Major Choice and the Aggregate Effects of College Subsidies

Furkan Sarikaya*

[Job Market Paper]

[Click here for the latest version]

Abstract

Higher education subsidies are primarily distributed through need-based programs, without differentiating by college major. However, labor market outcomes vary significantly across majors. Science and Engineering graduates tend to earn the highest wage premiums and face the lowest unemployment rates, while there is a strong prior pattern of ability selection into these majors. I study the aggregate effects of higher education subsidies, taking into account key differences across college graduates by major. These differences include ability selection, patterns of skill formation, and frictions in post-college labor markets. I develop an equilibrium labor market search model with two-sided multidimensional heterogeneity and endogenous college and major decisions. In my model, individuals are initially sorted into college majors based on their multidimensional abilities (math, verbal, and social) and preferences. These decisions lead to differential human capital accumulation across all ability dimensions. I use data from the NLSY79 and O*NET to calibrate the model, which I then use to evaluate the effects of subsidies targeted at specific college majors. My findings indicate that Science and Engineering and Business and Economics majors demonstrate limited responsiveness to subsidies compared to Humanities and Social Sciences majors. This is because Humanities and Social Sciences majors tend to attract individuals who might otherwise opt out of college. The expenditure-neutral, welfare-maximizing subsidy scheme, which allows for differential subsidies based on college major while maintaining fixed total subsidy costs, leads to a 0.5% increase in overall welfare. This policy also results in a 35% increase in the number of Science and Engineering graduates.

Keywords: Higher education subsidies, college major choice, ability selection, job search, equilibrium

*Arizona State University, *E-Mail:* fsarikay@asu.edu. *Acknowledgments*: I am extremely grateful to my advisor, Gustavo Ventura, for his invaluable comments, guidance, and support. I also thank my committee members, Galina Vereshchagina and Domenico Ferraro, for their insightful comments and feedback. This paper has benefited from helpful discussions with Esteban Aucejo, Tomas Larroucau, Xincheng Qiu, Alessandra Peter, Cristiano Mantovani, and participants of the ASU macro workshop. All errors are my own.

1 Introduction

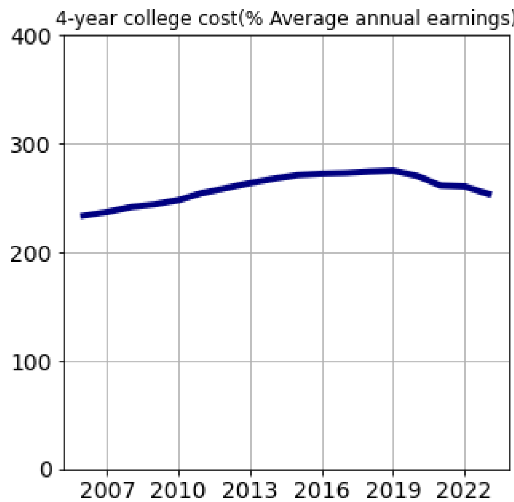
The U.S. higher education system operates on a high-tuition, high-aid model. Pursuing a college degree in the United States entails a significant financial burden, with the average published cost of a four-year college exceeding $150,000, which is about 2.5 times the annual mean earnings (see Figure 1a). Financial subsidies are also substantial. From 2006 to 2023, the average annual financial aid per student amounted to around 45% of the average published cost (see Figure 1b). Most of this financial aid is provided to students through need-based programs, with some also allocated through merit-based programs..[1] Notably, there is currently no widespread, nationally implemented policy offering subsidies based on college major choice in the United States. On the other hand, graduates with different college majors exhibit varying performances in the labor market, including differences in earnings, unemployment rates, and skill mismatches in their jobs, with outcomes generally favoring science and engineering majors. While some of these differences can be explained by ability sorting into majors (*"better"* students prefer *"better"* majors), most of the disparities persist even after accounting for this sorting (Altonji et al. (2012), Webber (2014)).

These differences have garnered policy attention towards prioritizing subsidies for certain majors, specifically for science and engineering. For instance, in Florida, policymakers have sought to steer students toward certain fields by adjusting tuition fees—lowering them for STEM majors and raising them for liberal arts. Similar incentives are offered in at least 15 other states, as reported by the National Conference of State Legislatures (Cohen (2016)).

