T-\tau}\bar{U}_{Tlm}(\mathbf{X})$, as a control for the continuation value.

Next, we turn to the estimation of the type distribution. Students are observed to submit ranked lists of field preferences. Conditional on being of type ℓ, the likelihood of a given ranked list follows a nested logit form. To define the full likelihood, we simplify notation by letting $N_{i,t,r,m}$ denote the number of programs available to individual i at time t, in position r, within field m, given prior choices. As students move down the ranked list, $N_{i,t,r,m}$ will also be updated accordingly. The full likelihood of observing the ranked list $(m_1, m_2, \ldots, m_R)$ for student i is then:

$$L_{i,t}(m_1,\ldots,m_R,\mathbf{X}_i) = \prod_{r=1}^{R} \frac{N^{\upsilon}_{i,t,r,m_r}\exp\big(\bar{U}_{t\ell m_r}(\mathbf{X}_i)\big)}{\sum_{m'\in\mathcal{C}_i} N^{\upsilon}_{i,t,r,m'}\exp\big(\bar{U}_{t\ell m'}(\mathbf{X}_i)\big)}.$$

We construct the full log-likelihood by aggregating over all ranks, periods, and individuals. To estimate the model, we implement the EM algorithm.[22]

of the difference that does not require knowledge of some terminal value.

[21] Specifically,

$$V^F_{t+5}(m,\mathbf{X}) = \sum_k \pi(k|m,\mathbf{X})\left\{-(1-\beta)C^e_{okm} + E_t\left(w_{o,t+5}H(o,k,m) - \nu\log\lambda^e_{t+5}(o|k,m) - \beta\nu\log\left(\frac{\lambda_{t+6}(o|o,k,m)}{\lambda^e_{t+6}(o|k,m)}\right)\right)\right\}.$$

[22] Applying the EM algorithm is completely analogous to its application in the estimation of labor market types: one iterates on optimizing the likelihood conditional on the distribution of types, q_{il}, and an update of the distribution using Bayes' Rule, $q^{(g+1)}_{i\ell} = \frac{\exp(LL^{(g)}_{i\ell})q^{(g)}_{\ell}}{\sum_{\ell'}\exp(LL^{(g)}_{i\ell'})q^{(g)}_{\ell'}}$.

5.2 Identification

While all parameters are ultimately estimated jointly through the likelihood function, we provide here an intuitive overview of how key parameters are identified.

First, wage parameters are identified using the panel structure of wage data, analogous to a fixed-effects strategy. If occupations were ignored, individuals could be clustered into labor market types based on long-run average wage quantiles.[23] With occupations, identification exploits the wage patterns of occupational switchers. If workers switch occupations yet remain at stable wage ranks, they possess absolute advantages in both occupations. If their rank declines post-switch, this suggests a comparative advantage in their initial occupation. These cases are illustrated in Figure A.2.[24]

Second, switching cost parameters are identified from the regression specification linking occupational switching to changes in wage differentials. In particular, the response of *net* flows to changes in wage *differentials* identifies ν. For example, if flows between two occupations remained unchanged after the wage gap grew, this would imply a small $1/\nu$. The level of *gross* flows pins down the average level of switching costs, $C_{oo'km}$. Finally, differences in C across occupation pairs are identified by variation in switching patterns relative to wage differentials. For example, if most switching occurs between o and o' rather than o and o'', despite similar wage gaps, then $C_{oo'}$ is estimated to be smaller than $C_{oo''}$. See Artuç et al. (2010) for more discussion of these points.

Intuition for the identification of the field choice parameters can be drawn by breaking apart the steps of the EM algorithm. Given a known distribution of types, θ_l is identified from changes in field choices following shifts in the observed returns to that field, XV^F. Similarly, the fixed effects, ϑ_{lm}, reflect long-term sorting patterns under type-specific choice sets.

The type distribution itself is identified using information from students' ranked choice lists. The key insight is that, without latent types, the independence of idiosyncratic shocks implies that second choices are statistically independent of first choices: $P(m_{r_2} \mid m_{r_1}) = P(m_{r_2})$. Latent types restore this dependence structure by explaining systematic choice patterns across ranks.[25] Thus, observing ranked lists is *crucial*: without them, single-choice observations could not separately

[23]Specifically, we would fit a Gaussian mixture model, assuming each worker's draw was fixed at career start and shifted earnings in all occupations by the same amount. This approach resembles Bonhomme et al. (2019), and indeed the EM algorithm underlies practical implementations of k-means clustering.

[24]This simple intuition abstracts from the role of career profiles. Workers with similar careers are also grouped together. Seo and Oh (2024) discusses how observing switching at different horizons for the same cohort identifies persistent latent heterogeneity.

[25]See Greene and Hensher (2010) or Conlon et al. (2024) for more on the identification of unobservable preference parameters from ordered choice data.

identify heterogeneity in preferences from idiosyncratic choice shocks.

5.3 Estimation Results

We report the field-choice parameters first, then the labor market parameters. Two parameters are set outside the estimation. The exit rate is $\delta = 0.03$, implying an expected working life of $1/\delta \approx 33$ years after labor market entry, so a graduate entering in her late twenties retires in her early sixties.[26] The time discount factor is $\tilde{\beta} = 0.95$, so the effective discount factor is $\beta = \tilde{\beta}(1 - \delta) = 0.92$.

5.3.1 Field Choice

In this subsection, we emphasize two findings: (i) students' responsiveness to income shocks, and (ii) the importance of latent heterogeneity for generating realistic substitution patterns. The complete set of parameter estimates, including how different student types sort across fields is provided in Appendix B.7.

For our estimation, we consider four latent preference types, i.e., $L = 4$. We briefly summarize all four types, which have a natural interpretation, while Table A.6 contains the complete set of estimated preferences. Type 1 students predominantly prefer education, care, and medium health; Type 2 students focus almost exclusively on high-demand health fields; Type 3 students are drawn primarily to STEM and business fields; and Type 4 students are oriented towards humanities and social sciences.

Table 4 presents our estimates of θ for all types, along with their distribution. Additionally, we compute the mean semi-elasticity of students' first-choice field with respect to a shock for each student i:

$$\mathcal{E}^f_{il} \equiv \frac{d\pi_{im(i)l}}{d\log V_t(m|\mathbf{X})} = \pi_{im(i)l}(1 - \pi_{im(i)l}), \tag{15}$$

where $m(i)$ is the students' field choice. This value is the percentage point decrease in the share of students choosing a particular field in response to a 1% decline in that field's NPV. It varies across individuals due to differences in observable characteristics, so our final measure is the weighted average across students, where the weights reflect each student's probability of belonging to a given type, q_{il}.[27]

[26] This is within the range implied by Danish effective retirement ages over our sample period. Source: OECD, average effective age of labor market exit, 2000–2020.

[27] Calculating this elasticity requires knowing V^e in levels. Since thus far we have only measured V^e in differences,

Table 4
Field Choice Parameters

Parameter	Type 1	Type 2	Type 3	Type 4
θ	0.030	0.144	0.052	0.031
	(0.001)	(0.022)	(0.004)	(0.001)
Q_l	0.183	0.197	0.301	0.320
$\mathcal{E}^f$	0.071	0.352	0.115	0.074

Notes: The table reports the results of field choice estimation by latent preference type l. The first row reports the estimates of θ_l in (12), with bootstrapped standard errors in parentheses below. The second row reports the share of each type in the population. The third row reports the mean semi-elasticity of students' first choice field with respect to a shock, which is the mean of (15) across individuals.

We draw two lessons from our estimation. First, students respond to income differences. The type-weighted average of the final row—approximately 0.14—implies that a 10% decrease in the net present value of a field leads to a 1.4 pp decline in the probability of choosing that field. While modest, our effect is in line with other estimates from structural models, as well as estimates using experimental variation.[28] For example, Wiswall and Zafar (2015) estimate the elasticity of the log odds ratio of switching fields with respect to an income shock to be 1.6, while the equivalent number in our estimation would be 2.5. As another comparison, Beffy et al. (2012) estimates that a 10% increase in returns to STEM fields only increases entry by a quarter of a percentage point. This is smaller than our estimate, but likely reflects our choice of aggregation coupled with the fact that students' preferences tend to be clustered. For example, nearly all STEM students are of only one type. For those students, the probability of choosing STEM following a 10% shock for this group only declines by 0.9 pp. The love-of-variety parameter is estimated at $\upsilon = 1.363$ (Table B.4), indicating that students value the breadth of programs within a field.

We also examine aggregate substitution patterns across fields of study. For each field m, we calculate how students' choices shift following a shock to that field's NPV. For example, if the returns to Education increase by 1%, we calculate the fraction of students switching to Medium Health, Care, and other fields. In a standard logit model without heterogeneity, substitution patterns are proportional to initial conditional probabilities:

$$\frac{d\pi_{m'}/d\log V_m}{d(1-\pi_m)/d\log V_m} = -\pi_{m'}/(1-\pi_m).$$

we assume the entry cost into Retail equals one-quarter of a year's earnings, reflecting the average time required to find a retail job. Our results are not particularly sensitive to this normalization. Appendix B.4 provides details on our approximation, including how we take averages over time.

[28]Some care is required as we are measuring the responsiveness to NPV, which includes earnings and the expectation of shocks and costs along the career path.

Table 5

Substitution Matrix $\left(\frac{d\pi_{m'}/d\log V_m}{d(1-\pi_m)/d\log V_m}\right)$

	Initial Field														
Substitute Field	Education	Medium Health	Care	Arts	Humanities	Communication	Social Sciences	Business Admin	Law	Life Sciences	Physics/Math	ICT	Engineering	Architecture	High Health
Education	.	7.53	60.13	2.63	3.58	2.87	3.04	1.61	1.76	0.78	0.72	0.79	0.86	1.11	1.08
Medium Health	12.11	.	16.30	1.40	0.92	1.32	1.22	1.11	1.39	19.51	1.23	0.47	0.53	0.65	47.82
Care	63.93	17.40	.	2.04	1.51	1.55	1.33	0.58	0.71	0.43	0.31	0.28	0.38	0.56	0.53
Arts	1.24	1.04	2.43	.	7.73	4.24	7.36	1.96	3.18	0.28	0.49	1.83	0.65	1.32	0.13
Humanities	5.58	2.27	4.43	26.17	.	15.14	35.76	4.28	13.47	0.92	1.73	1.86	1.29	2.75	1.41
Communication	4.48	3.97	5.47	18.27	21.36	.	16.53	24.37	13.92	3.68	7.39	5.02	7.34	6.45	1.21
Social Science	3.99	3.78	3.39	25.68	39.49	14.65	.	11.68	25.09	2.58	3.66	5.66	3.36	7.22	6.57
Business Admin	1.36	2.42	1.73	11.02	9.76	33.95	10.64	.	23.27	9.31	25.28	21.25	26.19	22.26	1.50
Law	0.76	1.82	0.91	6.02	8.69	7.97	12.80	12.19	.	2.61	3.22	5.05	5.18	6.38	2.19
Life Sciences	0.76	12.03	1.20	0.91	0.83	4.99	0.97	6.66	3.35	.	11.30	8.59	14.52	9.14	29.61
Physics/Math	0.40	1.17	0.63	0.58	0.92	2.23	1.42	7.09	2.01	5.00	.	6.49	11.46	6.28	4.05
ICT	0.10	0.11	0.21	1.47	0.71	1.35	0.70	5.43	1.41	1.59	4.92	.	9.03	7.69	0.22
Engineering	0.55	0.83	0.71	0.83	0.74	2.16	1.17	11.86	3.47	10.52	26.76	26.85	.	25.21	2.91
Architecture	0.62	0.63	0.82	1.73	1.89	2.93	3.71	7.47	3.52	3.53	7.69	13.70	15.47	.	0.76
High Health	4.12	45.00	1.65	1.25	1.85	4.66	3.35	3.72	3.46	39.25	5.30	2.15	3.75	2.98	.

Notes: Displays the estimated unconditional substitution patterns of students in each field who face a 1% shock to the NPV of that field. Columns are initial field, rows are substitute field. Each cell is the fraction of switchers moving from the initial field to a new field.

However, with heterogeneity, this relationship no longer holds, becoming:

$$\frac{d\pi_{m'}/d\log V_m}{d(1-\pi_m)/d\log V_m} = -\frac{\sum_i \sum_l q_{il}\pi_{im'}\pi_{im}\theta_l V^e_{im}}{\sum_i \sum_l q_{il}(1-\pi_{im})\pi_{im}\theta_l V^e_{im}},$$

where the subscript i indexes individual student characteristics, including latent type.

Table 5 shows the substitution matrix, with shocked fields in columns and alternatives in rows (each column sums to 100%). In line with earlier discussions, the patterns reveal distinct preference clustering. Two striking examples are Education and Engineering. For Education, substitution occurs mostly toward Care and Medium Health. Similarly, following a decline in the NPV of Engineering, 25% of switching students opt for Physics/Math or Life Sciences—fields accounting for only 5.5% of total students—capturing realistic clustering patterns. This matrix is largely in line with the second choice data that we plot in Table 1.

Table A.7 shows substitution without type heterogeneity, relying solely on first-choice data, while still constraining choice sets by high school track and GPA bounds. In this simplified model, unrealistic patterns emerge—for all fields, a shock pushes 10% of switchers into Education. This stark contrast emphasizes the critical role of modeling preference heterogeneity.

5.3.2 Labor Market

We now discuss estimates of labor market parameters governing workers' occupational reallocation. Appendix B.7 contains the full set of parameter estimates, with standard errors, and additional tables on the responsiveness of workers by fields of study.

Table 6 displays estimates of the labor supply parameters that govern how workers reallocate across occupations. The first column is estimated by ignoring worker heterogeneity, as in Artuç et al. (2010), while the second reflects our full model, with four labor market types ($K = 4$), test scores where available, and field of study. Controlling for heterogeneity in type k and field m dramatically raises the measured responsiveness of occupation switching to wage differentials, and consequently leads to much smaller estimated switching costs.

The first row is our estimate of $1/\nu$, scaled so that average wages are equal to 1.[29] The second and third rows report our estimated switching costs, both averaged over workers who switch and weighted by their observed moves. The second row reports C relative to last period's earnings; the third reports C net of the expected preference shock a mover draws, $E(\varepsilon_{o'} - \varepsilon_o)$.[30] Accounting for

[29]For each individual, we take the q_{ik} weighted average across types.

[30]We normalize by ε_o so that the expected cost of staying is 0. For a stayer the destination and origin shocks coincide and cancel. Using the Gumbel expressions for the expected shocks conditional on the choice, the expected

human capital lowers both the scale and variance of costs. In our full model the average switching cost is on the order of 2–3 years of income, depending on whether expected shocks are included. Ignoring human capital raises C to about 26 years of income, and the variance of the shocks is then so large that the expected cost net of shocks turns large and negative.

Table 6
Labor Supply Parameters

Parameter	(1)	(2)
$\bar{w}/\nu$	0.25	1.67
	(.100)	(.160)
C/w_{t-1}	26.13	3.08
	(33.47)	(5.80)
$(C - E(\varepsilon_{o'} - \varepsilon_o))/w_{t-1}$	-4.79	2.22
	(15.39)	(5.42)

Notes: The first column reports the estimates using Artuç et al. (2010), while the second reports the main specification. In the first column, standard errors are clustered by year, ignoring first-stage error in estimating average returns. In the second column, standard errors are calculated using a bootstrapping procedure that adjusts for both first-stage errors and clusters by year (see Appendix B.3). The first row presents the estimates for $1/\nu$ in (4), scaled by the mean wage. The second row presents the population-weighted average costs paid by switchers relative to their previous period's earnings. The third row presents the population weighted average costs paid by switchers net of expected shocks relative to their previous period's earnings. In both cases, the parentheses are the population standard deviation in costs paid by switchers (relative to earnings).

The reason controlling for this heterogeneity matters is comparative advantage. A worker's human capital $H_o(k, m)$ varies with her field m, through the field–occupation productivities α_{om}, and with her latent type k, through her type–occupation productivities, α_{ok}. Because she is most productive, and best paid, in the occupations suited to her type and field, she tends to stay in them. A model that treats workers as homogeneous, as in Artuç et al. (2010), cannot see this and reads the resulting persistence as high switching costs. As a result, workers move only on extreme draws. In fact, the movers are so selected on large draws that their average draws exceeds the fixed cost, and therefore the expected cost net of shocks turns negative. Controlling for k and m separates comparative advantage from genuine frictions, which is why our estimated costs are both smaller and far less dispersed. The types are also economically meaningful: Appendix B.6 shows our four types sort sharply across occupations, matching the comparative advantage implied by the estimated α_{ok}.

In order to make the parameters economically interpretable, we calculate the semi-elasticity of exiting an occupation with respect to a temporary income shock. That is to say, we look at the

cost of a move from o to o' is

$$C_{oo'km} + \nu \left(\log \lambda(o' \mid o, k, m) + \frac{\lambda(o \mid o, k, m)}{1 - \lambda(o \mid o, k, m)} \log \lambda(o \mid o, k, m) \right),$$

averaged across observed switchers.

percentage point increase in exit following a 1% temporary income shock. Given the structure of the model, this can be calculated as,

$$\frac{d(1-\pi_{oo})}{d\log w_o} = -\frac{w_o}{\nu}\pi_{oo}(1-\pi_{oo}). \tag{16}$$

We can calculate this for each worker, given their background and type, and take a weighted average across the economy. The semi-elasticity is 0.13, meaning that a 1% temporary decline in earnings raises the exit rate by 0.13 pp. While this figure may seem small, it reflects the small magnitude of a 1% temporary shock. A permanent wage shock can be approximated by $w_o/(1-\beta)$.[31] In this case, a 1% permanent decline in earnings would raise the exit rate by approximately 1.7 pp—a significant increase given that mean turnover in a year is 15%. Table A.5 displays (16) for each field of study, showcasing substantial heterogeneity in the responsiveness across fields.

In summary, ignoring field heterogeneity, selection, and latent worker differences biases the estimated switching costs upward and makes workers look less responsive to wage changes than they are. The next section uses our estimates in a set of counterfactuals, tracing how field choice and occupational switching shape the aggregate and distributional effects of two labor market shocks.

6 Counterfactual Analysis

This section uses the estimated model to study how the economy absorbs labor market disruptions. We first close the model by specifying labor demand, then introduce two disruptions, a trade war and an AI-driven shock to the value of skills. For each disruption, we report the long-run aggregate impact, the transition, the value of education reform, and the distributional consequences.

6.1 Closing the Model

We specify goods markets and labor demand, with full details, including equilibrium conditions, in Appendix C.1. Let $s \in \mathcal{S}$ index industries, and denote Denmark and Foreign by D and F, respectively. There is a final non-traded consumption good, which is a nested CES aggregate of industrial outputs, combining both domestic and foreign varieties. At the top level, the final good aggregator, Y_t, is defined as:

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

[31] Technically, the value function changes both directly through the wage decline and indirectly through ripple effects on switching probabilities. For simplicity, we focus only on the direct effect.

where Y_{st} represents industrial outputs, and μ_s is an exogenous demand shifter for industry s. Industrial outputs are further disaggregated into non-traded combinations of domestic and foreign varieties:

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

where Y_{st}^D and Y_{st}^F are domestic and foreign varieties, respectively, with corresponding price indices P_{st}^D and P_{st}^F. The elasticity parameters σ and η satisfy the standard condition $\eta > \sigma > 0$.

In addition to domestic demand, there is external demand for Danish exports, captured by the following demand curve:

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

where A_{st}^F is an exogenous demand shifter for exports.

Domestic Production

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 combining production and non-production services. We classify occupations into two groups: production and non-production, denoted by $\mathcal{O}^P$ and $\mathcal{O}^N$, respectively. The production process follows the CES function:

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

where L_{st}^N denotes the total output of non-production occupation services, L_{st}^P represents the total output of production occupation services, and Q_{st}^M is the aggregated imported material inputs. The variables A_{st}, A_s^N, A_s^P, and A_s^M denote the sector-specific total factor productivity (TFP) and the productivities of non-production services, production services, and imported inputs, respectively. The parameter $\alpha > 0$ represents the elasticity of substitution between L_{st}^N and the production aggregate, while $\rho > 0$ denotes the elasticity of substitution between L_{st}^P and Q_{st}^M.

Both production and non-production inputs are modeled as Cobb-Douglas aggregates over occupations. For $f = \{N, P\}$, we have:

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

where $\sum_{o\in\mathcal{O}^f} \beta_{so} = 1$, and L_{sot} represents the efficiency units of human capital within each occupation and industry. Since workers are indifferent across sectors within an occupation, the total human capital in an occupation is given by: $L_{ot} = \sum_s L_{sot}$. Finally, consistent with the labor supply model, workers allocate their entire income within the period it is earned. Thus, aggregate expenditure equals aggregate income:

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

We calibrate the demand-side parameters from aggregate labor shares or rely on values derived from existing literature. Further details are provided in Appendix C.2.