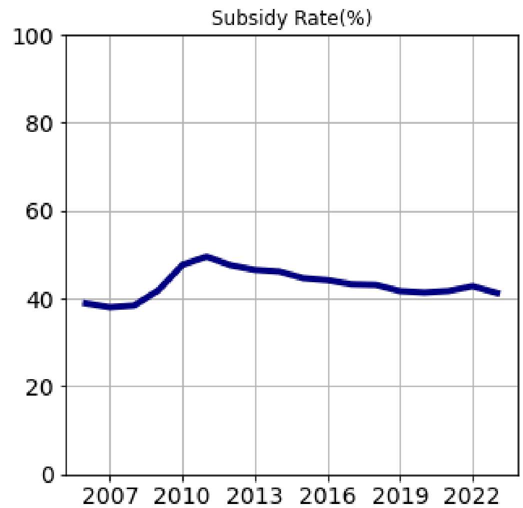
However, the aggregate effects of such policies are not straightforward. On one hand, encouraging students to pursue majors where human capital is more highly valued by the labor market could improve overall labor market outcomes. On the other hand, the extent to which individuals will respond to these incentives is uncertain, as college major choices are influenced by personal abilities and interests. Even if these policies successfully alter individual decisions, they could shift the ability composition within the prioritized majors. In other words, lower-skilled individuals may be drawn to these majors due to the increased subsidies, potentially resulting in a less skilled pool of graduates. Furthermore, changing the distribution of college graduates may have general equilibrium effects, potentially leading to imbalances in the supply and demand for certain skills across the labor market.

[1] For a brief overview of financial aid programs in the U.S., see Appendix A. For a more detailed analysis, see Dynarski et al. (2023).

Figure 1: College Cost and Subsidies



(a) College Cost



(b) College Subsidies

Sources: *Trends in College Pricing and Student Aid - College Board (2022)* and *National Occupational Employment and Wage Estimates - Bureau of Labor Statistics (2023)*. **Notes:** The average total cost for a year of college at a four-year school includes tuition, fees, on-campus room and board, books, supplies, and other expenses. Subsidies include financial aid, federal loans, grants, education tax credits and deductions, and Federal Work-Study (FWS). All prices are in 2022 dollars.

To this end, this paper studies the aggregate effects of higher education subsidies, with a special focus on college major-targeted subsidies by considering college major heterogeneity. Although there is a vast literature on the effects of higher education subsidies, it typically neglects the college major perspective by treating postsecondary education as homogeneous.[2] Departing from this approach, I build an equilibrium labor market search model with two-sided multidimensional heterogeneity and endogenous college major decisions by workers. The purpose of the model is to account for observed differences between graduates of different college majors and to evaluate the aggregate effects of higher education subsidies. Moreover, the model provides a natural environment to evaluate the impact of subsidies targeted specifically at college majors.

In the model, individuals are sorted across college majors according to their multidimensional abilities and preferences, or they may choose not to attend college at all. The decision to pursue a particular major shapes their human capital formation; for example, opting for an engineering major may result in greater accumulation of math skills com-

[2]For example, Heckman et al. (1998), Abbott et al. (2019), and Shephard and Sidibe (2019) study the general equilibrium effects of college subsidies.

pared to verbal skills, whereas choosing a liberal arts major could lead to the opposite outcome. After developing their human capital, workers and firms participate in a search within a frictional labor market where jobs have multidimensional requirements. Firms, each with heterogeneous job requirements, decide on job creation. In this environment, matchings are subject to both extensive margin (unemployment) and intensive margin (skill mismatch of workers and jobs) inefficiencies.

Modeling the endogenous choice of college majors based on individual abilities and preferences allows for evaluating the effectiveness of subsidies targeted at specific majors. In particular, the model framework suggests that human capital accumulation depends not only on the chosen major but also on individuals' pre-college abilities, which are exogenous to the model. Thus, when individuals alter their major preferences due to differential subsidies, they carry their initial abilities with them. Moreover, firms play a crucial role in shaping skill demand, as they decide which types of jobs to create in response to shifts in the composition of college graduates, or, more broadly, the changing skill composition of the labor force.