6.2 Setting Up the Experiment

We specify the two disruptions, then the three scenarios that govern how freely entering cohorts may re-choose fields.

6.2.1 The Two Shocks

The two disruptions operate through different channels: the trade war on the demand for occupations, the AI shock on the value of skills.

A trade war We model the trade war as a 30% increase in the cost of imported goods and offshored inputs, together with a 30% reduction in foreign demand for Danish exports. The scenario captures renewed protectionist policies and the retaliation they invite. Although the shock is uniform across sectors, exposure varies widely across occupations.[32] Production occupations and tradable services are the most exposed (Figure C.1a), and the dispersion carries over to fields of study (Figure C.1b).

An AI shock to skills AI is expected to substitute for some skills and complement others, many of them acquired in college.[33] As AI automates document analysis, for instance, a law degree loses value in every occupation its holder might enter. We therefore model AI as an unanticipated shock to the estimated field–occupation productivities α_{om}, the component of human capital a field of study contributes in each occupation, defined in (2). The shock scales all of a field's productivities by a single factor, calibrated to the occupation-level AI exposure of Eloundou et al. (2023) (Figure

[32] We measure each occupation's exposure before any equilibrium adjustment, as an average of its sectors' net reliance on foreign markets (export intensity minus imported-input intensity), weighting each sector by its share of the occupation's wage bill. Field exposure aggregates this over the occupations each field's graduates enter. Appendix C.3.1 gives the exact construction.

[33] Webb (2020), Felten et al. (2021), and Eloundou et al. (2023) construct occupation-level measures of exposure to AI; Deming and Noray (2020) document rapid obsolescence of skills in technology-intensive majors.

C.2a), aggregated over the field's entry occupations. Fields whose entry occupations are more exposed than average lose human capital (e.g., Law, ICT, Physics/Math, Life Sciences), while fields below average gain (e.g., Medium Health, Education, Care). The gains capture the rising relative return to skills AI cannot replicate (Figure C.2b). We normalize the shock so that aggregate human capital is unchanged at the initial field composition, which separates the *reallocation* consequences of AI from any assumption about its aggregate effect. Appendix C.3.2 gives the full construction.

For both shocks, the ex-ante exposure is the direct effect in partial-equilibrium. It need not translate into the realized impact, because workers can reallocate across occupations and new cohorts across fields. The general-equilibrium model determines whether they do, and through which margin.

6.2.2 The Counterfactual Scenarios

How much of a disruption the economy absorbs depends in part on how freely entering cohorts can re-choose fields, and that freedom is set by education policy. We compare three scenarios that differ only in how the education margin may respond. The first is a benchmark that shuts the margin off. The other two are the policies of interest.

- *No Reallocation*: entrants keep the field choices of the pre-shock steady state.
- *Current System*: entrants re-choose their fields under the two restrictions of the Danish admission system (Section 2.3): the high-school tracking that limits which programs a student may apply to, and the capacity caps on oversubscribed programs.
- *Flexible Response*: both restrictions are removed when the shock arrives, so entrants re-choose freely.[34]

We model admission probabilistically: a student with test score s and background h enters field m with probability $P(m \mid h, s)$, recovered from historical acceptance rates and held fixed across counterfactuals.[35] Removing the capacity constraints raises these probabilities to one; removing tracking expands the set of programs a student may rank.

Each counterfactual requires two computations, the new steady state and the transition to it. The baseline for both is the pre-shock steady state, the stationary equilibrium the model reaches

[34] We also study reforms that lift only one of the two restrictions, and a reform that holds the High Health field fixed. Section 6.3.3 and Appendix C.6 report these variants.

[35] This is a tractable stand-in for the deferred-acceptance assignment used in estimation, adopted because we do not model program acceptance conditional on field choices. Appendix C.4 provides more detail and shows that the resulting steady state reproduces the observed field–occupation allocation closely.

from the calibrated 2018 economy with no shock. We solve both in relative changes from this baseline, using exact hat algebra, which requires only the calibrated expenditure and labor shares rather than the levels of the demand-side primitives (Dekle et al., 2008).[36] The post-shock steady state is the set of wages, prices, and stationary distribution of workers over fields, types, and occupations at which all markets clear under the new fundamentals. Comparing it to the pre-shock steady state gives the long-run aggregate effects, after the workforce has fully turned over and every worker is a new entrant.

We then solve for the transition between the two steady states. Each disruption arrives unanticipated in the pre-shock steady state, after which workers and students hold perfect foresight over the adjustment path.[37] We compute the path with a shooting method, solving continuation values backward from the post-shock steady state, rolling the distribution of workers over fields, types, and occupations forward from the pre-shock one, and iterating the wage path until labor markets clear in every period. Appendix C.8 gives the full solution algorithm.

Welfare is the mean change in log lifetime labor income of a cohort, relative to the same cohort in the pre-shock steady state. In the steady-state comparison, every worker is a new entrant. Along the transition we distinguish two groups. Entrants are the first cohort to choose a field after the shock. Incumbents are workers already in the labor market when it arrives, whose fields are fixed by construction.[38]

6.3 Counterfactual Results

6.3.1 The Long-Run Aggregate Impact

The long-run impact is the change between the pre and post-shock steady states, once the workforce has fully turned over. Panel A of Table 7 reports it under the three scenarios, for two income measures: mean lifetime income (the welfare of the college cohorts that choose fields) and GDP (the

[36]In estimation, wages are measured relative to the average wage, so the price level plays no role (Table 6). In the counterfactuals prices move across equilibria, and we deflate wages, and the field values computed from them, by aggregate price index from the demand side, so all reported changes are real.

[37]Perfect foresight is the standard benchmark in this literature (e.g., Dix-Carneiro, 2014; Caliendo et al., 2019; Traiberman, 2019). The belief protocol is a design choice. Under myopic beliefs, for instance, agents would expect current conditions to persist. Under perfect foresight, agents make no systematic forecasting errors about aggregates once the shock has arrived. Individual careers remain stochastic, but education and career decisions respond to the wage and price paths the economy realizes.

[38]We simulate cohorts with identical preference and productivity draws across all scenarios, so differences reflect the shock and the policy environment alone.

aggregate real income of all workers).[39] The gap between No Reallocation and the Current System isolates the education margin: what field choice contributes under the admission constraints. We take the two shocks in turn, reporting in each case the aggregate outcome and then the occupation-wage and field-value changes that produce it.

Table 7
Long-Run Aggregate Effects of the Two Disruptions

	Trade War			**AI**		
	No Reall.	Current	Flexible	No Reall.	Current	Flexible
Panel A: long-run aggregate impact						
Δ Mean lifetime income (%)	−16.17	−16.36	−11.63	+1.62	+2.98	+5.26
Δ GDP (%)	−14.32	−14.37	−13.28	+0.78	+1.15	+1.70
Field reallocation (%)	0.0	0.5	19.2	0.0	3.3	19.6
Panel B: value of the education policy margin						
Value of field adjustment (Current − No Reall.)		−0.19			+1.36	
Value of Flexible Response (Flexible − Current)		+4.73			+2.28	
of which standing reform value		+4.46			+4.46	
of which shock–reform interaction		+0.27			−2.18	

Notes: Post-shock steady state relative to the pre-shock steady state, for a trade war and an AI shock to skills. The three scenarios are *No Reallocation* (entrants keep their pre-shock fields), *Current System* (entrants re-optimize under the admission constraints), and *Flexible Response* (the constraints are removed). *Panel A* reports long-run aggregate impacts. *Mean lifetime income* is the mean change in log lifetime labor income ($\times$100) of a cohort living in the given steady state, where every worker is a new entrant, and covers the college workers who choose fields. *GDP* is aggregate real income across all workers, which includes those without a college degree. *Field reallocation* is the share of students whose chosen field differs from the pre-shock base. *Panel B* decomposes the change in entrants' mean lifetime income, the top row of Panel A, into the value of the field-choice margin and of the reform, in percentage points. The *value of field adjustment* is the Current System entry minus the No Reallocation entry. The *value of the Flexible Response* is the Flexible Response entry minus the Current System entry, split into the *standing reform value* (removing the constraints with no shock) and the *shock–reform interaction* (the remainder).

The trade war lowers mean income by 16.36% under the Current System.[40] With fields frozen, the loss is nearly identical, at 16.17%. The loss is slightly larger under the Current System, by 0.19 pp, from congestion as the few students who switch crowd into the fields they enter. Output falls by about the same amount in the two scenarios, 14.37% against 14.32%. Removing the admission constraints cuts the income loss to 11.63%. Most of this gain is the standing value of the reform, not adaptation to the trade war, a distinction we develop in Section 6.3.3.

The education margin does little because the trade war does not change the relative value of fields. The shock falls unevenly across occupations, where occupation wages fall by 16.2% on average with a standard deviation of 6.0 log points, but it falls nearly evenly across fields, where field values fall by 3.8% on average with a standard deviation of only 0.7 (Table C.2). When wages fall in the exposed occupations, workers shift toward the less affected ones, so no field bears the full loss of the

[39] The two differ because workers without a college degree, who enter GDP but do not choose fields, are less exposed to both disruptions.

[40] All income, wage, and value changes in this section are log changes $\times$100. We refer to individual changes as percent and state dispersion across occupations or fields in log points.

occupations its graduates start in. Because every field declines by about the same amount, switching field gains little, and only 0.5% of students change field. The estimated occupational switching elasticity governs how far workers can arbitrage a shock across occupations, so the divergence between the two shocks is what our estimates imply.

The AI shock raises mean lifetime income by 2.98% under the Current System. With fields frozen, the gain is 1.62%, so re-choosing fields nearly doubles it and is worth +1.36 pp. Output rises as well, 1.15% against 0.78%. Removing the admission constraints raises the gain in lifetime income further to 5.26%.

The education margin matters here because AI changes the relative value of fields. AI lowers the value of a field's skills rather than the demand for occupations, so a worker carries the loss into every occupation she might enter and cannot move away from it. The shock barely moves occupation wages, with a cross-occupation standard deviation of 1.1 log points against 6.0 for the trade war. It falls instead on fields. The shock is normalized to hold aggregate human capital fixed, so average field value is unchanged by construction, but they disperse sharply—a standard deviation of 8.9 against 0.7 under the trade war (Table C.2). Because exposed fields lose and complemented fields gain, field choice now responds with 3.3% of students changing field. Appendix C.5 provides more details.

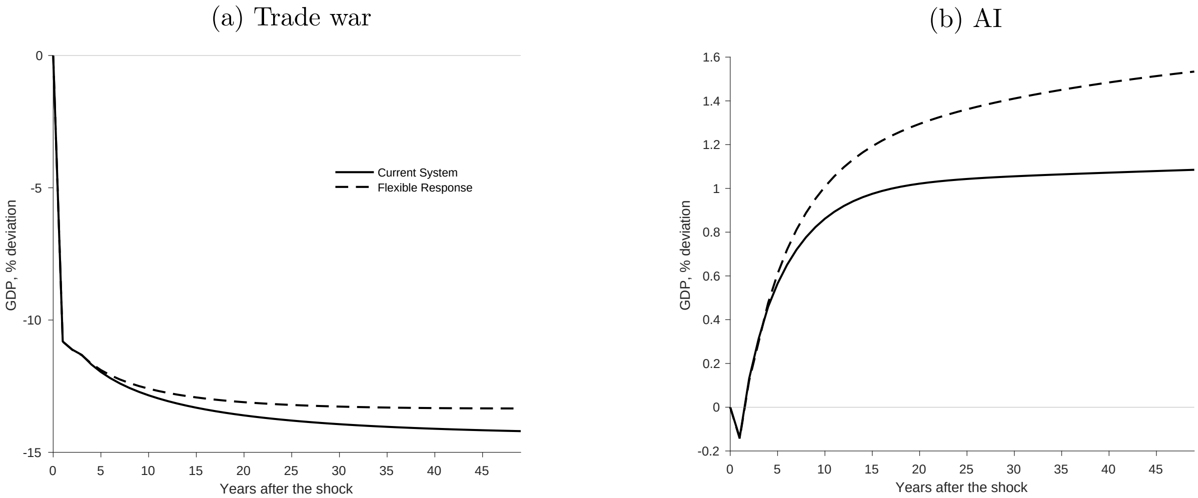
6.3.2 The Transition

The steady-state comparison sets aside the transition path between the two steady states, where the two shocks adjust at different speeds. Figure 2 traces GDP after each. The trade war is absorbed by occupational mobility, with workers moving across occupations as soon as the shock arrives, so most of the long-run loss is realized on impact: GDP falls by about 11% immediately and settles at about −14% within a decade. The AI shock cannot be absorbed this way. It works through the field choices of entering cohorts, so it builds up only as the workforce turns over: GDP is flat on impact and rises to about +1%, half within five years and most within two decades.

Entrants and incumbents fare differently along this path. Under the trade war, entrants do worse than incumbents, −16.24% against −13.65% under the Current System, because the loss deepens slightly over the transition.[41] Under AI the order reverses, with both groups gaining. But

[41] An incumbent is already working when the shock arrives and earns through the early years while the loss is still shallow. An entrant reaches the labor market only after she graduates, by which point the decline has set in, so she spends her whole career on the deepened path.

Figure 2
Adjustment of GDP under the Two Disruptions



Notes: GDP (percent deviation from the no-shock economy) along the perfect-foresight transition, by years after the shock. Panel (a): trade war; panel (b): AI. Solid line: Current System; dashed line: Flexible Response.

entrants gain more than twice as much, +3.18 against +1.43, because only entrants can choose the AI-complemented fields. These averages hide wide differences within each group, which Section 6.3.4 takes up.

6.3.3 The Value of Education Policy

We now ask what the *Flexible Response*, which enhances field choice flexibility by removing the admission constraints, is worth. We evaluate it in normal times and under each shock. The value in normal times is the reform's *standing value*. The remainder is specific to each shock. The Flexible Response is best read as a limiting case, the upper bound of flexibility an education reform can deliver. Our model also does not take a stand on the objective a planner should pursue. It instead gives the economic and distributional consequences of the reform, which are an input to any such objective.[42]

The standing value is large: removing the constraints raises mean lifetime income by 4.46 pp with no shock (Table 7, Panel B). Under the trade war the reform is worth 4.73 pp, almost all of it the standing value, as the shock adds only +0.27. Under AI it is worth 2.28 pp, about half its standing value, so the shock-specific part is −2.18 pp.

[42]We remain agnostic on the source of national objectives: they could be paternalistic, if student preferences reflect incomplete information (Hastings et al., 2016), or fiscal, if there are budgetary reasons to direct students toward higher-output fields. Two caveats apply to the reform itself. We abstract from the cost of expanding capacity, and we assume students complete the programs they enter, which may overstate graduation in some fields (Ahn et al., 2024; Arcidiacono et al., 2025).

The difference lies in which fields the constraints keep students out of. Both restrictions bear on the selective, high-return fields: the tracking bars students from fields outside their secondary-school track, and the caps ration seats in the oversubscribed ones. The students they turn away enroll instead in fields such as Medium Health, Education, and Care. With no shock, removing the constraints lets these students enter the high-return fields, Law, ICT, Physics/Math, and Life Sciences, where their income more than doubles. AI lowers the value of these high-return fields and raises the value of the unconstrained fields that had absorbed them. Entering the high-return fields no longer raises income, so only students with a large enough preference for them still do. The access the reform grants is therefore worth far less than in normal times. The trade war, by contrast, does not change the relative field values (Table C.2), so the rationed fields are no more affected than the rest and the reform keeps its standing value.

This value builds only over the long run. Because the reform works through the field choices of entering cohorts, the dashed lines in Figure 2 coincide with the solid ones on impact and separate only slowly, reaching about a quarter of the eventual gap within a decade and half within two decades. The policy thus does little for the generation a disruption hits, and its benefit accrues to later cohorts.

Two further results are in the appendix. First, lifting the reform's two restrictions separately shows that almost all of its value comes from removing the academic tracking rather than the capacity caps (Appendix C.6). The same appendix also reports a variant that holds the High Health field fixed, since our framework may fit it poorly and the field draws many students. This variant leaves the conclusions unchanged. Second, none of the counterfactual results is specific to Denmark's admission constraints. In an economy that never had them, hit by the same two shocks, re-choosing a field is again valuable under AI and negligible under the trade war (Appendix C.7).

6.3.4 Who Bears the Disruptions

Section 6.3.2 shows that a disruption falls differently across generations. Within a generation, the distributional consequences are just as uneven. Figure 3a plots the change in lifetime income under the Current System against pre-shock lifetime income, for each shock.

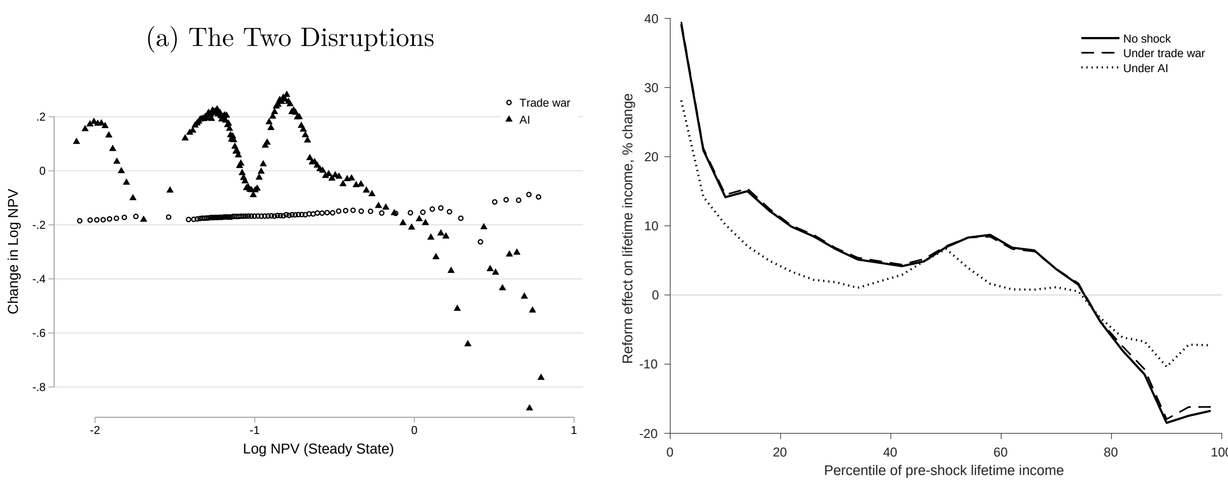
The trade war falls on everyone similarly, with losses between 10 and 18% over the whole distribution, larger at the bottom. This is the distributional counterpart of the aggregate result in Section 6.3.1. The trade war pushes all field values down together, so no field, and no part of the income distribution, escapes.

The AI shock does the opposite. It raises income across most of the distribution but lowers it sharply at the top, where the highest earners lose between 40 and 80% of their lifetime income. In between, gains and losses sit at similar income levels. This is because income is set mainly by a worker's latent type k, so every field holds workers throughout the distribution. Workers from the fields AI raises and the fields it lowers mix at every income level, in similar numbers through the middle. However, the mix shifts at the top, where the devalued fields have the highest average returns. The top decile comes mostly from these fields, with 83% of the workers there holding a degree in a devalued field, twice the share in the workforce as a whole. AI therefore narrows the distribution.

Figure 3
Distributional Incidence of the Disruptions and the Reform

(a) The Two Disruptions

(b) The Education Reform



Notes: Binned scatter of the change in log lifetime NPV against log pre-shock lifetime NPV for workers with college degrees. *Panel (a)*: the post-shock Current System relative to the pre-shock economy—open circles, trade war; filled triangles, AI. *Panel (b)*: the reform's effect on lifetime income, the Flexible Response minus the Current System, against log pre-shock lifetime NPV. Solid, no shock; dashed, trade war; dotted, AI. Alternative reforms are in Appendix C.6.

The admission reform redistributes as well, raising incomes in the lower half of the distribution and lowering them at the top (Figure 3b).[43] Opening the selective, high-return fields admits students the constraints had excluded, which raises the bottom, while their entry bids down returns in those fields and lowers the top.[44] The disruptions barely change this shape. Under the trade war the incidence almost coincides with the no-shock case, because the trade war moves field values together.

[43]The profile is not perfectly monotone, with the gain reaching a local peak just above the median. The peak reflects a large flow of students out of Care, the largest field and among the lowest-return, into the high-return fields the reform opens. The movers' gains are largest for those just above the median, and netting them out leaves a broadly progressive profile.

[44]Appendix C.6 reports the same exercise for the academic tracking and the capacity caps separately and finds similar patterns.

Under AI the same force that shrinks the reform's value also flattens the incidence. AI lowers the value of the high-return fields the reform opens, so entering them no longer raises income. Students moving up from the low-return fields gain less, and the lower half rises less. At the same time, fewer students enter those fields, and the top falls less.