I calibrate the model using data from the National Longitudinal Survey of Youth 1979 (NLSY79) and the Occupational Information Network (O*NET). The joint distribution of pre-college abilities—specifically in the *math*, *verbal*, and *social* skill dimensions—is extracted from NLSY79. Additionally, I combine ONET and NLSY79 data to reconstruct the joint distribution of job requirements across these same three dimensions. For college majors, I focus on three distinct fields: *science and engineering* (S&E), *business and economics* (B&E), and *humanities and social sciences* (HSS). The model is then calibrated to replicate observed differences in pre-college abilities, college premiums, unemployment rates, and skill mismatch across these majors. In summary, the benchmark model captures ability selection into college majors as well as differences in the labor market outcomes across college majors.

Findings The model distinguishes college majors based on skill accumulation patterns. Calibrated values of the model parameters reveal that Science and Engineering (S&E) is math-intensive, where graduates accumulate substantial math human capital, and enrolling in this major requires strong pre-college math abilities. Humanities and Social Sciences (HSS), in contrast, are verbal-intensive. Business and Economics (B&E) lies between these two, involving moderate levels of both math and verbal skills.

The model also reveals key differences among ability types. The accumulation of social skills during college is slower and relatively uniform across all majors, suggesting that social skills are not developed to the same extent as math or verbal skills. Although social skills play a smaller role in college major selection compared to math and verbal abilities, they are critical in the labor market, particularly in worker-job matching, and offer higher returns than math or verbal skills. Among the latter two, the returns to math are significantly greater, which helps explain why S&E graduates, whose major is math-intensive, tend to experience better labor market outcomes.

Given these characteristics, I conduct several counterfactual analyses on college subsidies. First, I examine the effects of uniform changes in subsidy rates across all college majors. While uniform increases in subsidies significantly boost college participation, the composition of college graduates by major remains relatively stable. Specifically, when the benchmark college subsidy rate of 43% is raised to a full-subsidy, free-college policy, the overall college graduation rate increases by 8.4 percentage points. However, the distribution of graduates across different majors remains nearly unchanged.

Second, I allow for differential subsidy rates across majors to better understand their effects. Varying subsidies by major proves to be an effective tool for altering the composition of college graduates, and the responsiveness of each major differs significantly. For example, making the HSS major free while keeping subsidies for other majors at the benchmark rate leads 7.6% of individuals, who would have otherwise pursued different educational paths (including non-college options), to choose HSS majors. Targeting SE results in a 6.2% shift toward SE majors, while targeting BE leads to a 5.2% shift toward BE majors. These results suggest that the composition of college graduates is highly sensitive to major-specific subsidies, with HSS showing the greatest responsiveness.

Not only do sensitivities to subsidies vary, but the profile of individuals attracted to each major through targeted subsidies also differs. For instance, 70 percent of individuals drawn to HSS by a 50% increase in HSS subsidies—while keeping other subsidy rates at the benchmark—are those who would have chosen non-college pathways under the benchmark economy. In comparison, this figure is around 55% for S&E and B&E majors. The higher responsiveness of non-college individuals to HSS subsidies is attributable to the distinct characteristics of the majors. Since S&E and B&E are more math-intensive, lacking math skills incurs a higher cost for individuals pursuing these majors. This discourages many non-college individuals from choosing S&E or B&E, even when these majors offer a cost advantage.

As a result, targeting S&E and HSS yields higher and comparable aggregate output gains compared to targeting B&E. This is because HSS subsidies attract a larger share of individuals from the non-college pool. On the other hand, S&E majors offer greater accumulation of valuable skills, leading to significant gains in productivity. In contrast, B&E subsidies attract fewer individuals from the non-college pool compared to HSS and result in more limited increases in the most productive skills compared to S&E.

Finally, I calculate welfare-maximizing subsidies while keeping the total subsidy expenditure the same as the benchmark and allowing for differential subsidies across college majors. Under this scheme, the subsidy rate for S&E increases significantly to 78%, while the subsidy for HSS is reduced to 21%, and no subsidy is allocated to B&E. This adjustment yields a 0.5% welfare gain without any additional expenditure, as the total subsidy cost remains constant. Additionally, this welfare-maximizing scheme leads to a 4.3 percentage point increase in the share of S&E graduates in the economy and a modest 0.8 percentage point decrease in the overall college enrollment rate due to the reduced subsidies for other majors. Welfare gains from differential college subsidies can increase to 1% if an additional 20% in subsidy expenditure is allocated.