7 Conclusion

We model and estimate how students choose a field of study, and how that choice shapes their later careers, using administrative data from Denmark that let us separate heterogeneity in preferences over fields from heterogeneity in comparative advantage across occupations. The estimates show that fields differ in the skills and comparative advantages they confer, that accounting for this raises the estimated responsiveness of workers to income and lowers their estimated switching costs, and that students respond to expected income but only modestly, concentrating their choices within clusters of related fields.

We put the estimated model to work on two labor market disruptions, a trade war and an AI shock to skills. The experiments show that field choice is an active margin of adjustment only when a shock changes the relative value of fields. The trade war does not. Workers reallocate across occupations until its incidence is a common decline across fields, so field choice barely responds and the loss is shared across the income distribution. The AI shock does. It lowers the value of the skills some fields provide, so field choice responds and the losses fall on the holders of the devalued degrees. Whether education policy helps therefore depends on the shock. Removing Denmark's admission constraints increases lifetime income by 4.46 pp in normal times. This value barely changes under the trade war and roughly halves under AI, which devalues the very fields the constraints had rationed.

A natural next step is to endogenize the choices we hold fixed. We take high-school track as given, but it, like the decision to enter college at all, responds to the same labor-market incentives the model already captures, and bringing those margins inside the model would extend its reach. Other simplifications invite the same treatment: institution quality, family background, and the assumption that students forecast future income rationally rather than under limited information, as well as the supply side of education, whose costs a full reckoning of the reform's value would need. We see these as promising directions for future research.

References

Abramitzky, Ran, Victor Lavy, and Maayan Segev, "The Effect of Changes in the Skill Premium on College Degree Attainment and the Choice of Major," *Journal of Labor Economics*, 2024, *42* (1), 245–288.

Acemoglu, Daron and David H. Autor, "Skills, Tasks and Technologies: Implications for Employment and Earnings," in Ashenfelter O. and Card D., eds., *Handbook of Labor Economics*, 1 ed., Vol. 4B, Elsevier, 2011, chapter 12, pp. 1043–1171.

Adão, Rodrigo, Martin Beraja, and Nitya Pandalai-Nayar, "Technological Transitions with Skill Heterogeneity Across Generations," NBER Working Papers 26625, National Bureau of Economic Research, Inc 2020.

Adao, Rodrigo, Martin Beraja, and Nitya Pandalai-Nayar, "Fast and Slow Technological Transitions," *Journal of Political Economy Macroeconomics*, 2024, *2* (2), 183–227.

Agarwal, Nikhil and Paulo Somaini, "Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism," *Econometrica*, 2018, *86* (2), 391–444.

Ahn, Tom, Peter Arcidiacono, Amy Hopson, and James Thomas, "Equilibrium Grading Policies With Implications for Female Interest in STEM Courses," *Econometrica*, 2024, *92* (3), 849–880.

Alon, Titan, Natalie Cox, and Minki Kim, "Debt, Human Capital, and the Allocation of Talent," 2024. Revise and resubmit, *American Economic Review.*

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.

_ , Esteban Aucejo, Arnaud Maurel, and Tyler Ransom, "College Attrition and the Dynamics of Information Revelation," *Journal of Political Economy*, 2025, *133* (1), 53–110.

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.

Backus, David K., Patrick J. Kehoe, and Finn E. Kydland, "Dynamics of the Trade Balance and the Terms of Trade: The J-Curve?," *The American Economic Review*, 1994, *84* (1), 84–103.

Beffy, Magali, Denis Fougere, and Arnaud Maurel, "Choosing the field of study in postsecondary education: Do expected earnings matter?," *Review of Economics and Statistics*, 2012, *94* (1), 334–347.

Blanchard, Emily and Gerald Willmann, "Escaping a protectionist rut: Policy mechanisms for trade reform in a democracy," *Journal of International Economics*, 2011, *85* (1), 72–85.

_ **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.

_ **and** _ , "College Major Restrictions and Student Stratification," Technical Report, CESifo Working Paper Series No. 11601 2024.

Bonhomme, Stephane, Thibaut Lamadon, and Elena Manresa, "A Distributional Framework for Matched Employer Employee Data," *Econometrica*, 2019, *87* (3), 699–739.

Bordon, Paola and Chao Fu, "College-Major Choice to College-Then-Major Choice," *The Review of Economic Studies*, 05 2015, *82* (4), 1247–1288.

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 and Fernando Parro, "Estimates of the Trade and Welfare Effects of NAFTA," *Review of Economic Studies*, 2015, *82* (1), 1–44.

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

Calonico, Sebastian, Matias D Cattaneo, Max H Farrell, and Rocio Titiunik, "rdrobust: Software for regression-discontinuity designs," *The Stata Journal*, 2017, *17* (2), 372–404.

Caucutt, Elizabeth M. and Lance Lochner, "Early and Late Human Capital Investments, Borrowing Constraints, and the Family," *Journal of Political Economy*, 2020, *128* (3), 1065–1147.

Conlon, Chris, Julie Mortimer, and Paul Sarkis, "Estimating Preferences and Substitution Patterns from Second-Choice Data Alone," Technical Report, New York University 2024. Accessed: 2024-08-14.

Conlon, John J and Dev Patel, "What jobs come to mind? stereotypes about fields of study," 2022.

Dekle, Robert, Jonathan Eaton, and Samuel Kortum, "Global rebalancing with gravity: Measuring the burden of adjustment," *IMF Staff Papers*, 2008, *55* (3), 511–540.

Deming, David J and Kadeem Noray, "Earnings Dynamics, Changing Job Skills, and STEM Careers*," *The Quarterly Journal of Economics*, 06 2020, *135* (4), 1965–2005.

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

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

Dvorkin, Maximiliano and Alexander Monge-Naranjo, "Occupation Mobility, Human Capital and the Aggregate Consequences of Task-Biased Innovations," Working Papers 2019-13, Federal Reserve Bank of St. Louis 2019.

Eckardt, Dita, "Training Specificity and Occupational Mobility: Evidence from German Apprenticeships," 2024.

Eloundou, Tyna, Sam Manning, Pamela Mishkin, and Daniel Rock, "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models," 2023.

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.

Felten, Edward, Manav Raj, and Robert Seamans, "Occupational, industry, and geographic exposure to artificial intelligence: A novel dataset and its potential uses," *Strategic Management Journal*, 2021, *42* (12), 2195–2217.

Ferriere, Alex, Ricardo Reyes-Heroles, and Gaston Navarro, "Escaping the Losses from Trade: The Impact of Heterogeneity and Skill Acquisition," Technical Report 2021.

Gemici, Ahu and Matthew Wiswall, "Evolution of Gender Differences in Post-Secondary Human Capital Investments: College Majors," *International Economic Review*, 2014, *55* (1), 23–56.

Greene, William H and David A Hensher, *Modeling ordered choices: A primer*, Cambridge University Press, 2010.

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

Grigsby, John, "Skill Heterogeneity and Aggregate Labor Market Dynamics," Mimeo, Princeton University 2021.

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.

_ , _ , Anely Ramirez, and Seth D. Zimmerman, "(Un)informed college and major choice: Evidence from linked survey and administrative data," *Economics of Education Review*, 2016, *51*, 136–151. Access to Higher Education.

Heckman, James J., John Eric Humphries, and Gregory Veramendi, "Returns to Education: The Causal Effects of Education on Earnings, Health, and Smoking," *Journal of Political Economy*, 2018, *126* (S1), S197–S246.

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

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

Hsieh, Chang-Tai, Erik Hurst, Charles I. Jones, and Peter J. Klenow, "The Allocation of Talent and U.S. Economic Growth," *Econometrica*, 2019, *87* (5), 1439–1474.

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

_ , Jakob R. Munch, and Pernille Plato, "Changing Tracks: Human Capital Investment after Loss of Ability," *American Economic Review.* Forthcoming.

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.

Imbens, Guido W and Thomas Lemieux, "Regression discontinuity designs: A guide to practice," *Journal of econometrics*, 2008, *142* (2), 615–635.

Kapor, Adam J., Christopher A. Neilson, and Seth D. Zimmerman, "Heterogeneous

Beliefs and School Choice Mechanisms," *American Economic Review*, May 2020, *110* (5), 1274–1315.

Kapor, Adam, Mohit Karnani, and Christopher Neilson, "Aftermarket frictions and the cost of off-platform options in centralized assignment mechanisms," *Journal of Political Economy*, 2024, *132* (7), 2346–2395.

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

Khanna, Gaurav and Nicolas Morales, "The IT Boom and Other Unintended Consequences of Chasing the American Dream - Working Paper 460," Working Papers 460, Center for Global Development August 2017.

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.

Larroucau, Tomas and Ignacio Rios, "Dynamic College Admissions," 2022. Working paper.

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

Lise, Jeremy and Fabien Postel-Vinay, "Multidimensional Skills, Sorting, and Human Capital Accumulation," *American Economic Review*, August 2020, *110* (8), 2328–76.

Llull, Joan, "Immigration, wages, and education: A labour market equilibrium structural model," *The Review of Economic Studies*, 2018, *85* (3), 1852–1896.

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

Matsuyama, Kiminori, "A Simple Model of Sectoral Adjustment," *The Review of Economic Studies*, 02 1992, *59* (2), 375–387.

Monras, Joan, "Immigration and Wage Dynamics: Evidence from the Mexican Peso Crisis," *Journal of Political Economy*, 2020, *128* (8), 3017–3089.

Moscarini, Giuseppe and Kaj Thomsson, "Occupational and Job Mobility in the US*," *The Scandinavian Journal of Economics*, 2007, *109* (4), 807–836.

Ransom, Tyler, "Selective Migration, Occupational Choice, and the Wage Returns to College Majors," *Annals of Economics and Statistics*, 2021, (142), 45–110.

Robles, Verónica C. Frisancho and Kala Krishna, "Affirmative Action in Higher Education in India: Targeting, Catch Up, and Mismatch," NBER Working Papers 17727, National Bureau of Economic Research, Inc January 2012.

Ruhl, Kim J., "The International Elasticity Puzzle," Working Paper 08-30, New York University, Leonard N. Stern School of Business 2008.

Seo, Jaeeun and Ryungha Oh, "Sectoral Shocks and Labor Market Dynamics: A Sufficient Statistics Approach," 2024. Working paper.

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

vom Lehn, Christian, Cache Ellsworth, and Zachary Kroff, "Reconciling Occupational Mobility in the Current Population Survey," *Journal of Labor Economics*, 2022, *40* (4), 1005–1051.

Webb, Michael, "The Impact of Artificial Intelligence on the Labor Market," *SSRN Electronic Journal*, 01 2020.

Wiswall, Matthew and Basit Zafar, "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.

Appendix For Online Publication

Field Choice, Skill Specificity, and Labor Market Disruptions

Valerie Smeets
Aarhus University & CEPR

Lin Tian
INSEAD & CEPR

Sharon Traiberman
NYU & NBER

Appendix A: Additional Tables and Figures

A.1 Additional Tables

Table A.1
Field of Study Classification

Field	Field Name	ISCED-F Code	College	Notes
1	Education	11	✓	
2	Arts	21	✓	
3	Humanities	22, 23	✓	
4	Social Sciences	31	✓	
5	Communication	32	✓	
6	Business Admin	41	✓	
7	Law	42	✓	
8	Life Sciences	51, 52	✓	
9	Physics/Math	53, 54	✓	
10	ICT	61	✓	
11	Engineering	71, 72	✓	Combined only for highest level of education
12	Manufacturing	72		
13	Architecture	73	✓	
14	Agriculture	80		
15	Medium Health	91	✓	Separated between university and professional bachelor's
16	Care	92	✓	
17	Personal Services	101, 102		
18	High Health	91	✓	Separated between university and professional bachelor's
19	Security	103		
20	Transport	104		

Notes: This table presents the authors' classification of fields of study, aggregated from the International Standard Classification of Education Fields of Education and Training (ISCED-F) codes. The ISCED-F is a framework developed by UNESCO to classify educational programs and related qualifications by fields of study, facilitating international comparability of education statistics and policies. Not all fields are part of the college (ISCED 6+) system, some are exclusively present in short-cycle education programs (ISCED 5). Checkmarks denote the fields of study for college education. See Section 2.2 for relevant discussions.

Table A.2
Occupation Classification

Occupation	**Occupation Name**	**DISCO-08 2-Digit Codes**	**Notes**
1	Management (Business + Admin)	10, 11, 12	
2	Management (Activity)	13, 14	
3	Engineering	21	
4	Medicine	22	
5	Education	23	
6	Finance, Admin, and Sales	24	
7	Software and ICT	25	
8	Social and Cultural Work	26	
9	Law	26	Split out of 3-digit code 261
10	Science Technicians	31	
11	Healthcare Assistance	32	
12	Admin	33, 34	
13	IT	35	
14	Secretarial	41	
15	Customer Services	42	
16	Accounting	43	
17	Other Admin	44	
18	Services	51, 94	
19	Retail	52	
20	Social Work	53	
21	Security	54	
22	Agriculture	60, 92	
23	Machinery	72	
24	Electrical	74	
25	Craftsperson	71, 73, 75	
26	Operators	81, 82	
27	Transportation	83	Includes 9621
28	Other Manual Work	91, 96	Excludes 9621
29	Construction	93	

Notes: This table presents the authors' classification of occupations, aggregated from the Danish version of the International Standard Classification of Occupations (DISCO-08). DISCO-08 is Denmark's national implementation of ISCO-08, a system developed by the International Labor Organization (ILO) to classify and compare occupations across countries based on the type of work performed and required skill levels. The classification enables standardized occupational data analysis in labor market research. See Section 2.2 for relevant discussions.

Table A.3
Excess Demand for Field of Study and Log Income

	Dep. Var: Excess Demand			
	(1)	(2)	(3)	(4)
$\log \text{Income}_t$	0.482***			0.476***
	(0.145)			(0.150)
$\log \text{Income}_{t+5}$		0.322**		0.281*
		(0.157)		(0.154)
$\log \text{Income}_{t+10}$			0.580***	
			(0.212)	
N	330	270	195	270
Within R^2	0.085	0.100	0.143	0.137

Note: Regressions of log excess demand on mean log future income of an education field at various time horizons. All regressions include education field and year fixed effects, as well as mean GPA entry cut-off as a control. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. See Section 3 for relevant discussions.

Table A.4
Excess Demand for Field of Study and Income: Robustness

	Dep. Var: Excess Demand					
	(1)	(2)	(3)	(4)	(5)	(6)
$Income_t$	-0.575		0.204		0.216*	
	(0.482)		(0.126)		(0.120)	
$Income_{t+5}$		-0.297		0.285**		0.303**
		(0.497)		(0.145)		(0.137)
Controls						
Capacity	No	No	Yes	Yes	Yes	Yes
Future Shares	No	No	No	No	Yes	Yes
N	330	270	330	270	315	270
Within R^2	0.005	0.134	0.932	0.938	0.943	0.946

Note: Regressions of log excess demand on mean future income of an education field at various time horizons. All regressions include education field and year fixed effects, as well as mean GPA entry cutoff as a control. Robust standard errors in parentheses: *** p<0.01, ** p<0.05, * p<0.1. See Section 3 for relevant discussions.

Table A.5
Elasticity values by Education Field

Field of Study	$\mathcal{E}_m$
Business Admin	0.281
Social Science	0.254
Engineering	0.237
ICT	0.224
Physics/Math	0.200
Law	0.192
Life Sciences	0.187
Architecture/Construction	0.177
Arts	0.140
Humanities	0.137
Communications	0.136
High Health	0.084
Care	0.063
Education	0.052
Med Health	0.048

Notes: This table calculates the partial elasticity of occupational switching for each field of study. The partial elasticity of occupational switching is defined as the percentage point change in the exit rate from the most popular occupation in a given field in response to a one percent decline in the earnings in that occupation. See Section 5.3.2 for relevant discussions.

Table A.6
Distribution of First Priority by Student Type

	Type				Population
	1	2	3	4	
Education	50.7	0.8	1.4	1.8	10.4
Medium Health	7.7	57.7	0.9	0.6	13.2
Care	31.2	1.8	15.7	0.9	11.1
Arts	0.6	0.2	0.7	5.9	2.2
Humanities	2.0	0.4	1.0	21.0	7.4
Communication	1.8	0.7	9.9	13.7	7.8
Social Science	2.0	1.2	2.7	27.5	10.2
Business Admin	0.5	0.5	25.9	10.5	11.4
Law	0.4	0.6	4.7	7.5	4.0
Life Sciences	0.4	6.0	6.5	1.1	3.6
Physics/Math	0.2	0.6	4.7	1.1	1.9
ICT	0.0	0.0	3.4	1.5	1.5
Engineering	0.3	0.5	14.1	2.2	5.1
Architecture	0.4	0.2	6.0	3.2	2.9
High Health	1.9	28.7	2.5	1.5	7.2

Notes: Columns (1)-(4) report the share of individuals of each type (weighted by the probability of being that type, q_{il}) choosing each field as their first choice. The final column reports the population shares of each top priority. See Section 5.3.1 for relevant discussions.

Table A.7

Substitution Matrix $\left(\frac{d\pi_{m'}/d\log V_m}{d(1-\pi_m)/d\log V_m}\right)$

Substitute Field	Education	Medium Health	Care	Arts	Humanities	Communication	Social Sciences	Business Admin	Law	Life Sciences	Physics/Math	ICT	Engineering	Architecture	High Health
	Initial Field														
Education	.	15.15	19.35	10.47	11.97	13.24	9.57	11.49	8.91	8.60	7.13	8.07	9.03	8.83	7.50
Medium Health	18.53	.	24.78	15.69	13.98	15.37	11.65	15.44	11.71	10.71	9.11	11.79	10.89	11.37	9.49
Care	18.53	19.59	.	11.51	10.91	13.17	7.55	12.88	7.47	6.85	5.57	6.03	6.01	6.12	5.40
Arts	1.88	2.68	2.51	.	2.14	2.16	2.19	2.06	2.15	1.57	1.59	2.13	1.43	1.64	1.98
Humanities	9.73	9.52	9.50	8.93	.	10.71	10.23	7.31	8.33	7.13	7.57	3.47	5.24	4.85	8.90
Communication	9.96	9.21	9.33	8.65	11.30	.	9.24	8.99	8.35	5.28	5.23	3.04	4.09	3.74	6.57
Social Science	9.59	9.16	7.53	10.33	12.97	12.12	.	15.99	15.59	11.27	13.47	7.95	10.22	9.29	16.65
Business Admin	11.03	12.21	11.21	11.45	11.27	12.76	13.79	.	14.23	10.80	11.35	9.26	10.22	9.82	12.12
Law	3.53	3.82	3.17	5.06	4.51	4.37	6.28	6.13	.	4.41	4.38	5.17	5.27	4.98	5.64
Life Sciences	3.15	3.29	2.65	2.90	3.41	2.45	4.37	2.62	3.48	.	6.06	7.00	9.21	6.97	5.99
Physics/Math	1.25	1.35	0.94	1.14	1.59	1.01	2.76	1.91	2.02	3.59	.	2.69	3.90	2.99	4.24
ICT	0.67	1.09	0.70	1.16	0.70	0.58	0.99	1.19	1.03	1.90	1.56	.	3.45	3.30	1.53
Engineering	3.56	4.35	2.22	3.13	3.65	2.02	5.42	4.21	4.43	12.50	10.87	16.22	.	16.59	9.46
Architecture	2.17	2.63	1.73	2.14	2.10	1.54	2.91	2.59	2.49	5.57	4.81	8.80	9.84	.	4.52
High Health	6.41	5.96	4.37	7.44	9.50	8.51	13.05	7.18	9.82	9.82	11.30	8.37	11.19	9.51	.

Notes: This table displays the estimated substitution patterns of students in each field who face a 1% shock to the NPV of that field. These are calculated assuming *one* type of student. Columns are initial field, rows are substitute field. Each cell is the fraction of switchers moving from the initial field to a new field. See Section 5.3.1 for relevant discussions.

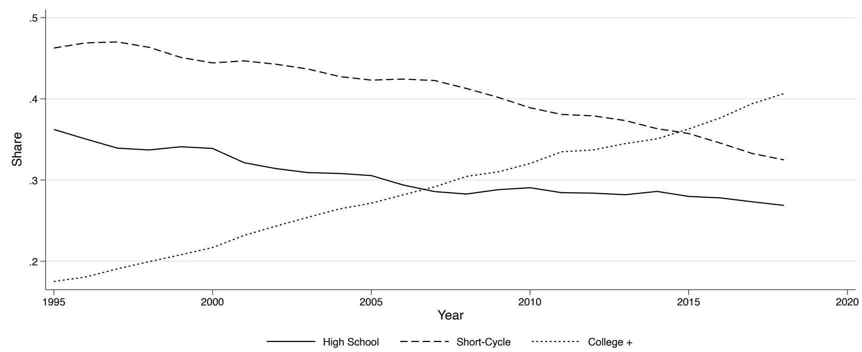
Table A.8
Field Reallocation from Removing the Capacity Caps

	Initial Field														
Counterfactual Field	Education	Medium Health	Care	Arts	Humanities	Communication	Social Sciences	Business Admin	Law	Life Sciences	Physics/Math	ICT	Engineering	Architecture	High Health
Education		0.04	11.83		0.08				0.24	0.16	0.32	0.11	0.97		
Medium Health	5.01		5.12			0.34		0.09	0.24	6.31	0.95	0.21	1.55	0.34	33.33
Care	2.00							0.17	0.48	0.41	0.95	0.42	0.97		3.03
Arts	12.64	0.14	4.55		6.40	5.03	5.33	3.66	11.84	3.28	3.89	4.53	4.18	5.76	6.06
Humanities	8.39	0.63	1.78	5.93		7.09	1.64	7.15	9.42	8.11	8.83	8.95	8.01	6.78	9.09
Communication	2.50	0.18	4.80	0.52	0.23		0.41	0.26	1.93	1.72	2.42	2.53	2.69	1.02	3.03
Social Sciences	36.30	3.47	8.24	39.18	41.31	40.57		43.07	56.04	23.91	32.28	34.74	29.31	36.95	12.12
Business Admin	3.50	0.21	10.08	3.61	2.36	3.43	0.61		7.00	4.83	6.41	5.26	6.12	3.73	
Law	14.39	1.40	13.58	44.59	44.05	33.37	83.20	37.84		25.14	27.66	31.58	24.04	34.58	24.24
Life Sciences	2.13	0.56	8.01	0.77	0.38	0.80	0.61	0.61	2.17		3.36	2.95	6.12	1.69	3.03
Physics/Math	0.38	0.04	1.40	0.26	0.23	0.69		0.35	0.48	0.57		0.42	0.40		
ICT	0.50		4.96	0.26	0.76	1.60	0.41	0.61	0.48	1.39	2.10		1.66	0.34	3.03
Engineering	0.13		2.48		0.08	0.11		0.09	0.97	0.98	0.84	0.74			3.03
Architecture	2.25	0.14	4.71	1.55	0.84	1.60	1.23	1.05	3.62	2.70	4.31	2.00	3.84		
High Health	9.89	93.21	18.47	3.35	3.28	5.37	6.56	5.06	5.07	20.48	5.68	5.58	10.13	8.81	

Notes: Each column is a student's field under the Current System; each row is the field that student chooses when the capacity caps are removed (caps lifted, high-school tracking retained), in the no-shock economy. Each cell is the share of those who switch out of the column field, conditional on switching, who move to the row field, $s_{ff'}/(1 - s_{ff})$. Switchers are identified by simulating individuals with the same characteristics and idiosyncratic shocks and computing optimal choices under each policy. The reform reallocates 8.3% of students. This table is discussed in Appendix C.6.

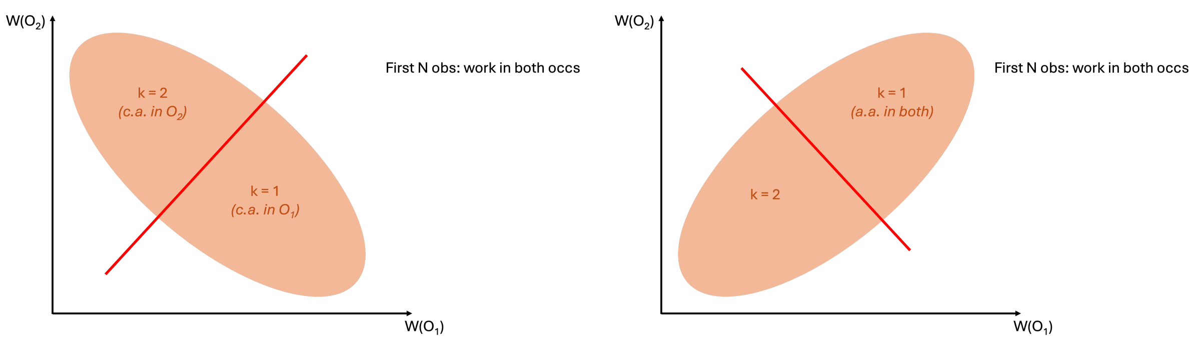
A.2 Additional Figures

Figure A.1
Education Level of New Cohorts at Labor Market Entry



Notes: This figure plots the times series of the shares of highest attained level of education across workers in their first year on the labor market. Year of labor market entry is defined as the first year of employment after the worker achieves their highest level of education (i.e., work during schooling is not counted). See Section 2.3 for relevant discussions.

Figure A.2
Intuition for the EM Algorithm



(a) Negative Correlation

(b) Positive Correlation

Notes: Panels (a) and (b) plot theoretical and appropriately time-discounted wages in occupation 1 (x-axis) against wages in occupations 2 (y-axis) for workers switching occupations. In panel (a), workers' rank changes with occupation switching, indicating likely differences in comparative advantages between the labor market types. In panel (b), workers switch occupations yet remain at stable wage ranks, which suggest one type possesses absolute advantages in both occupations over the other type. See further discussions in Section 5.2.

Appendix B: Estimation Appendix

We begin by detailing the estimation of the labor supply block, including implementation of the EM algorithm. We then discuss how we adjust the estimated labor supply parameters for workers without a college education. Next, we describe the estimation procedure for the field choice block. Finally, we outline our bootstrap approach for computing standard errors.

B.1 Labor Market Block Details

B.1.1 Log Likelihood and Initial Conditions

To identify individual types k, transition probability λ, and other human capital parameters, we first write down the following function:

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

where q_{ik} is the probability that individual i is type k, and the hats atop λ's indicate that these are estimated directly from the data rather than derived from the model.

Since initial occupations are non-random, we deal with the initial conditions problem by following Wooldridge (2005). We estimate a function, $\lambda(o_{i0}|m_i, k_i, t_{0i})$—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 are students' GPA (which we observe for a subset of workers). We parameterize $\lambda(o_{i0})$ with a logit link function. Formally,

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

We also include a dummy for when test scores are missing, which we treat as occurring exogenously.

B.1.2 Stage 2 of Labor Supply

To derive the estimating equation for the remaining parameters, we exploit the fact that the inclusive value of being in state (o, k, m, t), $V_t(k, m, o)$, is exactly the denominator in the logit probabilities. This is seen by combining equations (4) and (5). We thus may write choice probabilities as,

$$\lambda_t(o'|o, k, m) = \frac{\exp\left(\frac{-C_{oo'km} + w_{o't} H_{o'}(k,m) + \beta E_t V_{t+1}(k,m,o')}{\nu}\right)}{\exp\left(V_t(k, m, o)/\nu\right)}. \tag{B.2}$$

We then exploit the log-linearity of the resulting expression for the choice probabilities. First, take the logarithm of the relative probability of staying in an occupation versus switching to some occupation o':

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

Then we take the similar ratio in the next period, but this time consider the probability of staying in the new occupation o', versus switching from o to o':

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

Combining, we get,

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

where $\xi_{oo'kmt+1}$ is a time $t+1$ expectational error that arises because we have replaced $E_t\lambda_{t+1}$ with its realization throughout. This relationship, exploited by Artuç et al. (2010), is a special case of finite dependence, which allows for similar identification strategies in a wide class of dynamic discrete choice models (Arcidiacono and Miller, 2011; Traiberman, 2019). Under rational expectations, $\xi_{oo'kmt+1}$ is orthogonal to time t variables. In order to keep costs weakly positive, we run the following unweighted, transformed regression, via NLLS:

$$\log\left(\frac{\hat\lambda_{oo'tmk}}{\hat\lambda_{ootmk}}\right) + \beta\log\left(\frac{\hat\lambda_{o'o't+1,mk}}{\hat\lambda_{oo't+1,mk}}\right) = \underbrace{-\exp\left(\gamma_{km} + \sum_z \gamma_z|\zeta_{o'z} - \zeta_{oz}|\right) \times (1-\beta)\tilde\nu}_{\text{Moving Costs, } C} + \\ \tilde\nu\left(e^{\hat\mu_{o't}}\hat H(o',m,k) - \underbrace{e^{\hat\mu_{ot}}}_{w_{ot}}\hat H(o,m,k)\right).$$

To derive (11), we start from the same relationship over time:

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

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

B.1.3 Parameters for Non-College Workers

We estimate the labor supply block for workers without a KOT college degree in the same way as for college workers, assigning each worker a field and an education level g, and allowing for the latent types k within each group. Because we do not observe how workers select their level of education, we do not model that choice: the distribution of types within each education group is fixed across equilibria.

Estimating group by group identifies the log skill prices μ_{otg} only up to a group-level normalization of the productivities (in each group, the productivity of type 1 in occupation 1 is set to zero), so the groups are not measured in common units, while in equilibrium every group must face the same skill price per occupation. We therefore set each group's skill prices equal to the college group's and absorb the difference into its field–occupation productivities, in the year from which the counterfactuals are initialized:

$$\tilde{\alpha}_{omg} = \alpha_{omg} + \mu_{o0g} - \mu_{o0c},$$

where c denotes the college group. This measures every group's human capital in the college group's units.

B.2 Details on Field Choice

B.2.1 Deriving the Recursive Representation

First we derive the transformation for the entry costs regression by expressing the entry value function in terms of future entry values and current payoffs.

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

Group observable or estimated terms—wages, costs, and transition probabilities—into $V_t^F(k,m,\mathbf{X})$, where dependence on the base occupation o is suppressed for convenience. Thus we have,

$$V_t^e(k,m) = V_t^F(k,m,\mathbf{X}) + \beta V^e_{t+1}(k,m).$$

Averaging over k we have,

$$V_t^e(m,\mathbf{X}) = \left[\sum_k \pi(k|m,\mathbf{X})V_t^F(k,m,\mathbf{X})\right] + \beta V^e_{t+1}(m,\mathbf{X})$$

Hence,

$$\begin{aligned}
\bar{U}_t(l,m,\mathbf{X}) =& \vartheta_{lm} + \theta_l V_t^e(m,\mathbf{X}) \\
=& (1-\beta)\vartheta_{lm} + \theta_l\left[\sum_k \pi(k|m,\mathbf{X})V_t^F(k,m,\mathbf{X})\right] + \beta\left[\vartheta_{lm} + \theta_l V^e_{t+1}(m,\mathbf{X})\right]
\end{aligned}$$

$$=(1-\beta)\vartheta_{lm}+\theta_l\left[\sum_k \pi(k|m,\mathbf{X})V_t^F(k,m,\mathbf{X})\right]+\beta\bar{U}_{t+1}(l,m,\mathbf{X}).$$

Iterating forward yields the expression for arbitrary t, and we fix $\bar{U}_T$ as the terminal value for all periods.

B.2.2 Likelihood Conditional on l

The FOCs are given by,

$$\frac{\partial LL}{\partial \vartheta_{lm}}=\sum_t(1-\beta^{T-6-t})\sum_i q_{itl}\sum_r\left[\mathbf{1}(m_{r_i}=m)-\frac{N_{irm}^{\upsilon}\exp(U_{ml,t}(\mathbf{X}_i))}{\sum_{m'\in\mathcal{C}_i}N_{irm'}^{\upsilon}\exp(U_{m'l,t}(\mathbf{X}_i))}\right] \tag{B.3}$$

$$\frac{\partial LL}{\partial \theta_l}=\sum_t\sum_i q_{itl}\sum_r\left[XV_{t,T-6}^F(m_{r_i},\mathbf{X}_i)-\frac{\sum_{m\in\mathcal{C}_i}XV_{t,T-6}^F(m,\mathbf{X}_i)\times N_{irm}^{\upsilon}\exp(U_{ml,t}(\mathbf{X}_i))}{\sum_{m\in\mathcal{C}_i}N_{irm}^{\upsilon}\exp(U_{ml,t}(\mathbf{X}_i))}\right] \tag{B.4}$$

$$\frac{\partial LL}{\partial \upsilon}=\sum_t\sum_l\sum_i q_{itl}\sum_r\left[\log N_{irm}-\frac{\sum_{m\in\mathcal{C}_i}\log N_{irm}\times N_{irm}^{\upsilon}\exp(U_{ml,t}(\mathbf{X}_i))}{\sum_{m\in\mathcal{C}_i}N_{irm}^{\upsilon}\exp(U_{ml,t}(\mathbf{X}_i))}\right]. \tag{B.5}$$

We iterate on these FOCs to solve for each parameter. The first FOC states that the fixed effects match mean choices, the second that observed and predicted mean NPVs agree, and the third that observed and predicted mean varieties agree.

B.3 Bootstrapping

We calculate standard errors using the Bayesian block bootstrap, drawing weights for different stages. For the first stage of the labor block, the EM algorithm, by arguments in Arcidiacono and Miller (2011), is $\sqrt{N}$ consistent, so long as conditional on year, individuals' remaining errors are independent. While Arcidiacono and Miller (2011) provide analytical covariance matrices, we use a bootstrap procedure so that in the second stage, where we bootstrap instead over year, we can capture first stage uncertainty. For each bootstrap iteration, we draw from a Dirichlet distribution of length N with $\alpha_j=1\forall j$. We multiply these weights by N and then re-estimate the first stage.

In the second stage, we allow for clustering by year since expectational errors may be correlated. To do so we block bootstrap in the time dimension. Specifically, note that we may write the regression as depending on three periods $(t-1,t,t+1)$. We draw weights from a $T-2$ Dirichlet for the parameters in this estimation. We employ the same Bayesian bootstrapping procedure (with newly drawn weights) for the Field Choice estimation.

B.4 Approximating V^e

Given the Gumbel shocks, we derive above the following expression:

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

where λ are choice probabilities, and o is any choice of base occupation. Approximating in a stationary environment where $V_t^e\approx V_{t+1}^e$:

$$V_t^e(k,m)\approx\frac{-(1-\beta)C_{okm}^e+w_{ot}H(o,k,m)-\nu\log\lambda_t^e(o|k,m)-\beta\nu\log\frac{\lambda_{t+1}(o|o,k,m)}{\lambda_{t+1}^e(o|k,m)}}{1-\beta}.$$

The issue with the approximation is that we cannot estimate C^e_{okm} in levels. We assume that for retail, $C^e_{okm} \approx .25$, which is one quarter of the average annual income in the economy (which has been normalized to 1). Given this normalization, we can calculate the above for every value of o and every period t in the economy, and take the average. And so, calling the right hand side, $\tilde{V}^e_{ot}(k,m)$, where the $\tilde{V}$ refers to the approximated value evaluated at o, we define,

$$V^{e,approx}(k,m) = \frac{1}{|\mathcal{O}| \times T} \sum_o \sum_t \tilde{V}^e_{ot}(k,m).$$

We then have,

$$V^{e,approx}(m,\mathbf{X}) = \sum_k \pi(k|m,\mathbf{X}) V^{e,approx}(k,m).$$

B.5 Return Elasticities

We are interested in the returns to an additional GPA point across majors in order to see if there is a complementarity to overall ability and labor market returns. To this end we wish to calculate, $dV_m/dGPA$ for each major. Ignoring general equilibrium feedback due to resorting, which is second order,

$$\begin{aligned}\frac{d}{dGPA} V_m =& \frac{d}{dGPA} \sum_k V(m,k)\pi(k|m,GPA) \\ =& \sum_k V(m,k) \frac{d\pi(k|m,GPA)}{dGPA}.\end{aligned}$$

In our specification we use a flexible bin with 20 points for GPA. So this derivative is actually discrete whenever a student crosses the threshold to a new bin point. Two issues arise in calculating this object empirically: returns are time-varying, and there are other covariates. For the first we simply use the time average of $V_{m,k,t}$. For the second we calculate the elasticity for each field for all students and take the population-weighted average:

$$\frac{d}{dGPA} V_m = \frac{1}{N} \sum_i \frac{d}{dGPA} V_{mi},$$

where $\frac{d}{dGPA} V_{mi}$ is the return to a field for each person i. If a type is precluded from a field, we remove them from the calculation. As such, this is the mean across all students with the choice set available. Hence, we ignore the effect of GPA on choice set itself.

B.6 Sorting of Latent Types Across Occupations

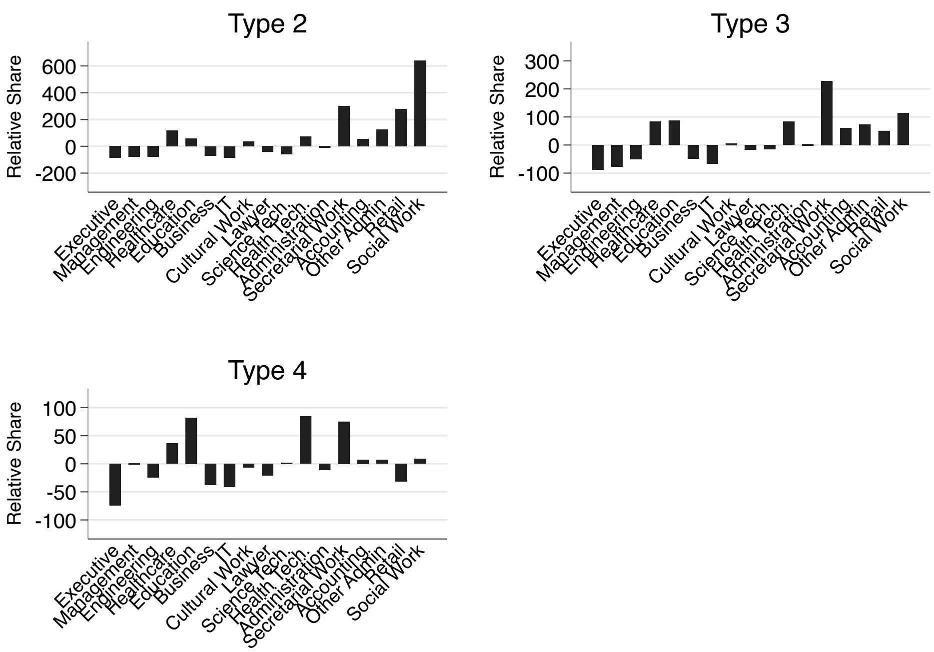
We provide more details to the discussion in Section 5.3.2. Our estimation recovers four latent worker types k that differ in productivity across occupations through the parameters α_{ok}. Because the types are latent, this appendix documents how they sort across occupations in the data, and whether that sorting aligns with the comparative advantage the model recovers.

We measure sorting relative to a base type (Type 1), since the level of each type's occupational distribution is not separately identified. For each type k and occupation o, the sorting index

$$\text{SortingIndex}_{ko} = \left(\frac{\sum_{i\in o} q_{ik} / \sum_i q_{ik}}{\sum_{i\in o} q_{i1} / \sum_i q_{i1}} - 1 \right) \times 100$$

is positive when type-k workers are over-represented in occupation o relative to the base type. Figure B.1 plots it for every occupation with at least a 0.5% population share, and the types sort clearly across them: Type 1 workers concentrate in management and Type 4 workers are absent from office and service work.

Figure B.1
Occupational Sorting Patterns by Type



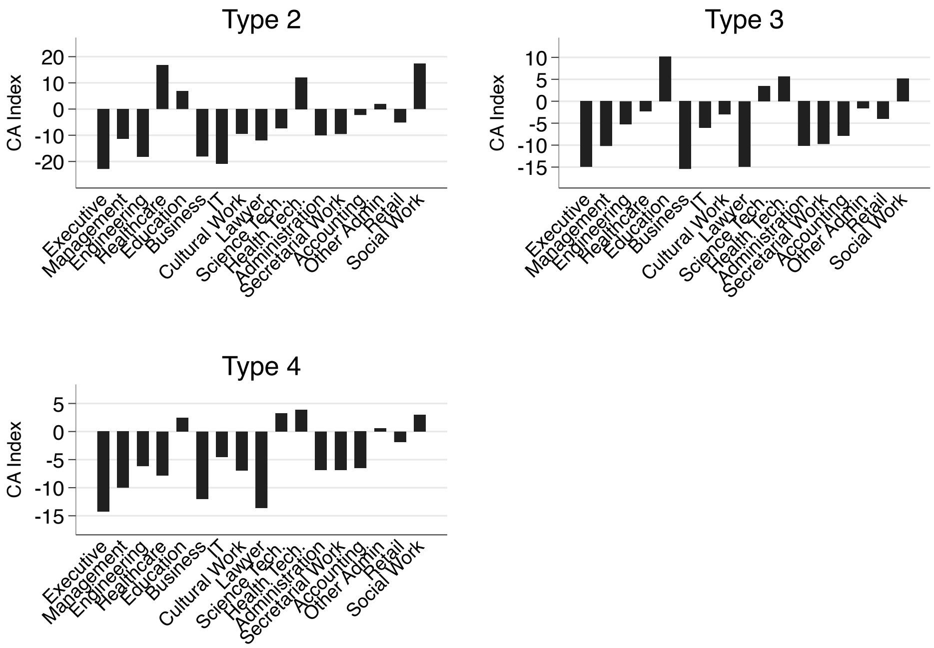
Notes: Each panel plots the data sorting index SortingIndex_{ko} (in %) by occupation, for latent worker types $k = 2, 3, 4$ relative to the base Type 1 (zero by construction, and omitted). A positive bar means type k is over-represented in that occupation relative to Type 1; a negative bar, under-represented. The index is built from the estimated posterior type probabilities q_{ik}, as defined above, and is shown for the occupations with at least a 0.5% population share. This figure is discussed in Section 5.3.2.