Related literature This study contributes at least to three strands of literature. First, it adds to the extensive body of work on college major decisions and their impact on labor market outcomes (for a comprehensive survey, see Patnaik et al. (2020)). Numerous micro-level studies have explored the determinants of these decisions, including work by Arcidiacono (2004), Arcidiacono et al. (2012), Altonji et al. (2015), and Wiswall and Zafar (2015). The literature highlights two key findings: first, individuals' abilities play a critical role in their choice of college major, and second, college majors significantly influence individual earnings even after accounting for ability sorting. This paper builds on these themes by incorporating ability sorting across college majors and differential human capital accumulation within majors. Additionally, it extends the analysis into a general equilibrium framework with frictions, providing insights beyond micro-level investigations.This framework allows me to analyze potential large-scale policies, such as subsidies targeted at specific college majors.

Second, there is a vast literature focusing on the general equilibrium effects of higher education subsidies (see, for example, Heckman et al. (1998), Lee (2005), Abbott et al. (2019), and Shephard and Sidibe (2019)). Among these, the recent work by Shephard and Sidibe (2019) is closely related to this paper. Their research explores the impact of policies aimed at increasing educational supply on the matching between workers and firms, as well as the wages of different skill groups within an equilibrium model where workers endogenously invest in education. This paper, however, diverges by focusing on college major decisions rather than overall college attendance choices, and by evaluating the effects of major-specific subsidies as opposed to uniform college cost subsidies. Additionally, a distinctive feature of my model is the inclusion of differential human capital accumulation across college majors, made possible through the multidimensionality of human capital.

Additionally, this paper aligns with research on the mismatch between workers and firms, a topic extensively explored in numerous studies examining its causes and outcomes (e.g., Sanders (2012), Perry et al. (2014), and Fredriksson et al. (2018)). Particularly relevant are the works of Guvenen et al. (2020) and Lise and Postel-Vinay (2020), which investigate the effects of multidimensional skill mismatch on wages and welfare. This paper contributes to the literature by linking skill mismatch phenomena to individuals' college major decisions, where the composition of their skills is shaped during their university education.

Outline The remainder of the paper is organized as follows: Section 2 introduces the theoretical model used for the quantitative analysis. Section 3 describes the data sources and presents insights from the data. Section 4 discusses the calibration strategy. Section 5 analyzes the aggregate effects of higher education subsidies. Section 6 examines the welfare-maximizing subsidies. Section 7 explores the role of various model properties in determining the results. Finally, Section 8 concludes the paper.

2 Model

I develop an equilibrium labor market search model with two-sided multidimensional heterogeneity and an endogenous college major decision. The model has two important features. First, individuals are sorted across college majors, or they can opt not to pursue college, based on their multidimensional innate abilities. Endogenous choice of a college major determines their human capital bundle before entering the labor market. For example, individuals with stronger math abilities are more likely to choose math-oriented majors such as engineering and tend to accumulate more math skills during college compared to verbal skills. Second, individuals are sorted across jobs in the labor market based on their post-college human capital bundles. There is a tendency for individuals to match with jobs that align closely with their skill set. However, this sorting process is not perfect, meaning individuals may choose to work in jobs that do not exactly match their skill profile in order to secure employment more rapidly.

2.1 Environment

Time is discrete, and the discount rate for the future is denoted as β. The economy consists of a unit measure of individuals who have linear preferences over income and face a probability ρ of retirement. Retired individuals are replaced by mass of ρ newborns each period. Individuals are *ex-ante* heterogeneous in their multi-dimensional innate abilities, represented by the vector $\boldsymbol{a} = \{a_n\}_{n=1}^N$, where there are N types of abilities.[3] Individuals draw their innate ability vector from an exogenous distribution: $\boldsymbol{a} \sim g_a(\boldsymbol{a})$. Note that bold letters will be used for N-dimensional vectors throughout the text. Similarly, firms have multidimensional job requirements, denoted by $\mathbf{r} = \{r_n\}_{n=1}^N$.

Immediately after they are born, individuals decide endogenously whether to attend college and, if so, which major to pursue. This decision results in a multidimensional human capital $\mathbf{h} \sim g_w(\mathbf{h})$, where $g_w(\mathbf{h})$ represents the endogenous human capital distribution.[4] To ensure conceptual consistency, I use the terms *ability*, *innate ability*, and *pre-college ability* interchangeably to refer to individuals' abilities at birth, $\boldsymbol{a}$. Similarly, I refer to individuals' human capital after college, $\mathbf{h}$, as *human capital*, *skills*, or *post-college human capital*.

After making their educational decisions, individuals initially enter the labor market as unemployed and transition to employment upon receiving and accepting job offers. In this model, only unemployed individuals engage in random job search; employed workers do not actively search. Matches experience exogenous separation shocks at a rate δ.

Distributions The distribution of (pre-college) abilities, $g_a(\boldsymbol{a})$, is exogenous to the model, while the distribution of (post-college) human capital, $g_w(\mathbf{h})$, emerges as an equilibrium outcome of individuals' education decisions. The free entry condition pins down the distribution of vacant firms, denoted as $g_v(\mathbf{r})$. In other words, firms post vacancies as long as their expected benefit is at least as big as the cost of vacancy posting $c(g_v(\mathbf{r}))$, where $c(.)$ is the convex function of mass of vacancies of type-$\mathbf{r}$ firms. The distribution of active matches between individuals with human capital $\mathbf{h}$ and jobs with requirements $\mathbf{r}$ is denoted as $g_m(\mathbf{h}, \mathbf{r})$. Table 1 illustrates the relationships between active match distribution and the equilibrium distributions of unemployed workers with human capital $\mathbf{h}$, denoted as $g_u(\mathbf{h})$. The overall counts of vacancies (V) and unemployed workers (U) are computed by integrating over the distributions of vacant firms or unemployed workers, respec-

[3] To clarify, when calibrating the model, I set $N = 3$ with three types of abilities: *math, verbal, and social*.

[4] Human capital is denoted as $\mathbf{h}$ for notational convenience. Essentially, human capital is a function of individuals' abilities and their chosen college majors, as further explained in the next subsection.

tively. Integrating $g_m(\mathbf{h}, \mathbf{r})$ over either dimension provides the distributions of employed individuals, $g_e(\mathbf{h})$, and producing jobs, $g_p(\mathbf{r})$.

Table 1: Densities

Description	Density/Distribution Function	Aggregate Value
Active matches	$g_m(\mathbf{h}, \mathbf{r})$	$M = \iint g_m(\mathbf{h}, \mathbf{r})\, d\mathbf{h}\, d\mathbf{r}$
Employed workers	$g_e(\mathbf{h}) = \int g_m(\mathbf{h}, \mathbf{r})\, d\mathbf{r}$	$E = \int g_e(\mathbf{h})\, d\mathbf{h}$
Unemployed workers	$g_u(\mathbf{h}) = g_w(\mathbf{h}) - g_e(\mathbf{h})$	$U = \int g_u(\mathbf{h})\, d\mathbf{h}$
Producing firms	$g_p(\mathbf{r}) = \int g_m(\mathbf{h}, \mathbf{r})\, d\mathbf{h}$	$P = \int g_p(\mathbf{r})\, d\mathbf{r}$
Vacant firms	$g_v(\mathbf{r})$ - free entry	$V = \int g_v(\mathbf{r})\, d\mathbf{r}$

2.2 College Major Choice

At the beginning of their lives, individuals decide whether to attend college and, if they choose to do so, they must also decide on their college major. There are I different college majors.[5] Therefore, individuals have a total of $I + 1$ education choices indexed by $i \in \{0, 1, 2, \ldots, I\}$, where $i = 0$ represents the choice of not going to college. Each college major i is characterized by its unique skill content: $\mathbf{c}^i = \left\{c_n^i\right\}_{n=1}^N$. As elaborated below, the skill content determines both the human capital accumulation of individuals who choose major i and the effort required to obtain a degree in this major.