To ask whether this sorting reflects comparative advantage, we construct a second index from the estimated productivities alone, again normalized to Type 1:

$$\text{CAIndex}_{ko} = \frac{\exp(\alpha_{ok}) / \sum_{o'} \exp(\alpha_{o'k})}{\exp(\alpha_{o1}) / \sum_{o'} \exp(\alpha_{o'1})}.$$

Figure B.2 shows that this index tracks the sorting observed in the data. Type 2 workers, for example, are under-represented in management and over-represented in social work on both measures. The estimated productivities account for only part of the magnitude of sorting, with the rest reflecting preferences and wages, but the patterns line up across the two. The sorting in the data therefore follows the comparative advantage the model recovers.

Figure B.2
Comparative Advantage and Sorting by Type



Notes: Each panel plots the comparative-advantage index CAIndex_{ko} by occupation, for types $k = 2, 3, 4$ relative to the base Type 1, built from the estimated occupational productivities α_{ok} alone. Occupations and panels match Figure B.1. The sign pattern lines up with the data sorting there: where a type is over-represented in an occupation, the model also gives it a comparative advantage in it. This figure is discussed in Section 5.3.2.

B.7 Complete Parameter Estimates

Table B.1
α_k

Task	Type			
	1	2	3	4
0	0.000	-0.773	-0.251	0.132
	(0.000)	(0.020)	(0.014)	(0.010)
1	0.000	-0.040	-0.030	-0.037
	(0.000)	(0.002)	(0.001)	(0.001)
2	0.000	-0.026	-0.008	-0.006
	(0.000)	(0.002)	(0.000)	(0.000)
3	0.000	0.081	0.044	0.031
	(0.000)	(0.002)	(0.001)	(0.001)
4	0.000	0.023	-0.035	-0.017
	(0.000)	(0.002)	(0.001)	(0.001)

Notes: This table displays the absolute advantage (task 0) and comparative advantage parameters (tasks 1-4) of different types. The first type is normalized in all parameters to 0.

Table B.2

$$\alpha_{om}$$

Occupation								Field							
	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15
1	0.000	-0.130	0.004	0.226	0.018	0.268	0.145	0.138	0.211	0.070	0.257	0.115	-0.013	-0.182	0.243
	(0.000)	(0.021)	(0.014)	(0.010)	(0.011)	(0.008)	(0.011)	(0.009)	(0.014)	(0.016)	(0.007)	(0.011)	(0.013)	(0.011)	(0.013)
2	0.000	-0.303	-0.113	0.124	-0.114	0.131	0.106	-0.008	0.076	-0.023	0.130	-0.002	-0.284	-0.207	0.141
	(0.000)	(0.020)	(0.011)	(0.010)	(0.012)	(0.007)	(0.011)	(0.014)	(0.015)	(0.024)	(0.005)	(0.008)	(0.012)	(0.004)	(0.009)
3	0.000	-0.067	-0.000	0.192	0.051	0.248	0.147	0.155	0.241	0.125	0.280	0.157	0.044	-0.011	0.201
	(0.000)	(0.017)	(0.017)	(0.018)	(0.019)	(0.016)	(0.026)	(0.016)	(0.017)	(0.018)	(0.016)	(0.015)	(0.021)	(0.032)	(0.022)
4	0.000	0.082	0.061	0.258	0.112	0.260	0.159	0.332	0.364	0.209	0.434	0.216	0.102	-0.265	0.458
	(0.000)	(0.049)	(0.029)	(0.030)	(0.031)	(0.030)	(0.039)	(0.023)	(0.024)	(0.064)	(0.024)	(0.028)	(0.022)	(0.035)	(0.021)
5	0.000	-0.364	-0.108	-0.058	-0.118	-0.030	-0.057	-0.039	-0.020	-0.080	-0.010	-0.060	-0.223	-0.235	-0.036
	(0.000)	(0.006)	(0.004)	(0.005)	(0.004)	(0.005)	(0.011)	(0.004)	(0.004)	(0.012)	(0.006)	(0.007)	(0.007)	(0.002)	(0.005)
6	0.000	-0.162	-0.084	0.106	-0.073	0.154	0.113	0.044	0.094	-0.075	0.126	0.015	-0.026	-0.174	0.163
	(0.000)	(0.012)	(0.011)	(0.007)	(0.009)	(0.006)	(0.009)	(0.010)	(0.011)	(0.024)	(0.009)	(0.012)	(0.011)	(0.014)	(0.017)
7	0.000	-0.099	0.040	0.156	0.018	0.214	0.147	0.078	0.120	0.087	0.223	0.103	-0.008	-0.124	0.098
	(0.000)	(0.021)	(0.014)	(0.014)	(0.013)	(0.013)	(0.018)	(0.014)	(0.016)	(0.013)	(0.012)	(0.015)	(0.017)	(0.027)	(0.023)
8	0.000	-0.137	0.019	0.215	0.096	0.202	0.234	0.151	0.185	0.004	0.201	0.152	0.052	-0.081	0.091
	(0.000)	(0.014)	(0.014)	(0.011)	(0.012)	(0.014)	(0.014)	(0.015)	(0.015)	(0.024)	(0.019)	(0.022)	(0.022)	(0.012)	(0.019)
9	0.000	0.012	-0.017	0.190	0.029	0.175	0.236	0.022	0.120	-0.039	0.149	0.007	0.038	-0.023	0.072
	(0.000)	(0.058)	(0.032)	(0.028)	(0.028)	(0.030)	(0.029)	(0.031)	(0.031)	(0.036)	(0.031)	(0.029)	(0.051)	(0.037)	(0.035)
10	0.000	0.003	-0.029	0.177	0.026	0.221	0.192	0.145	0.187	0.145	0.261	0.140	-0.142	-0.065	0.136
	(0.000)	(0.020)	(0.025)	(0.032)	(0.020)	(0.017)	(0.040)	(0.019)	(0.022)	(0.026)	(0.017)	(0.018)	(0.017)	(0.027)	(0.047)
11	0.000	-0.069	0.112	0.224	0.155	0.236	0.262	0.231	0.236	0.279	0.343	0.225	-0.039	-0.053	0.261
	(0.000)	(0.061)	(0.049)	(0.047)	(0.054)	(0.042)	(0.046)	(0.043)	(0.050)	(0.071)	(0.044)	(0.050)	(0.043)	(0.062)	(0.052)
12	0.000	-0.074	0.044	0.258	0.067	0.279	0.245	0.124	0.221	0.102	0.280	0.157	0.069	-0.108	0.131
	(0.000)	(0.013)	(0.009)	(0.010)	(0.009)	(0.009)	(0.012)	(0.013)	(0.017)	(0.016)	(0.010)	(0.010)	(0.010)	(0.009)	(0.023)
13	0.000	-0.091	0.049	0.201	0.058	0.250	0.148	0.086	0.130	0.100	0.238	0.151	0.045	-0.028	0.097
	(0.000)	(0.020)	(0.024)	(0.021)	(0.018)	(0.018)	(0.048)	(0.025)	(0.027)	(0.026)	(0.018)	(0.025)	(0.027)	(0.032)	(0.039)
14	0.000	-0.025	0.070	0.212	0.162	0.316	0.274	0.144	0.177	0.162	0.289	0.193	0.073	-0.099	0.138
	(0.000)	(0.022)	(0.013)	(0.016)	(0.012)	(0.012)	(0.018)	(0.018)	(0.022)	(0.024)	(0.015)	(0.017)	(0.014)	(0.014)	(0.027)
15	0.000	-0.006	0.075	0.114	0.091	0.291	0.195	0.070	0.024	0.086	0.225	0.211	0.063	-0.055	0.066
	(0.000)	(0.042)	(0.027)	(0.034)	(0.026)	(0.025)	(0.046)	(0.035)	(0.077)	(0.068)	(0.032)	(0.043)	(0.028)	(0.036)	(0.053)
16	0.000	-0.006	0.101	0.193	0.066	0.267	0.283	0.126	0.167	0.132	0.231	0.101	0.107	-0.120	0.003
	(0.000)	(0.035)	(0.035)	(0.028)	(0.020)	(0.022)	(0.026)	(0.032)	(0.033)	(0.055)	(0.023)	(0.030)	(0.023)	(0.032)	(0.062)
17	0.000	-0.078	-0.018	0.134	0.105	0.233	0.179	0.057	0.065	0.089	0.265	0.167	-0.020	-0.042	0.049
	(0.000)	(0.026)	(0.025)	(0.024)	(0.019)	(0.018)	(0.030)	(0.033)	(0.041)	(0.043)	(0.019)	(0.022)	(0.021)	(0.022)	(0.043)
18	0.000	0.030	-0.011	0.132	0.119	0.263	0.166	0.069	0.110	0.157	0.227	0.345	0.165	-0.027	0.139
	(0.000)	(0.032)	(0.039)	(0.047)	(0.024)	(0.022)	(0.039)	(0.070)	(0.064)	(0.055)	(0.030)	(0.026)	(0.024)	(0.022)	(0.047)
19	0.000	-0.052	-0.003	0.228	0.161	0.388	0.192	0.217	0.241	0.264	0.393	0.241	0.177	-0.216	0.237
	(0.000)	(0.027)	(0.028)	(0.031)	(0.023)	(0.020)	(0.047)	(0.027)	(0.036)	(0.038)	(0.024)	(0.027)	(0.021)	(0.026)	(0.046)
20	0.000	-0.036	-0.079	-0.040	0.008	0.021	-0.036	-0.056	0.022	-0.078	0.022	0.058	0.061	0.007	0.060
	(0.000)	(0.016)	(0.019)	(0.021)	(0.022)	(0.017)	(0.043)	(0.033)	(0.045)	(0.062)	(0.030)	(0.028)	(0.015)	(0.013)	(0.029)
21	0.000	-0.257	-0.167	0.012	-0.014	0.113	0.078	0.056	0.067	0.054	0.036	0.153	0.051	-0.050	-0.007
	(0.000)	(0.046)	(0.053)	(0.053)	(0.042)	(0.033)	(0.060)	(0.047)	(0.065)	(0.101)	(0.047)	(0.043)	(0.040)	(0.036)	(0.049)
22	0.000	-0.049	0.217	0.278	0.017	0.170	0.267	0.282	0.264	0.167	0.304	0.308	-0.022	-0.010	0.350
	(0.000)	(0.063)	(0.055)	(0.052)	(0.043)	(0.051)	(0.119)	(0.041)	(0.065)	(0.123)	(0.038)	(0.038)	(0.040)	(0.042)	(0.049)
23	0.000	0.095	0.115	0.109	0.110	0.178	0.192	0.226	0.290	0.288	0.386	0.259	0.180	0.031	0.494
	(0.000)	(0.109)	(0.108)	(0.102)	(0.097)	(0.067)	(0.236)	(0.111)	(0.097)	(0.117)	(0.058)	(0.064)	(0.095)	(0.073)	(0.108)
24	0.000	0.059	0.149	0.147	0.244	0.114	0.338	0.015	0.025	0.177	0.178	0.089	0.144	-0.005	0.222
	(0.000)	(0.094)	(0.095)	(0.094)	(0.090)	(0.090)	(0.080)	(0.095)	(0.188)	(0.108)	(0.079)	(0.098)	(0.106)	(0.088)	(0.093)
25	0.000	0.083	0.120	0.062	0.111	0.179	0.160	0.216	0.224	0.042	0.347	0.167	0.139	-0.123	0.236
	(0.000)	(0.024)	(0.037)	(0.043)	(0.034)	(0.027)	(0.051)	(0.029)	(0.037)	(0.051)	(0.020)	(0.020)	(0.025)	(0.027)	(0.073)
26	0.000	0.111	0.148	0.129	0.172	0.162	0.124	0.086	0.193	0.022	0.362	0.213	0.113	0.024	0.173
	(0.000)	(0.038)	(0.044)	(0.050)	(0.032)	(0.026)	(0.056)	(0.044)	(0.045)	(0.090)	(0.023)	(0.030)	(0.032)	(0.032)	(0.065)
27	0.000	0.012	0.023	0.070	0.058	0.193	0.101	0.152	0.072	0.181	0.243	0.170	0.156	-0.026	0.096
	(0.000)	(0.042)	(0.041)	(0.039)	(0.035)	(0.027)	(0.053)	(0.045)	(0.066)	(0.071)	(0.027)	(0.028)	(0.040)	(0.030)	(0.076)
28	0.000	0.091	0.166	0.154	0.185	0.367	0.297	0.287	0.330	0.222	0.379	0.427	0.209	-0.054	0.232
	(0.000)	(0.032)	(0.032)	(0.042)	(0.028)	(0.028)	(0.056)	(0.042)	(0.053)	(0.073)	(0.028)	(0.026)	(0.032)	(0.025)	(0.072)
29	0.000	0.073	0.091	0.238	0.207	0.203	-0.070	0.108	0.135	0.096	0.289	0.374	0.068	-0.035	0.113
	(0.000)	(0.062)	(0.045)	(0.078)	(0.034)	(0.042)	(0.105)	(0.072)	(0.085)	(0.127)	(0.037)	(0.033)	(0.038)	(0.041)	(0.081)

Table B.3

ϑ_{lm}

Field	Type			
	1	2	3	4
1	-0.500	0.073	0.908	-1.522
	(0.029)	(0.054)	(0.030)	(0.023)
2	-2.773	-2.463	1.340	0.988
	(0.020)	(0.009)	(0.018)	(0.036)
3	-4.536	-1.298	-0.243	0.904
	(0.034)	(0.034)	(0.019)	(0.026)
4	-3.564	0.116	0.263	1.531
	(0.037)	(0.040)	(0.025)	(0.026)
5	-5.163	-1.493	1.892	-0.717
	(0.028)	(0.040)	(0.035)	(0.030)
6	-6.318	-2.335	2.448	-0.087
	(0.020)	(0.017)	(0.040)	(0.023)
7	-4.190	0.591	4.852	3.421
	(0.020)	(0.057)	(0.043)	(0.042)
8	-4.086	1.984	4.779	0.734
	(0.017)	(0.109)	(0.061)	(0.033)
9	-4.440	-0.814	3.761	0.737
	(0.031)	(0.132)	(0.081)	(0.039)
10	-6.987	-3.427	4.167	0.936
	(0.004)	(0.005)	(0.031)	(0.035)
11	-4.596	-1.994	3.603	-0.105
	(0.010)	(0.027)	(0.036)	(0.026)
12	-4.162	-1.804	3.148	0.391
	(0.026)	(0.014)	(0.040)	(0.025)
13	-5.225	3.488	-0.894	-3.810
	(0.053)	(0.112)	(0.016)	(0.023)
14	-1.171	0.491	5.091	-3.424
	(0.034)	(0.079)	(0.021)	(0.029)
15	-4.827	3.711	3.178	-0.942
	(0.037)	(0.100)	(0.042)	(0.027)

Notes: Displays the type-field preference fixed effect.

Table B.4

Additional Parameters

Parameter	Estimate
υ	1.363
	(0.008)

Appendix C: Counterfactual Appendix

This appendix gives the demand-side model and its calibration, the algorithm for the counterfactual transitions, and additional results from the experiments.

C.1 Labor Demand Model Details

We provide more details to the labor demand model discussed in Section 6.1. 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:

$$H^P_{sot} = A_{st} \times \frac{\beta^P_{so}}{W_{ot}} \times A^P_s \times Q_{st} \times \left(P^D_{st}\right)^{\alpha} \times \left(P^P_{st}\right)^{\rho-\alpha} \times \left(W^P_{st}\right)^{-\rho}, \quad \forall o \in \mathcal{O}^P; \tag{C.1}$$

$$H^N_{sot} = A_{st} \times \frac{\beta^N_{so}}{W_{ot}} \times A^N_s \times Q_{st} \times \left(P^D_{st}\right)^{\alpha} \times \left(W^N_{st}\right)^{-\alpha}, \quad \forall o \in \mathcal{O}^N; \tag{C.2}$$

and demand for imported goods is

$$Q^M_{st} = A_{st} \times A^M_s \times Q_{st} \times \left(P^D_{st}\right)^{\alpha} \times \left(P^P_{st}\right)^{\rho-\alpha} \times \left(P^M_{st}\right)^{-\rho}, \tag{C.3}$$

where

$$Q_{st} = A^F_{st}\left(P^D_{st}\right)^{-\eta} + \mu_s E_t \times \left(\frac{P_{st}}{P_t}\right)^{1-\sigma} \times \frac{\left(P^D_{st}\right)^{-\eta}}{\left(P^D_{st}\right)^{1-\eta} + \left(P^F_{st}\right)^{1-\eta}}; \tag{C.4}$$

$$W^f_{st} = \prod_{o \in \mathcal{O}^f} w^{\beta_{so}}_{ot}, \quad f = \{P, N\}; \tag{C.5}$$

$$P^P_{st} = \left[A^P_s\left(W^P_{st}\right)^{1-\rho} + A^M_s\left(P^M_{st}\right)^{1-\rho}\right]^{\frac{1}{1-\rho}}; \tag{C.6}$$

$$P^D_{st} = \left[A^N_s\left(W^N_{st}\right)^{1-\alpha} + \left(P^P_{st}\right)^{1-\alpha}\right]^{\frac{1}{1-\alpha}} \tag{C.7}$$

$$P_{st} = A^{-1}_{st}\left[\left(P^D_{st}\right)^{1-\eta} + \left(P^F_{st}\right)^{1-\eta}\right]^{\frac{1}{1-\eta}} \tag{C.8}$$

$$P_t = \left[\sum_s \mu_s \left(P_{st}\right)^{1-\sigma}\right]^{\frac{1}{1-\sigma}} \tag{C.9}$$

$$E_t = \sum_o w_{ot}\left(\sum_s L_{sot}\right) \tag{C.10}$$

Finally, occupation wages, w_{ot}, adjust to clear the labor market, for $k = \{P, N\}$

$$H_{ot} = \sum_s H^k_{sot} \tag{C.11}$$

C.2 Demand Side Parameters Calibration

Table C.1 summarizes the full set of parameters to be calibrated to close the model. All production coefficients are estimated as the time-averaged expenditure shares.

Table C.1
Labor Demand Parameters and Variables

Parameters	Meaning	Value/Data Source
β_{so}^{P}, β_{so}^{N}	Labor inputs coefficients	IDA
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	Backus et al. (1994)

Notes: These parameters are discussed in Section 6.1.

Total wage bills in occupation-industry pairs identify the labor input coefficients. We obtain raw wage data from the Integrated Database for Labor Market Research (IDA) at Statistics Denmark. Occupation wages are computed by first running a mincerian regression and then dividing the residuals by the computed human capital using (2). We set elasticity across sectors σ at 0.2, following Atalay (2017) and the trade elasticity η at 1.5, a standard value for the elasticity of substitution between domestic and foreign varieties in the literature (Backus et al., 1994; Ruhl, 2008).

C.3 Construction of Labor Market Shocks

This appendix describes how we construct the trade war and AI shocks.

C.3.1 Trade War Shock

We model the trade war as a 30% increase in the price of imported goods (P_{st}^{F}) and offshored inputs (P_{st}^{M}), along with a 30% reduction in foreign demand for Danish exports (A_{st}^{F}). The three components capture higher costs of imported consumption goods, higher prices of offshored inputs, and reduced foreign demand for domestic exports. We choose a uniform 30%, a large but plausible disruption, rather than calibrate it to a particular episode, so the experiment identifies how the model adjusts to such a shock rather than the quantitative effect of any actual trade war.

Figure C.1 reports the trade war's *ex-ante* exposure—a measure of each occupation's and field's direct reliance on foreign markets, before any equilibrium adjustment. Panel (a) shows occupation exposure, computed as a trade-weighted average of each occupation's sectors' net foreign exposure; Panel (b) aggregates this to fields, weighting by the occupations each field's graduates enter. Production-related occupations and tradable services are the most exposed, and Engineering, Law, Architecture, and Business Admin are the most exposed fields.