Human capital accumulation Individuals accumulate multidimensional human capital as a result of the interaction between their abilities and their choice of college major. Specifically, the human capital of an agent with the ability vector $\boldsymbol{a}$ after completing education at i is given by:

$$h(\boldsymbol{a}, i) = \left\{a_n + a_n^{\gamma_n} c_n^{i(1-\gamma_n)}\right\}_{n=1}^N \tag{1}$$

The choice of a college major determines the composition of individuals' human capital, i.e., each college major has a specific set of skill content $\mathbf{c}^i$. This implies that some college majors might be more intense in certain skill types than others. Specifically, post-college human capital in skill type n is the type n ability of an individual plus the interaction of the type n ability and related skill content of the chosen major i.[6] The second term in

[5]To clarify, when calibrating the model, I use three different college majors: *science and engineering*, *business and economics*, and *humanities and social sciences*.

[6]The human capital accumulation is indeed in the spirit of Ben-Porath (1967). However, instead of the interaction of ability and endogenous skill investment decision in a continuous space, as is the case

equation 1 implies that there is complementarity between the ability and skill content of the major across each dimension. The degree of complementarity may vary across skill types depending on the parameter γ_n. For example, it may be the case that the innate social abilities are relatively more important in social skill accumulation in college, while college education play a larger role in accumulation of verbal skills.[7]

I assume the skill content of not going to college is zero, denoted as $\mathbf{c}^0 = \mathbf{0}$. This normalization assumption implies that if an individual chooses not to go to college, their human capital stays at the level of their ability: $h(\boldsymbol{a}, 0) = \boldsymbol{a}$.

This form of human capital accumulation technology allows individuals to differ in skill composition based on their college major choice rather than associating each college major with a specific skill type. For example, a history major graduate may also possess a level of math skills, potentially not as much as an engineering graduate. Furthermore, even among individuals choosing the same major, post-college skills may differ based on their pre-college abilities. For instance, there could be engineering graduates who are better at social skills compared to their peers within the same major.

Cost of college: The cost of college education comprises three components: pecuniary, non-pecuniary (study effort), and time cost. Pecuniary cost refers to the monetary expenses associated with college, net of subsidies, and is denoted by:

$$p^i(1-\tau^i) \qquad \forall i \in \{1, 2, \ldots, I\} \tag{2}$$

where p^i and τ^i represent monetary expenses and subsidies for major i, respectively.[8]

The non-pecuniary (or study effort) cost is associated with the effort required by individuals to complete a college degree. This effort depends on the difference between individual abilities and the skill content of the chosen major:

$$d(\boldsymbol{a}, i) = \sum_{n=1}^{N} k_n [max\{c_n^i - a_n, 0\}]^2 \tag{3}$$

As individual abilities fall short of the skill contents of the chosen major, individuals are required to exert increasingly more effort.[9] For example, individuals with lower math

in Ben-Porath, here the interaction occurs between ability and a vector $\boldsymbol{c}$ characterizing college majors. Individuals choose the appropriate $\boldsymbol{c}$ vector, potentially from a discrete set of vectors.

[7] i.e $\gamma_{social} > \gamma_{verbal}$

[8] Note that $p_0 = 0$. At the benchmark, I will assume that tuition fees and subsidies are the same across majors, ensuring that the monetary cost of each major is identical.

[9] Similar ideas have been used in empirical literature exploring the determinants of college major choice.

abilities need to exert more effort when studying majors with high math skill content. The parameters k_n capture the idea that the effort needed to be exerted in certain skill types, in case of a skill shortage, may be larger than in others.

The final cost associated with college education is the opportunity cost. While those who decide not to attend college can directly enter the labor market, others must spend a certain amount of time in college, forgoing potential income from the labor market during this period. For future reference, the total pecuniary and study effort cost of college major i for an individual with ability $\boldsymbol{a}$ is given by:

$$\mathcal{C}(\boldsymbol{a}, i) = p^i(1 - \tau^i) + d(\boldsymbol{a}, i) \tag{4}$$

It is worth noting that while the pecuniary cost is the same for everyone within a given major, the non-pecuniary cost is private and varies based on each individual's pre-college ability set.

Education choice: Individuals have preferences over education options, denoted by $\boldsymbol{\epsilon} = \{\epsilon_1, \epsilon_2, \ldots, \epsilon_I\}$, which are drawn from a Gumbel distribution with a zero location parameter and a scale parameter σ. This stochastic component may captures any additional benefits and costs associated with an education option that are not described above, such as parental income, gender, and interest in the subject.[10]

I assume that workers enter the labor market as unemployed and denote the value of being unemployed for a given level of human capital, $V_u(\cdot)$. The optimal education choice of an individual with innate ability $\boldsymbol{a}$ and preference $\boldsymbol{\epsilon}$ is given by

$$V(\boldsymbol{a}, \boldsymbol{\epsilon}) = max_{i \in \mathcal{I}}\{-\mathcal{C}(\boldsymbol{a}, i) + \epsilon_i + \beta_i V_u(h(\boldsymbol{a}, i))\} \tag{5}$$

Notice that the discount factor, β_i, in this equation also depends on the education choice i and differs from the per-period discount factor β. The education option-specific discount factor β_i represents the opportunity cost of education. While individuals not going to college can directly enter the labor market and search for a job ($\beta_0 = 1$), those who choose

For example, Arcidiacono (2004) assumes that effort is a function of the individual's ability relative to their peers within the same major. Ahn et al. (2019) document that the grading policy is more strict in STEM majors.

[10]The recent empirical literature has studied several dimensions of this preference or 'taste'. For example, Xia (2016) finds that the probability of a student choosing a major that corresponds to the occupation of a family member is strongly correlated with the family member's wage at the time the major choice is made. For an exhaustive survey of the literature, see Patnaik et al. (2020).

to attend college must discount their value as they spend time in education.[11]

This formulation implies that the choice of college major (or college itself) depends on three components. First, the total cost associated with each education option, including pecuniary cost, effort cost, and the opportunity cost of time. Second, the preference (or taste) shock, a stochastic component representing other factors affecting education choice. Third, the value of being in the labor market with the skill bundle resulting from each education choice.

2.3 Preference and Technology

Individuals have linear preferences over income. For notational convenience, I use $\mathbf{h}$ to represent the after-college human capital of individuals, even though human capital is indeed a function of individuals' initial abilities and college major choices.[12]

The flow utility of an unemployed individual is $b\overline{\omega}$, implying that unemployed individuals receive a certain fraction of the average wage $\overline{\omega}$. This value includes unemployment benefits, home production activities, and the value of leisure. The flow utility of an employed type-$\mathbf{h}$ worker in a type-$\mathbf{r}$ job is the match-specific wage, $\omega(\mathbf{h},\mathbf{r})$, net of the disutility from work, $g(\mathbf{h},\mathbf{r})$.[13] Following Lise and Postel-Vinay (2020), the disutility of work arises only if the worker is overqualified for their job in a particular skill dimension:

$$g(\mathbf{h},\mathbf{r}) = \sum_{n=1}^{N} \phi_n [max\{h_n - r_n, 0\}]^2 \tag{6}$$

While overqualification results in disutility for individuals, being underqualified in a match leads to a loss of output. The production function is specified as follows:

$$y(\mathbf{h},\mathbf{r}) = \sum_{n=1}^{3} \left[\pi_n(r_n + h_n) - \eta_n[min\{h_n - r_n, 0\}]^2\right] \tag{7}$$

The production function is assumed to be increasing across all dimensions of job requirements and human capital. This implies that jobs with higher requirements are more productive, regardless of the worker's human capital, and similarly, individuals with higher skills are more productive, regardless of job requirements. Thus, the first term of

[11] Specifically, college is assumed to take four years, while the model period is a quarter, so $\beta_i = \beta^{16}$ for all $i \neq 0$.

[12] After education decisions are made, individuals enter the labor market, and at this stage, the only relevant factor is their human capital bundles. Therefore, it is more convenient to use $\mathbf{h}$ instead of $h(\boldsymbol{a}, i^*)$.

[13] An individual of type $\mathbf{h}$ refers to an individual with after-college human capital $\mathbf{h}$, irrespective of their college major choice.

the statement stands for the full potential of the match. On the other hand, the second term stands for the deviation from this full potential, specifically the output loss caused by the underqualification of the worker in the job. The parameter η_n allows the varying skill mismatch cost across different skills.