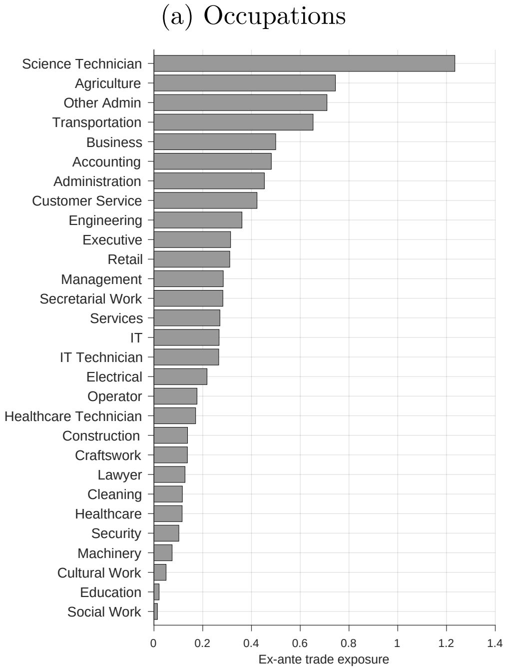
C.3.2 AI Shock

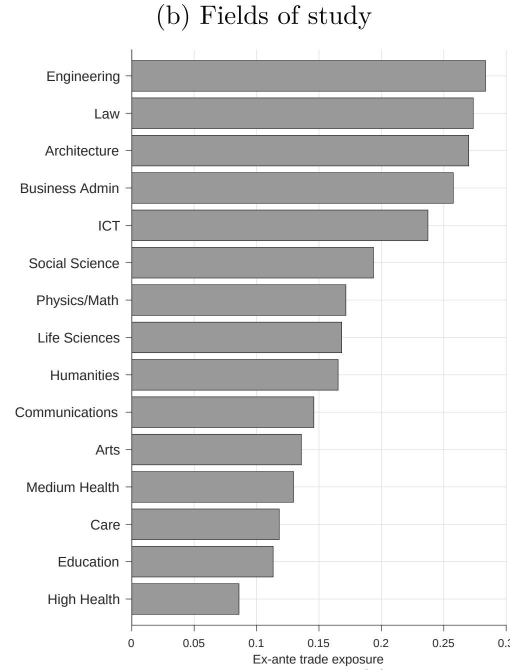
We model AI as a shock to the productivity of skills rather than to the demand for occupations. It scales the human capital a field provides, and because that scaling carries into every occupation the field's graduates enter, a worker cannot offset it by changing occupation. In general equilibrium it still shifts the demand for human capital across occupations, but its incidence is anchored to the field a worker studied, not the occupation she holds.

Concretely, the shock scales each field's human capital by a field-specific factor ψ_m, common across occupations and worker types:

$$H'(o,m,k) = \psi_m \, H(o,m,k) \qquad \text{for all } o, k,$$

Figure C.1
Ex-Ante Exposure to the Trade War Shock





Notes: Ex-ante exposure to the trade war, before any equilibrium adjustment, in the 2018 economy. Panel (a): occupation exposure $Exposure_o = \sum_i \left(\frac{X_i - M_i - O_i}{WB_i}\right) \frac{WB_{oi}}{WB_o}$, where X_i, M_i, and O_i are the export value, imported-goods value, and offshored-input value of sector i, and WB_i, WB_o, WB_{oi} are the wage bills at the sector, occupation, and sector-occupation level. Panel (b): field exposure $\sum_o Exposure_o \, \pi^e_m(o)$, where $\pi^e_m(o)$ is the entry-occupation distribution of field m's graduates. This figure is discussed in Section 6.2.1.

which is equivalent to shifting the estimated field–occupation productivities by a field-specific constant, $\alpha'_{om} = \alpha_{om} + \log \psi_m$. We calibrate the multipliers $\{\psi_m\}$ in three steps.

Step 1: field exposure. We measure a field's exposure to AI by aggregating occupation-level exposure over the occupations its graduates enter. Let $z_o \in [0, 1]$ denote the AI exposure of occupation o from Eloundou et al. (2023), and let $\pi^e_m(o)$ be the share of field m's new graduates whose first occupation is o, implied by the model's entry probabilities $\lambda^e(o \mid k, m)$ and the type composition of the field. Field exposure is

$$E_m = \sum_o \pi^e_m(o)\, z_o.$$

We use the entry-occupation distribution rather than realized lifetime employment, since it directly captures the skills a degree trains for.

Step 2: a two-sided multiplier. Let $e_m = (E_m - \bar{E})/\sigma_E$ be field exposure standardized across the fields. AI substitutes for the skills of above-average-exposure fields and complements those of below-average-exposure fields:

$$\psi_m = \max\big\{\, 1 - \kappa \max(e_m, 0) + \kappa \max(-e_m, 0),\ \ \underline{\psi} \,\big\},$$

with $\kappa = 0.30$ and a floor $\underline{\psi} = 0.05$. The complementarity term—the gain to low-exposure fields—captures the rising relative return to non-automatable interpersonal, manual, and creative skills as cognitive tasks are automated, in the task-based tradition of Acemoglu and Autor (2011) and Deming and Noray (2020).

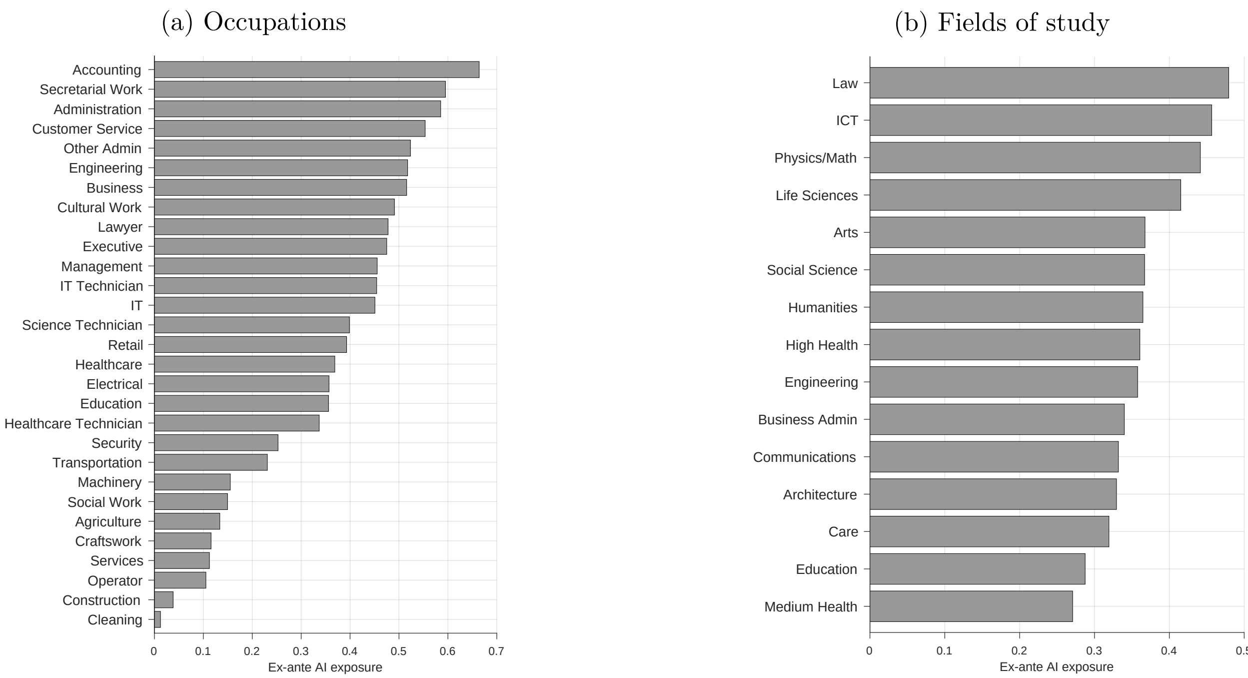
Step 3: net-neutral normalization. We rescale the multipliers so that aggregate human capital at the initial field composition is unchanged:

$$\tilde{\psi}_m = \psi_m \Big/ \exp\Big(\sum_{m'} s_{m'} \log \psi_{m'}\Big),$$

where s_m is field m's share of college employment. This holds the aggregate skill endowment fixed, isolating the reallocation consequences of AI—which fields gain and lose relative to one another—from any assumption about its effect on the aggregate level of productivity, on which the AI literature offers little consensus.

Figure C.2 reports the AI shock's ex-ante exposure across occupations (Panel a) and fields (Panel b). The occupation exposure measure is taken directly from Eloundou et al. (2023); field exposure aggregates it over each field's entry-occupation distribution. The most exposed occupations are cognitive and analytical roles, and the most exposed fields are Law, ICT, Physics/Math, and Life Sciences.

Figure C.2
Ex-Ante Exposure to the AI Shock



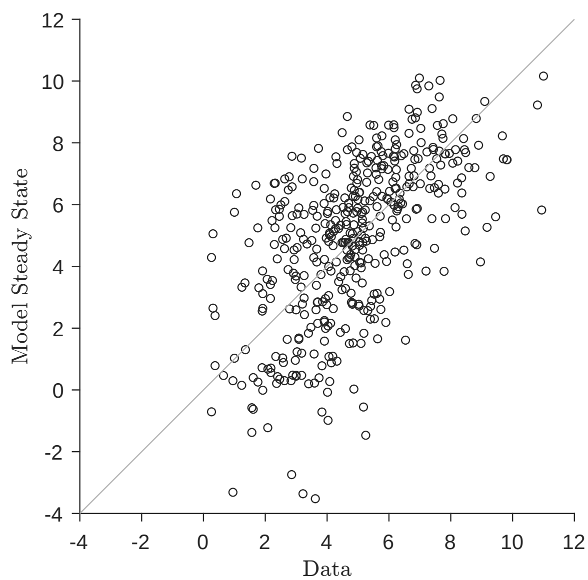
Notes: Ex-ante exposure to the AI shock, before any equilibrium adjustment. Panel (a): occupation-level AI exposure from Eloundou et al. (2023). Panel (b): field exposure, the same measure aggregated over the entry-occupation distribution of each field's graduates. This figure is discussed in Section 6.2.1.

C.4 Modeling Admissions and Fit

Because programs within a field are symmetric, there is no program-level sorting for a deferred-acceptance algorithm to reproduce, so we do not run the algorithm in the counterfactuals. For each field m we instead compute the ex-ante probability that a student with score s and track h is admitted to a program in field m, denoted $P(m \mid h, s)$. When a student lists field m she is admitted with this probability; if she is not, she draws her next application from the field-choice probabilities implied by (9), so she may apply to the same field more than once, as in the data. This lets the admission constraints operate on aggregated fields while preserving the score-dependent rationing the algorithm would impose.

To assess how well this procedure, and more generally our model of occupational sorting, reproduces the data, Figure C.3 plots the log data share of each field–occupation cell (m, o) against the model share in our simulated steady state. Because the model is initialized from the data shares, the figure asks whether the simulated economy stays near the data as it converges to its steady state. The correlation across cells is 0.59. A regression of the log model share on the log data share has a slope somewhat below one, so the model shares are a little flatter than the data. Model and data shares diverge most for the smallest cells, whose data shares are themselves measured from few workers.

Figure C.3
Data versus Model-simulated field–occupation shares



Notes: Each point is a field–occupation cell (m, o), plotting the log model share in the simulated steady state (vertical axis) against the log data share (horizontal axis). This figure is discussed in Section 6.2.2.

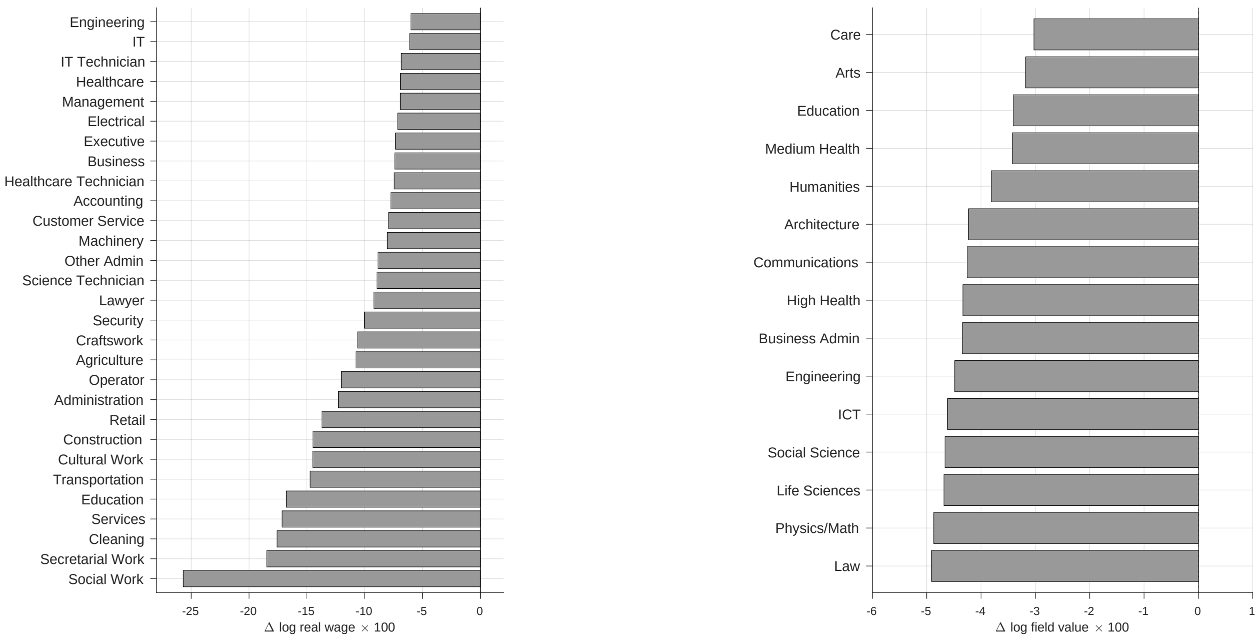
C.5 Additional Results on the Two Disruptions

Table C.2 reports the dispersion of each shock across occupations and fields, and Figures C.4 and C.5 plot the underlying distributions, across occupations (changes in occupation wage w_o) and fields (changes in the value V_m^e). The trade war moves occupation wages widely but leaves field values nearly common; the AI shock changes occupation wages little but disperses field values. Figure C.6 reports field reallocation under each disruption.

C.6 Alternative Education Reforms

Figure C.4
Equilibrium Incidence of the Trade War

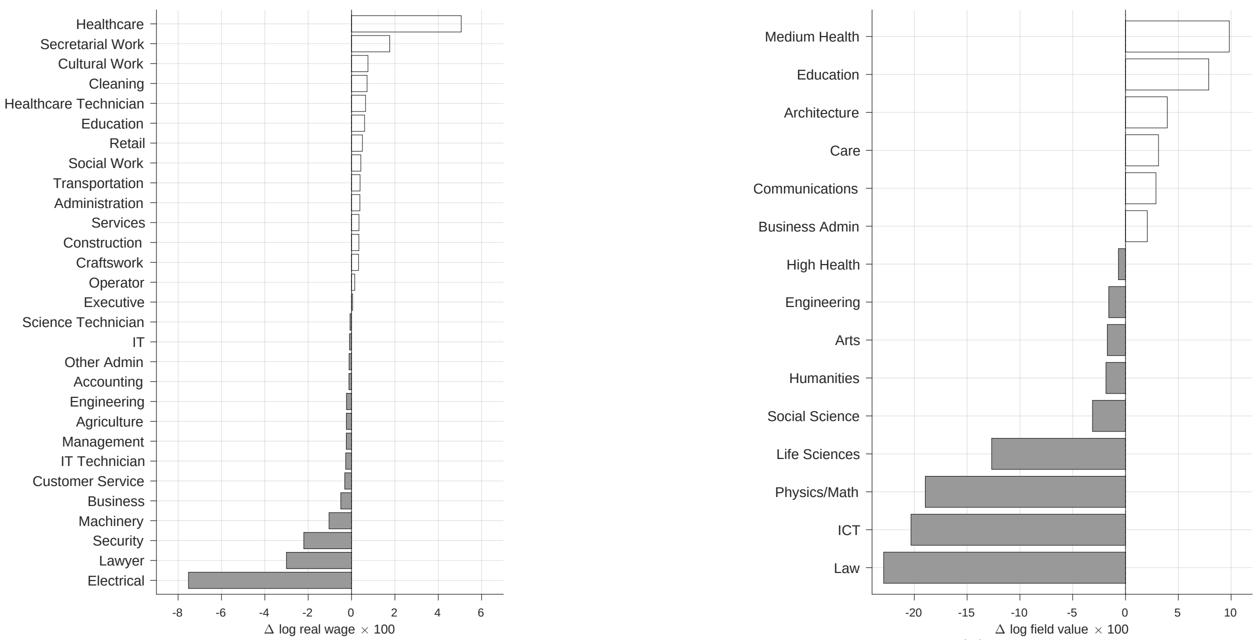
(a) Occupations (b) Fields of study



Notes: Post-shock steady state relative to the pre-shock steady state, Current System. Panel (a): change in log occupation wages w_o ($\times 100$), by occupation. Panel (b): change in log field values V_m^e ($\times 100$), by field. This figure is discussed in Section 6.3.1.

Figure C.5
Equilibrium Incidence of the AI Shock

(a) Occupations (b) Fields of study



Notes: Post-shock steady state relative to the pre-shock steady state, Current System. Panel (a): change in log occupation wages w_o ($\times 100$), by occupation. Panel (b): change in log field values V_m^e ($\times 100$), by field. This figure is discussed in Section 6.3.1.

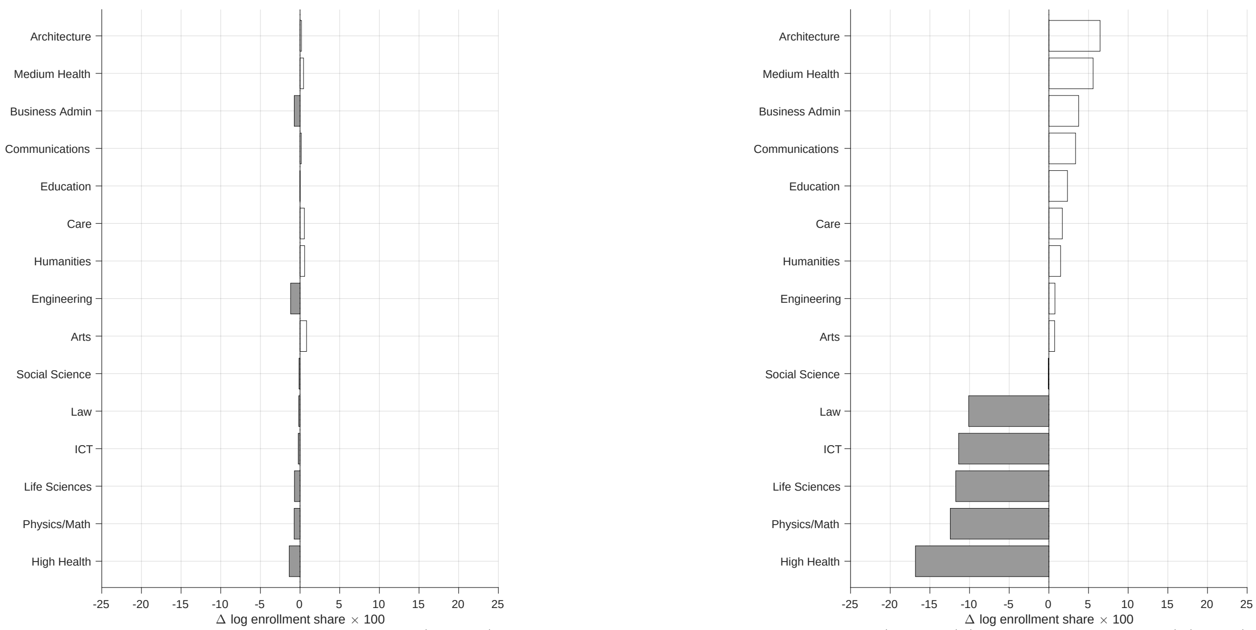
Table C.2
Dispersion of the Two Disruptions Across Occupations and Fields

	Trade War		AI	
	mean	s.d.	mean	s.d.
Δ log occupation wage (w_o) $\times 100$	-16.2	6.0	$+0.4$	1.1
Δ log field value (V_m^e) $\times 100$	-3.8	0.7	-0.0	8.9

Notes: Post-shock steady state relative to the pre-shock steady state, Current System. Occupation statistics are weighted by base employment; field statistics are weighted by enrollment. This table is discussed in Section 6.3.1.

Figure C.6
Field Reallocation Under the Current System

(a) Trade war (b) AI



Notes: Change in log enrollment shares ($\times$ 100) of entering cohorts under each shock (panel (a): trade war; panel (b): AI), Current System steady state relative to the pre-shock steady state. Fields are ordered by the AI response, on a common scale across panels. Under trade war, 0.5% of students change fields; under AI 3.3% do. This figure is discussed in Section 6.3.1.

The *Flexible Response* of the main text removes both restrictions at once: the academic tracking that ties applicants to the fields of their secondary-school program, and the capacity caps that ration entry into oversubscribed fields. These are separate instruments, and a planner could use either alone. Table C.3 reports each, together with the full reform and a variant that holds the High Health field (medicine, dentistry, and veterinary medicine) to the Current System. Outcomes are measured against the pre-shock economy, as in Table 7; and the value of a reform is its income gain over the Current System.

Almost all of the reform's value comes from removing the tracking, not the caps (Table C.3). The two restrictions bind differently: the tracking limits which fields a student may enter, so lifting it opens fields that were closed to her, while the caps only limit how many students each field admits. Removing the tracking alone is worth +4.89 pp under the trade war and +3.30 under AI. Removing the caps alone is worth +0.11 and −1.84. Under the trade war the caps do almost nothing, because field values move together and reallocating across fields gains little whether or not seats are rationed. Under AI removing the caps lowers welfare, because the caps ration the high-return fields the shock devalues, so lifting them moves students into fields whose value is falling. The full reform is therefore worth less under AI than removing the tracking alone, +2.28 against +3.30. Holding the High Health field to the Current System leaves the reform's value almost unchanged.

Table C.3
Alternative Education Reforms

	Trade War			AI		
	Income	GDP	Realloc.	Income	GDP	Realloc.
Panel A: post-shock levels						
Current System	−16.36	−14.37	0.5	+2.98	+1.15	3.3
Full reform	−11.63	−13.28	19.2	+5.26	+1.70	19.6
Remove tracking	−11.47	−13.28	18.5	+6.28	+1.92	19.1
Remove caps	−16.25	−14.16	8.3	+1.13	+0.82	9.1
Full, high-health fixed	−11.86	−13.37	18.1	+5.58	+1.79	19.0
Panel B: value over the Current System (income)						
Full reform		+4.73			+2.28	
Remove tracking		+4.89			+3.30	
Remove caps		+0.11			−1.84	
Full, high-health fixed		+4.51			+2.61	

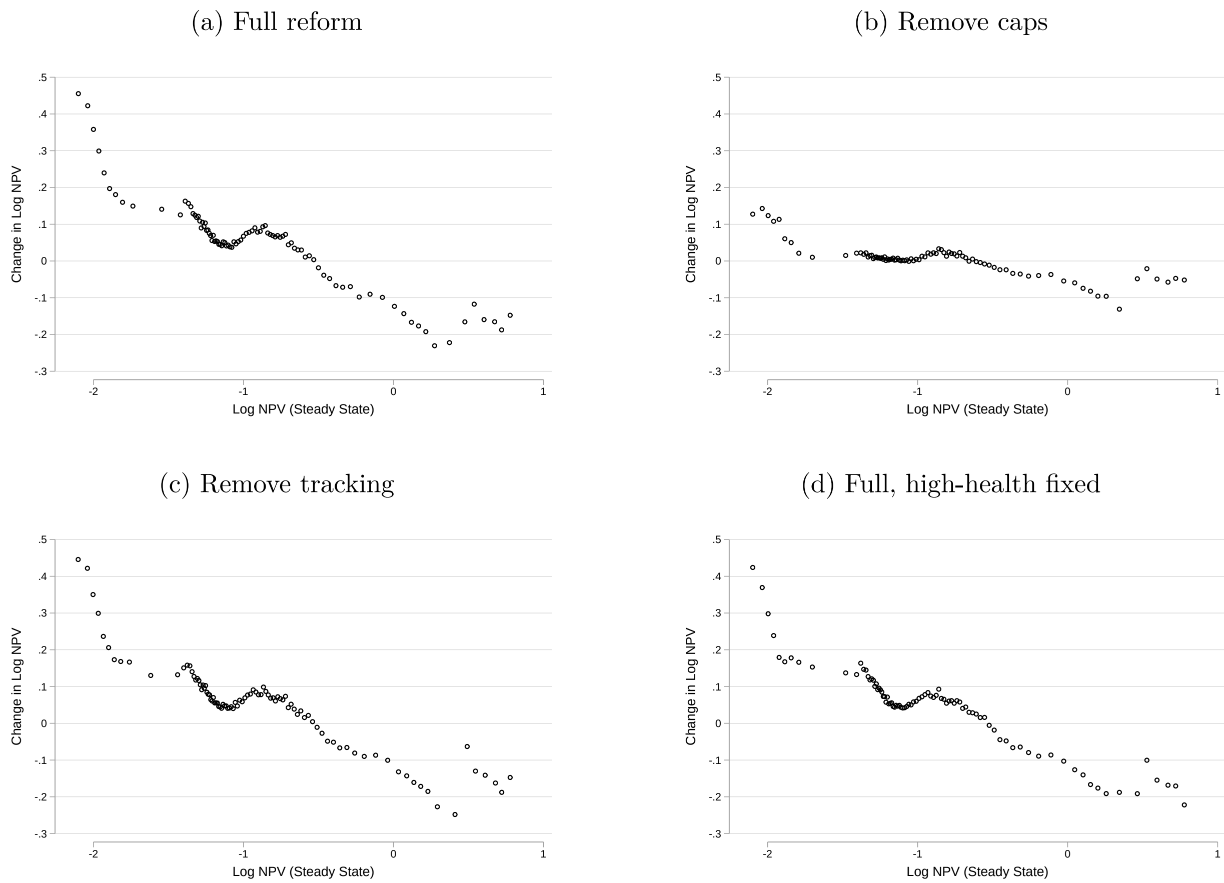
Notes: Post-shock steady state relative to the pre-shock steady state. *Income* is the mean change in log lifetime labor income (×100); *GDP* is aggregate real income; *Realloc.* is the share of students whose chosen field differs from the pre-shock base (%). *Full reform* (the Flexible Response of the main text) removes both restrictions; *Remove tracking* lifts the academic tracking but keeps the caps; *Remove caps* removes the caps but keeps the tracking; *Full, high-health fixed* is the full reform with the High Health field (medicine, dentistry, and veterinary medicine) held to the Current System. Panel B reports each reform's income gain over the Current System, the difference of the Income entries in Panel A. The Current System row reproduces Table 7. This table is discussed in Section 6.3.3.

The incidence of the alternative reforms follows the same pattern as the full reform, described in Section 6.3.4. Removing the tracking alone, removing the caps alone, and the full reform are all progressive (Figure C.7): each raises incomes at the bottom of the pre-shock distribution and lowers them at the top, through the same mechanism—access at the bottom, and the general-equilibrium compression of returns in the newly opened high-return fields at the top. For removing the tracking and the full reform this redistribution comes with large aggregate gains; removing the caps redistributes the same way but adds little to output.

C.7 Results on the Never-Constrained Economy

The main analysis projects the 2018 economy, with its current admission constraints, forward to the baseline pre-shock steady state and then applies each shock. Here we instead consider an economy that never had those constraints, and subject it to the same shocks. This asks whether a permanently flexible education system responds to disruptions differently. Table C.4 compares its post-shock steady state to its own pre-shock steady state, with and without entrants' field adjustment. The value of field adjustment is close to its Current System value: −0.22 against −0.19 under the trade war; +1.53 against +1.36 under AI, so the field margin reflects the structure of the economy rather than Denmark's admission institutions.

Figure C.7
Distributional Incidence of the Reforms (No Shock)



Notes: Binned scatter (binsreg) of the change in log lifetime NPV against log NPV in the steady state, for each reform with no shock (the standing reform). Each reform raises lifetime income at the bottom of the distribution and lowers it at the top. *Full reform* removes both the caps and the tracking; *Remove caps* removes the caps but keeps the tracking; *Remove tracking* lifts the tracking but keeps the caps; *Full, high-health fixed* is the full reform with the High Health field (medicine, dentistry, and veterinary medicine) held to the Current System. This figure is discussed in Section 6.3.3.

Table C.4
The Never-Constrained Economy

	Trade War	AI
Fields frozen	−15.87	−0.73
Fields re-chosen	−16.09	+0.80
Value of field adjustment	−0.22	+1.53

Notes: Mean change in log lifetime labor income (×100) in the never-constrained economy's post-shock steady state, relative to its own no-shock steady state, where every worker is a new entrant. *Fields frozen* holds entrants' field assignment at the pre-shock allocation; *fields re-chosen* lets them re-optimize, and the value of field adjustment is the difference. This table is discussed in Section 6.3.3.

C.8 Counterfactual Solution Algorithm

This appendix describes how we compute the economy's response to each shock. A shock is an unanticipated, permanent change in fundamentals. It arrives in the pre-shock steady state, the stationary equilibrium the model reaches from the calibrated 2018 economy with no shock, and agents thereafter have perfect foresight over the equilibrium path. The economy traces a transition from the pre-shock steady state to the post-shock steady state. We solve it in exact hat algebra, expressing all equilibrium conditions in relative changes from the pre-shock steady state (Dekle et al., 2008). We first define the equilibrium, then describe the shooting algorithm.

C.8.1 The perfect-foresight transition equilibrium

Let the shock arrive at $t = 1$, and denote relative changes from the pre-shock steady state by $\hat{x}_t = x_t/x_0$. A perfect-foresight transition equilibrium is a path, for $t = 1, \ldots, T$, of occupation wages $\{\hat{W}_{ot}\}$; goods and factor prices $\{\hat{P}_t, \hat{P}_{st}, \hat{W}^f_{st}\}$; worker value functions and inclusive values $\{V_t(k, m, o), V^e_t(k, m)\}$ and the field values $\{V_t(m \mid \mathbf{X})\}$ they imply; occupation-choice probabilities $\{\lambda_t(o' \mid o, k, m), \lambda^e_t(o \mid k, m)\}$; the field and type allocation of each entering cohort; and the resulting labor stocks $\{L_{omkt}\}$ and outputs $\{\hat{Q}_{st}\}$, such that:

(i) *Incumbents optimize.* Given the wage and price path, incumbent workers' occupation choices solve the dynamic discrete-choice problem with continuation values $V_t(k, m, o)$, and $\lambda_t(o' \mid o, k, m)$ are the implied transition probabilities.

(ii) *Entrants optimize.* Each entering cohort chooses a field of study when applying to college and, upon graduation, an initial occupation, to maximize expected lifetime utility given the inclusive values $V^e_t(k, m)$ and the field values $V_t(m \mid \mathbf{X})$, subject to the admission system.

(iii) *Stocks evolve.* The labor stock follows the law of motion with exit rate δ: each period a fraction δ of every cell is replaced by new graduates, and the remaining incumbents re-sort across occupations according to λ_t.

(iv) *Markets clear.* Occupation wages clear each labor market period by period, and goods prices and outputs satisfy the demand-side equilibrium conditions of Appendix C.1.

Two boundary conditions pin down the path. The labor allocation at $t = 0$ is the pre-shock steady state, including the cohorts already in the education pipeline: cohorts already in school when the shock arrives enter the labor market at $t \leqslant 4$ with their pre-shock field allocation and pre-shock entry-occupation behavior, since their field choices are irreversible; the first cohort able to re-optimize its field enters at $t = 5$. The economy converges to the post-shock steady state, which provides the terminal values for the backward recursion.

C.8.2 Algorithm

We solve for the equilibrium with a shooting method. We guess the path of occupation wages, solve the model forward and backward given that guess, and update the wage path until labor markets clear in every period. Each iteration proceeds in three blocks. In the *backward block* (Steps 3–6), taking the guessed wage path as given, we solve for prices by hat algebra and then for the workers' value functions and the implied field values, by backward recursion from the terminal steady state. In the *forward block* (Steps 7–9), we simulate the field and type composition of each entering cohort, compute occupation-choice probabilities, and roll the labor stock forward from the pre-shock steady state. In the *market-clearing block* (Steps 10–13), we compute aggregate expenditure, output, and the implied labor demand, and update the wage guess to move each occupation toward clearing. The full set of steps follows.

Step 1. Calculate the pre-shock steady state and the post-shock steady state

Step 2. Guess a path of wages $\{w_{ot}\}_{o,t=1,T-1}$

Step 3. Solve for the path of prices using the hat algebra:

$$\hat{W}_{st}^{f} = \frac{\prod_{o\in\mathcal{O}^f} W_{ot}^{\beta_{sot}}}{\prod_{o\in\mathcal{O}^f} W_{o0}^{\beta_{so0}}}; \quad f = P, N$$

$$\hat{P}_{st}^{P} = \left(S_{s0}^{P}\hat{A}_{stP}\left(\hat{W}_{st}^{P}\right)^{1-\rho} + (1-S_{s0}^{P})\left(\hat{P}_{st}^{M}\right)^{1-\rho}\right)^{\frac{1}{1-\rho}}$$

$$\hat{P}_{st}^{D} = \frac{1}{\hat{A}_{st}}\left[S_{s0}^{N}\hat{A}_{stN}\left(\hat{W}_{st}^{N}\right)^{1-\alpha} + (1-S_{s0}^{N})\left(\hat{P}_{st}^{P}\right)^{1-\alpha}\right]^{\frac{1}{1-\alpha}}$$

$$\hat{P}_{st} = \left[(1-S_{s0}^{I})\left(\hat{P}_{st}^{D}\right)^{1-\eta} + S_{s0}^{I}\left(\hat{P}_{st}^{F}\right)^{1-\eta}\right]^{\frac{1}{1-\eta}}$$

$$\hat{P}_{t} = \left(\sum_{s} S_{s0}\left(\hat{P}_{st}\right)^{1-\sigma}\right)^{\frac{1}{1-\sigma}}$$

$$P_t = \hat{P}_t P_0$$

Step 4. Solve the path of Bellman equations:

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

Step 5. Solve for the path of inclusive values of being type (k,m):

$$V_t^{e}(k,m) = \nu \log \sum_{o} \exp\left(\frac{-C_{o,km}^{e} + w_{ot}H_{o}(k,m)/P_t + \beta V_{t+1}(k,m,o)}{\nu}\right)$$

Step 6. Compute the NPV for each field m, for the cohort entering the labor market at t:

$$V_t(m|\mathbf{X}) = \sum_{k} V_t^{e}(k,m)\pi(k|m,\mathbf{X}).$$

This is the entry-date inclusive value the student evaluates when choosing her field: under perfect foresight, her expectation of the field's value at graduation equals its realized value at her entry date.

Step 7. Compute the field and type composition of each entering cohort, $\{L_{mkt}^{e}\}$, by Monte Carlo simulation

(a) For each individual i and period t, generate a test GPA s_i for i. We do not index by t because they will be copies each period.

(b) Generate a preference type l_i from the conditional CDF of the preference-type distribution given s_i:

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

where e is a uniform random variable.

(c) Draw one vector of Gumbel taste shocks over fields, ε_{im}, held fixed across ranks. For $r = 1, ..., R$, the r-th ranked choice is

$$m_{ir} = \underset{m \in \mathcal{C}_i}{\arg\max}\ \vartheta_{lm} + \theta_l V_t(m|\mathbf{X}) + \upsilon \log N_{irm} + \varepsilon_{im},$$

where $\mathcal{C}_i$ is the track-specific choice set and N_{irm} is the number of programs in field m still available to i at rank r. After each rank, the chosen field's program count is reduced by one, so a field may be listed more than once while it has programs remaining. This is the ranked-list model of the estimation.

(d) Assign the agent sequentially to a field in their choice set based on acceptance probabilities (conditional on s). If the agent is not admitted to the first $R-1$ choices, then they are automatically allocated to the R-th choice.

(e) Assign labor-market types analytically: given the assigned field m and GPA s, the individual's mass is distributed across types k with weights $\pi(k|m,s)$.

(f) Calculate the share of individuals belonging to each (k,m) cell and scale up to the mass of the college-entering cohort. The simulation covers college entrants; the (k,m) masses of non-college entrants are held at their pre-shock steady-state composition, with their entry-occupation choices still responding to the path.

Step 8. Using $V_t(k,m,o)$ solve for λ_{kmt} and $\lambda^e_t(o|k,m)$

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

$$\lambda^e_t(o|k,m) = \frac{\exp\left(\frac{-C^e_{o,km} + w_{ot}H_o(k,m)/P_t + \beta V_{t+1}(k,m,o)}{\nu}\right)}{\sum_{o'}\exp\left(\frac{-C^e_{o',km} + w_{o't}H_{o'}(k,m)/P_t + \beta V_{t+1}(k,m,o')}{\nu}\right)}$$

Step 9. Compute $\hat{H}^{supply}_o$

(a) Use the flow equation:

$$L_{omkt} = (1-\delta)\sum_{o'}\lambda_t(o|o',k,m)L_{o'mkt-1} + \delta\lambda^e_t(o|k,m)L^e_{mkt}$$

(b) Construct H_{omkt}:

$$H_{omkt} = L_{omkt} \times H(o,m,k)$$

(c) Compute changes in human capital supply across occupations

$$\hat{H}^{supply}_{ot} = \sum_{k,m} H_{omkt}/H_{o0}$$

Step 10. Calculate aggregate expenditure (and income)

$$\hat{E}_t = \sum_o S_{o0}\hat{H}_{ot}\hat{W}_{ot}$$

Step 11. Calculate changes in domestic output

$$\hat{Q}_{st} = S^{x}_{s0}\hat{A}^{F}_{st}\left(\hat{P}^{D}_{st}\right)^{-\eta} + (1 - S^{x}_{s0})\hat{E}_t\hat{P}^{\eta-\sigma}_{st}\hat{P}^{\sigma-1}_{t}\left(\hat{P}^{D}_{st}\right)^{-\eta}$$

Step 12. Calculate a new guess for wages, by enforcing labor market clearing and output market clearing. Choose $\hat{W}_o$ such that

$$\begin{aligned}\hat{H}_{ot} &= \sum_s \frac{H_{so0}}{H_{o0}}\hat{H}_{sot} \\ &= \sum_s \frac{H_{so0}}{H_{o0}}\frac{1}{\hat{W}_{ot}}\hat{Q}_{st}\left(\hat{P}^{D}_{st}\right)^{\alpha}\left(\hat{W}^{N}_{st}\right)^{-\alpha}\hat{\beta}^{N}_{so}\hat{A}_{sN},\end{aligned}$$

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

$$\begin{aligned}\hat{H}_{ot} &= \sum_s \frac{H_{so0}}{H_{o0}}\hat{H}_{sot} \\ &= \sum_s \frac{H_{so0}}{H_{o0}}\frac{1}{\hat{W}_{ot}}\hat{Q}_{st}\left(\hat{P}^{D}_{st}\right)^{\alpha}\left(\hat{P}^{P}_{st}\right)^{\rho-\alpha}\left(\hat{W}^{P}_{st}\right)^{-\rho}\hat{\beta}^{P}_{so}\hat{A}_{sP},\end{aligned}$$

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

Step 13. Update $\hat{W}_{ot}$ by the damped rule described under Implementation, and iterate from Step 3 again.

C.8.3 Implementation

Horizon, updating and convergence. We solve the path over a horizon of $T = 150$ periods. Each iteration updates the wage path by damped fixed-point iteration, $\hat{W}^{new}_{ot} = \chi\,\hat{W}^{clearing}_{ot} + (1 - \chi)\,\hat{W}^{old}_{ot}$ with $\chi = 0.03$, and we iterate until two criteria hold jointly: the wage path converges as a fixed point (a maximum per-period residual below 10^{-6}), and every slow-moving equilibrium object—wages, value functions, choice probabilities, the labor stocks and human-capital supplies, and the field allocation—is within a relative distance of 2.5×10^{-3} of the post-shock steady state throughout the final 20 periods, which verifies that the horizon is long enough. The binding object is the occupational labor stock L_{omkt}, which turns over only at the entry rate: with $\delta = 0.03$, an expected career is about 33 years. We set T by the most slowly converging experiment, the full admission reform of Section 6.3.3 run with no shock, which displaces the largest mass of students of any transition, and we verify that the welfare results do not change when we extend the horizon.

Welfare. We measure welfare by simulating individual careers along the equilibrium path and comparing them, agent by agent with identical draws, to the same agents in the pre-shock steady state. Incumbents enter with their pre-shock field and re-optimize occupations period by period; the entering cohort chooses a field under the path's field values $V_t(m \mid \mathbf{X})$ and then chooses occupations; the entrant cohort we track is the first free to re-optimize its field, entering the labor market at $t = 5$. Each simulated career lasts 50 years, so welfare depends on the path only through $t \approx 55$ and is unchanged by extending T. For each cohort we report the mean change in log lifetime labor income.

Supplementary Data Appendix

SA.1 More on Denmark's College System and High School Tracks

Denmark's college education system is structured to accommodate a diverse range of academic and professional aspirations. It includes two distinct pathways: Professional Bachelor's degrees (3-4 years) and University Bachelor's degrees (3 years + 2 years Master's).[45] Professional Bachelor's programs focus on technical and marketable skills that cater to immediate employment. Fields include Business Admin, Education, and Engineering, alongside vocational training. These programs are characterized by their integration of hands-on learning and workplace training, often through internships. University Bachelor's programs are more academically rigorous and provide a foundation for further studies or research-based careers.

Upper secondary schooling plays a critical role in determining students' higher education options. The Danish upper secondary education consists of several tracks, including Stx (general academic), Hf (general and flexible), Hhx (business and economics), and Htx (technical and scientific). Each track prepares students for specific fields of higher education. For instance, Hhx emphasizes business-related disciplines, while Htx focuses on engineering and technology. In 2008, reforms introduced greater flexibility to the upper secondary system, including supplemental summer courses (*Gymnasiale suppleringskurser*) to allow students to qualify for programs outside their original track. Despite these changes, the system retains a degree of rigidity, as the upper secondary specialization continues to influence the range of higher education programs accessible to students.

SA.2 Classifying Field of Study

The Danish register data contains several distinct codes, depending on database, that record field of study. These include the UDD codes (which are internal program codes) FAGOMRAADE (which map to ISCED-F codes) and HOVEDOMRAADE (which are program codes that include ISCED-F mappings). Statistics Denmark provides concordances between all codes, and we base our classification specifically on HOVEDOMRAADE. Specifically,

- HOVEDOMRAADE < 30 is level 1 (high-school education)[46]
- HOVEDOMRAADE $= 30, 40$ is level 2 (short-cycle education) — ISCED 5
- HOVEDOMRAADE $= 50+$ is level 3 (college and above) — ISCED 6

Table A.1 summarizes how we classify ISCED-F codes into our final definitions. Some codes are combined for sample size limitations, while health professions are divided between university and professional bachelor's.

SA.2.1 Concording KOT to UDDA

KOT codes map to a particular program-institution-year combination. There is no direct concordance from KOT codes to fields of study. As such, we construct our own concordance. Our goal is to map

[45] Denmark adheres to the Bologna Process. See https://eng.uvm.dk/upper-secondary-education/national-upper-secondary-education-programmes or https://www.uvm.dk/gymnasiale-uddannelser/uddannelser/gymnasial%20supplering/om-gymnasial-supplering for more background

[46] In Denmark, upper secondary education includes both general upper secondary education (Gymnasiale Uddannelser) and vocational basic courses (Erhvervsfaglige Grundforløb). The former is academically oriented, while the latter focuses on practical, job-specific training.

KOT to UDDA and then to IDA, thus we construct a mapping that is based on the connection between the students who receive admission to a particular KOT code, and the UDD code that these students start according to UDDA. The UDD codes can be mapped to fields. We have supplemented this concordance with one that we did by hand using the names of the KOT codes, as a check against our procedure. There are a few things to note.

First, KOT codes within a span of time tend to be consistent. That is to say, the KOT code for "Psychology in Copenhagen University" will be consistent in each year. However, programs can change. If a program exits, KOT codes can be recycled. We generate synthetic KOT codes by hand, based on the text strings of the KOTs. These line-up with changes in region attached a KOT's institution, but can be more stringent. The reason we would like KOT codes to be consistent over time is because the field associated to a program can change if an administrator records a program. This can lead to spurious changes in the composition of fields. Thankfully this is a rare occurrence, affecting on the order of 10% of codes. Henceforth, a KOT code refers to a synthetic KOT code which we consider to be a *unique* and *time-consistent* program (field x institution).

Second, the same KOT code can map to different field codes over time. In this case we assign a time-consistent field to avoid spurious changes in the composition of field. When there is no time consistency in the field codes, we use the following set of rules to determine the case (where, below, code means an ISCED-F 4 digit field code):

1. There is a one time switch in the code, the first year of the program is before 2000, and there are at least 3 years where the initial program is observed: In this case we use the *initial* field. This is because we want to ensure backwards compatibility with people who have completed the program before 1996. It is rare that the same program disappears for longer than 4 years, so that we assume if a program first appears in 2000 or later, it is a new program.

2. There is a one time switch in the code, the first year is before 2000, the *ISCED-F 2 Digit* code does not change, and there is less than 3 years that the initial program is observed: Assign the modal code.

3. There is a gap where a code switches and switches back: Assign the initial/final code

4. There is no change in the *ISCED-F 2 Digit* code and multiple switches: Assign modal code.

5. The initial year is 2000 or later (new program), there is a change in the program code that leads to a change in field, and it does not switch back: Assign modal code. Note that even if fields change, time consistency can be preferred because we update *everyone* who begins the program.

6. Remaining cases (of which there are less than 20 KOT codes out of $\sim$1400): Decide case-by-case, and provide a note explaining the choice.

After completing this concordance we have a unique synthetic KOT code that can be merged using true KOT-year combinations to a unique field of study. There are two remaining issues. First is that there are a small number of KOT codes that do not match to UDDA or a field directly. We assign these programs by hand based off of the text. Second, occasionally, there is a very poor match between the UDD and KOT. This occurs when students who gained admission to a KOT program end up assigned many different UDD codes. This can occur, for example, in cases where students do mixed-field-programs (e.g., "Chemistry and Physics"). For each KOT-year, we calculate the share of students in the modal UDD code for that KOT-year. If that is below 50%, we flag this and compare directly to the string in the text. We override the KOT-UDD match very rarely and note when and why.

At the end of the whole matching process, we still have two remaining issues to tackle. First, throughout, language programs are often reclassified. This occurs, for example, when universities move language programs from business departments to communications departments. It also happens within time-consistent KOT codes. Due to the tremendous variation in language code assignments, we hand-collected all language programs and assigned them to the most general super-set that many languages were classified, which was Communications. Thus, we create a synthetic field code that essentially combines language programs with communications, and moves a handful of language programs that are not directly learning a foreign language (e.g., linguistics), and combine this with humanities.

Second, some KOT codes could not be matched to UDD. Obviously these codes will be for small programs, as it is implied that a very small number of people (<=5) matched to the modal UDD code in that KOT. For these KOT codes we use the text and impute the code by hand. On the codes that we did by hand, the match rate for ISCED-F 2 digit fields between the hand-code and the modal UDD match was 96% for non-language based fields. Thus, we think the hand-coding is not particularly subject to error. The final concordance contains a KOT code, year, synthetic KOT code, original UDD-based field, share of students in the KOT who have the UDD-based field, assigned-field, and notes.

Merging Back to UDDA

With the concorded KOT codes, we can also map this back to UDDA. We find students who earned access to a particular KOT code. If their field of study begins with the dominant UDD code, or a UDD code that maps to a field in the same ISCED-F 1 digit grouping, we reassign them to the time-consistent, KOT-defined, field. Not including language codes, at the 4 digit level, this results in (possible) changes for 3.1% of code-years, while at the ISCED-F level, 2.0% of code-years are affected. In total, 1.36% of students—or about 13500 histories—need to be reassigned. Most of these individuals are not in KOT. This affects 0.11% of all person-years, and 0.65% of all person-years conditional on being college or more. We outline the procedure for updating below.

First, we define a variable for whether a student can be presumed to have started the field of study under their KOT. We do this based on three rules, all based on their first year of study in college after the KOT admission year:

1. they are in a UDD code that has at least 5 people in it, and at least 10% of people start this UDD code;

2. their UDD code is at least 50% of the total share of people; or

3. their UDD code is in the same 1-digit ISCED-F field as the field they will be reassigned to.

After this, 92.5% of students are considered to have started their KOT code. We then use the following two rules to update their code to the concordance-assigned code:

1. The 1 digit ISCED field assigned by their UDD does not change over time. (using the 2 digit code is almost identical). For example, if they start a program that is "natural science, n.e.s." according to UDD, and they are recorded as being in any natural science program *according* to UDD, we assume that their actual program has not changed and assign them the UDD based code.

2. If the UDD-based ISCED code switched to the KOT-Concordance based ISCED code at any point, we started using this UDD code going forward and reassign them to this code if it does not change. For example, if a certain program had been assigned as chemistry, and is reassigned to chemical engineering. Then we reassign people who are listed as chemistry to chemical engineering for as long as they are listed as having studied chemistry (and end their history thereafter).

Ultimately, 69% of people have codes updated (of a 1.36% subset of the full sample). Hence, a total of 63% of histories are updated. These people are considered to have started the program they applied for, finished it, and gone on to work using this field of study. What has changed is how we record the field of study relative to UDD. The remaining 37% are considered legitimate switches. That is to say, these are people who did not finish the degree for which they applied. While this is much larger than the baseline reoptimization rate, it is worth pointing out that the overwhelming number of programs where there is a discrepancy are small programs—often interdisciplinary—where students are more likely to have been reshuffled. For individuals with an updated code, and no switch later, we consider them to have graduated with the updated KOT program. We merge them to the final UDDA file.

SA.3 Defining Occupations

In order to define occupations we need to deal with three issues:

1. Between 2009 and 2010 there is a change in codes, as the ILO ISCO codes are changed to their 2010 version.

2. Many occupations are missing at various levels of digits (or have a "low quality" flag)

   For example, there are codes that are 3000, even though there is no "30" 2-digit heading. Instead, this implies that only the first digit could be ascertained with confidence.

3. Defining a level of aggregation

To address these issues, we proceed in three stages. First, we harmonize occupation codes across the ISCO revisions using the concordance of Humlum (2021). Second, we identify and correct spurious occupation transitions and impute remaining low-quality or missing occupation codes. Finally, we aggregate occupations to the classification used in the analysis.

Step 1: Harmonizing occupation codes across ISCO revisions. We rely on the concordance of Humlum (2021), which concords DISCO 6-digit codes across time. The difference between the ISCO's 4-digit occupations and the DISCO codes lies in the last two digits, which are allowed to be more specific. Humlum (2021)'s concordance is one-to-one. The match is nevertheless imperfect, as not all occupations have the full 6-digit code. As there are multiple concordances and the number of digits in a code changes across years, we do so in the following, we therefore construct the harmonized occupation codes as follows:

- First priority for a merge is given to exact 6-digit matches. We call these A merges. However, we do not allow switching to take place (or not take place) across ISCO-08 2-digit groupings *if* there is not a concurrent ISCO-88 2-digit grouping switch (or non-switch). That is to say, if there is a 2-digit change in the 88 codes, but not in the 08 codes, we do not accept the merge; if there is not a 2-digit change in the 88 codes, but there is in the 08 codes, we do not accept the merge. We call this the switching-consistency condition.

- Second priority for a merge is given by matching the modal 6-digit string in a spell (connected set of plant-years). We call these C merges. Potential A and C matches agree 98% of the time at the 4-digit level. We again impose the switching-consistency condition.

- Final priority for a merge is given by matching 4-digit codes using Humlum (2021)'s 4-digit concordance. We call these B merges. Potential A and B matches agree at the 4-digit level 88% of the time, and at the 2-digit level 95% of the time. We again impose the switching-consistency condition.

- After the main merge is done, we allow overriding of the switching-consistency condition whenever there is a consistent 2-digit ISCO-08 code within a spell. There are not many violations of the condition in the first place, and this is a very conservative imputation that affects 0.4% of the observations.

Step 2: Correcting spurious transitions and imputing missing occupation codes. After this merge is done, there is a consistent set of occupation codes over time. However, casual analysis of the data suggests significant spurious codes, especially around the break point 2009/2010. To demonstrate this, we define a "Mass Event" as a 15% change in the absolute value of the share of an occupation in the economy year-on-year (using only high quality codes, as we do not use imputed codes later in the analysis). We also look at 10% changes in large occupations. The median change is on the order of 5% in a given year. Thus, such events ought to be rare and indeed they are, affecting some 5% of the sample.

Nevertheless, the events are unevenly distributed over time, affecting some 25% of the sample in 2009/2010 and 20% of the sample in each of 1997 and 2003, and 10% of the sample in 1999, while affecting less than 2% (and often less than 0.5%) of the sample in most other years. The reason for these breaks is two-fold: there is a change in the source of occupational codes over time. Some of this is dealt with by only retaining high quality occupation codes, and ignoring imputations. Because of this, a great many of the jumps—especially those before 2009/2010—are driven primarily by workers who had low quality codes or missing codes having their code attached. In 2009/2010, there are jumps for several different reasons:

- Occasionally it is obvious that several 08 codes were mistakenly given their 88 classification and the correction only occurred later. As an example, there is a spike in workers in nursing homes (code 513 in the 88 ISCO system) becoming bartenders (513 in the 08 ISCO system) and then returning (532 in the 08 ISCO system). There is documentation from Statistics Denmark suggesting that in cases where it was not clear which nomenclature was being used, as long as the 08 code existed, it was assigned as the occupation code.[47]

- Occasionally there is a true reclassification, likely stemming from gaps in the concordance (on account of trailing digits). These often "make sense" at first glance. For example, there is large reclassification of bank tellers into financial NES workers.

The issues are also concentrated in some occupational codes more than others. In particular, ISCO codes that cover traditional "production" occupations (plant operation, machine work, etc.) suffer less from the problem because the occupation codes between 88 and 08 are almost the same for these occupations, making concording easier. The majority of issues are in occupations for higher

[47] Specifically see the document DISCO__koder__for__2010 available at https://www.dst.dk/da/TilSalg/Forskningsservice/Dokumentation/hoejkvalitetsvariable/personers-tilknytning-til-arbejdsmarkedet-set-over-hele-aaret--akm-/disco08-alle-indk.aspx.

skilled professions, as there were many changes in the coding for these occupations. For example, the creation of many IT and ICT occupations.

Unfortunately, detecting these changes is hard. We documented all cases in the spread sheet `concordanceNotes`.[48] We hand concord these cases. Since we are interested in occupational switching we take the following conservative stand: after visual inspection we back- or forward- fill occupation codes so that there is *no* switching occupation within a job spell. This is similar to the assumption of Moscarini and Thomsson (2007).

We now deal with the remaining low-quality imputations, missing codes, or codes with only one leading digit. We impose the following hierarchical rules (so that each subsequent rule only applies if no valid imputation has already been found):

1. If a worker switches firms from $t-1$ to t, the code is missing in t but not in $t+1$, and is assigned good in $t+1$, we assign o_{t+1}.
2. If a worker switches from $t-1$ to t, and the lag is good, impute the lag o_{t-1}.
3. Holes are filled. If o_t is missing and $o_{t-1} = o_{t+1}$ then o_{t-1} is assigned.
4. If the second digit is missing at t, and the leading digit between t and $t+1$ matches, and o_{t+1} is good, then o_{t+1} assigned.
5. If the second digit is missing at t, and the leading digit between t and $t-1$ matches, and o_{t-1} is good, then o_{t-1} assigned.

For codes that are not high quality, or not resolved by the above imputation scheme, we set the occupation codes to missing. In the estimation, we ignore missing data rather than integrate over it. This is not problematic (but perhaps inefficient) if missing data is random. In all of our choices we have aimed to be conservative, and assume less switching than actually occurs. Our final yearly switching rate at the 2 digit level is around 12%, reduced from 14%, which is in line with vom Lehn et al. (2022). The change is small because, as previously discussed, only a handful of years are ever affected. To this end, the effect across years is more varied. After the procedure, the switching rate is more uniform across years—including 1997, 1999, and 2003. However, switching rates remain elevated between 2009 and 2010. We exclude this from the estimation of switching elasticities and find small changes to estimates.

Step 3: Occupational aggregation. Finally, we aggregate occupations to the level used in the empirical analysis. We base our classification primarily on the Danish DISCO-08 2-digit occupation codes, with a small number of adjustments described in Table A.2. The resulting classification consists of 29 occupation groups, which are used throughout the paper.

SA.4 Tasks Construction

In addition to the Danish register data, 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.

We conduct principal component analysis on the Knowledge, Abilities, and Skills modules. We match from SOC codes to ISCO codes and use census weights to collapse to the ISCO 4 digit

[48] This is available at Sharon Traiberman's website, the servers on which the data can be accessed, and will be included in the replication package.

level. We remove management tasks from all non-management occupations as they can badly bias estimates.[49] We collapse from the ISCO 4 to the ISCO 2 level using Danish population weights.

SA.5 Constructing Education Histories

From UDG we can assign students a final exam score, as well as in which program they received the score. In our final analysis (which we call the high schooler sample), we only consider students who are within 5 years of high school graduation and who have applied at most twice. This also automatically drops students where this information is missing (for example, foreign students).[50] In our analysis we only consider students in the main 4 types of secondary program: Stx (split by specialty before 2008), Hhx, Htx, Hf. Ultimately, we keep 77% of KOT in our samples. For 10% of the sample, we do not have their high school track, or they went to a different track. For the remaining, we are missing other demographic information, or they do not meet the requirements to be included in the high schooler sample. A very small fraction of students are in specialized IB program or other programs (about 6200 students out of 1.4 million). We fold these into the Stx program. Similarly, there were in the past 1 and 2 year speciality Hhx programs, which affects about 18300 students. We combine these into a single Hhx program.

In the KOT database, we know which type of exam is required by a program (`kot_eksvo`). This exam highly correlates with the high school type assigned to students. We do not have information on conversion factors for students who apply from programs that do not match to the exam, so we treat their score as any other for the purposes of determining acceptance thresholds (described below).

SA.6 Constructing Choice Sets

SA.6.1 Determining Cutoffs

Students who apply to programs in Denmark can apply via Quota 1. This requires being allocated via one's exam score. In this case, we need to determine the set of schools that student's could *ex-post* have reasonably determined themselves eligible. As the published test score cutoffs have not been digitized (to the best of our knowledge), it is hard to determine the thresholds for admission. Moreover, even with this information, students can enter programs through alternative routes—for example through quota 2, standby, or other means. Thus we need a more consistent notion of cutoff rooted in the full set of pathways for admission to a program. We employ the following way to allocate a cutoff rule. We use this cutoff rule to then determine the number of programs in each field that each student could have entered. Specifically, we allow a 12/13 point (120/130 point scale) buffer.

1. Calculate the minimum score of a student admitted to a program with a quota 1 admission rule (otherwise assign a cutoff of 0) for each program, year, and grade scale (sometimes students overlap between the older and newer scales).

2. Calculate the median cutoff across years. If a given year's cutoff is more than 20 points below the median, we assign the first percentile of admitted scores.

[49] This occurs because occasionally production occupations will be matched with a "Supervisor" SOC code that is in turn given manager characteristics. Because of the structure of the match, where the census weight is all managers, including these occupation matches badly tilts all occupations towards managerial tasks.

[50] For the purpose of constructing test score cutoffs, we actually pull all students for whom we have high school track data, as this can be used to discipline cutoffs.

3. We update the median and recalculate. We then repeat replacing the first percentile with the 2.5th and the fifth.

4. If after this, there is a gap away from the median, we use the *largest* score of students who were applied on quota 1 and were rejected from the program (i.e., they received admission on a lower priority than the program in question). If the max is larger than the current cutoff, we use the 95th percentile.

5. For gaps far above the median cutoff, we use the score of the largest rejected applicant.

In general, there is a very high correlation between the minimum admit score and the largest reject score. The correlation between the minimum and the maximum is .62. This is high, but still reflects that there is a great deal of noise. This is especially the case in small programs, where one individual admitted via an alternative route (or perhaps misrecorded) can dramatically shift the cutoff. However, the correlation between the 5th percentile of admitted scores and 95th percentile of rejected scores has a correlation coefficient of 0.89, suggesting that this procedure is meaningfully capturing the difficulty of being admitted to some programs.

There is one additional issue: for a few years around 2008, there are students applying both under the old and new grading schema. Unfortunately, since we do not have the raw scores (only averages across classes), we cannot update the scales, and we do not know the conversion formula. However, for years when there is overlap, we take the cutoff for programs where people on both scales apply and regress one on the other. For example, we run,

$$\log GPA^7_{pt} = \beta_t + \beta \log GPA^{13}_{pt},$$

where GPA^7_{pt} is the cutoff under the 7 point system in year t for program p. These regressions in either direction produce R^2s over 0.7, suggesting a tight relationship between these cutoffs. We then impute the thresholds using this regression in years and programs where we do not observe people from one grade scale or the other applying. When this procedure is finished, we have a score cutoff for each GPA scale, for each program, in each year. There are also unconstrained programs, which have no threshold.

SA.6.2 Determining Available Options

Students may not be able to apply to all programs despite a high score. This can occur because they lacked the requisite preparation. Before 2008, there was considerable tracking at the high school level in Denmark, with upper secondary programs being divided into a mathematics and linguistics track. Tracking remains in place afterward but was mildly loosened to encourage more flexibility. Nevertheless, programs often have minimum requirements. For example, to apply for Economics at the University of Copenhagen one *needs* to have taken A-level mathematics. After 2008, the system was made to be more flexible. For example, summer programs have been expanded to allow students more flexibility. Nevertheless, rigidity remains. To determine which programs students can have entered we use their high school background. Before 2008 this is 5 different backgrounds—Hhx, Htx, Hf, Stx-Math, Stx-Linguistics—and after 2008 it is 4, as they combined the Stx programs. For each program, we look at all times this program was listed a priority for students from each background. We treat a program as eligible if it is allocated as a priority at least 5 times by students.

We know which exam students need to have taken to apply to a program, and this is *highly* correlated with educational background, so we believe that conditioning on high school background is an excellent proxy for programs which students can take. Of course, it is likely an overestimate given that we do not observe precise classes.

Once we have allocated the programs to which students with a background can apply for, we construct a choice set for each student based on a 12 (or 13, depending on scale), buffer around the calculated cutoff. Thus, for every year-high school-gpa-scale, we have a unique choice set that lists the programs available to each student within a field (including possibly 0). Our model implicitly assumes programs within a field are picked at random.