2.4 Search and Matching

In this frictional environment, only unmatched agents(unemployed individuals and vacant firms) participate in a random search. The matching technology is described in the Cobb-Douglas form with constant returns to scale, where ψ is the elasticity of matches with respect to unemployment:

$$M(U, V) = AU^{\psi}V^{1-\psi} \tag{8}$$

The rate at which vacant firms meet unemployed workers is given by $q_v(\theta) = A\theta^{\psi}$, where $\theta = \frac{V}{U}$ is the market tightness. Similarly, the probability that unemployed workers meet vacant firms is $q_u(\theta) = A\theta^{1-\psi}$.

In this setup, meeting does not necessarily imply matching. A successful match occurs when both the worker and the firm are mutually willing to consummate; otherwise, they may opt to continue their search for better matches. In equilibrium, agents will accept any job for which the joint surplus $S(\mathbf{h}, \mathbf{r})$ is positive. Then the set of acceptable jobs for a type-$\mathbf{h}$ worker can be defined as follows:

$$A_w(\mathbf{h}) = \{\mathbf{r}' : S(\mathbf{h}, \mathbf{r}') \geq 0\} \tag{9}$$

Similarly, the set of acceptable workers for a type-$\mathbf{r}$ firm:

$$A_f(\mathbf{r}) = \{\mathbf{h}' : S(\mathbf{h}', \mathbf{r}) \geq 0\} \tag{10}$$

As these sets get larger, the probability for agents to meet an acceptable partner (conditional on meeting a partner) increases.

2.5 Surplus Sharing

Let $V_u(\mathbf{h})$ represent the unemployment value for a type-$\mathbf{h}$ worker, $V_e(\mathbf{h}, \mathbf{r})$ is the value of a type-$\mathbf{h}$ worker employed at a firm of type-$\mathbf{r}$, $V_v(\mathbf{r})$ is the value of a vacancy for a firm of type-$\mathbf{r}$, and $V_p(\mathbf{h}, \mathbf{r})$ stands for the value of a firm of type-$\mathbf{r}$ employing a type-$\mathbf{h}$ worker.

The surplus resulting from a match between a worker of type-$\mathbf{h}$ and a firm of type-$\mathbf{r}$ is the sum of the differences between the values of being matched and unmatched for workers and firms:

$$S(\mathbf{h},\mathbf{r}) \equiv V_p(\mathbf{h},\mathbf{r}) - V_v(\mathbf{r}) + V_e(\mathbf{h},\mathbf{r}) - V_u(\mathbf{h}). \tag{11}$$

Following Shimer and Smith (2000), I assume that wages are determined by Nash bargaining based on the match surplus between workers and firms where the worker's bargaining power is α. This implies:

$$\alpha S(\mathbf{h},\mathbf{r}) = V_e(\mathbf{h},\mathbf{r}) - V_u(\mathbf{h}) \tag{12}$$

$$(1-\alpha) S(\mathbf{h},\mathbf{r}) = V_p(\mathbf{h},\mathbf{r}) - V_v(\mathbf{h}) \tag{13}$$

This sharing rules determine the match specific wages, $\omega(\mathbf{h},\mathbf{r})$:

$$\omega(\mathbf{h},\mathbf{r}) = \alpha\left(y(\mathbf{h},\mathbf{r}) + c(g_v(\mathbf{r}))\left[\frac{\int_{A_w(\mathbf{h})} g_v(\mathbf{r}) S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}}{\int_{A_f(\mathbf{r})} g_u(\mathbf{h}) S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{h}}\right]\right) + (1-\alpha)[g(\mathbf{h},\mathbf{r}) + b\overline{\omega}] \tag{14}$$

The match-specific wage is calculated as the weighted average of the value of the match and the worker's option value, with the weights determined by the worker's bargaining power, α. The term with integrals accounts for the competition among type-$\mathbf{r}$ firms to hire type-$\mathbf{h}$ workers. As more workers find type-$\mathbf{r}$ firms acceptable, the wage offered by these firms decreases. This suggests that by choosing a particular education option or type of human capital, individuals create an externality for others pursuing similar human capital. See Appendix B for the detailed derivation of wage function.

2.6 Recursive Formulation

The value function of an unemployed individual with human capital $\mathbf{h}$ is recursively defined as:

$$V_u(\mathbf{h}) = b\overline{\omega}+\beta(1\text{-}\rho)\left[\underbrace{(1\text{-}q_u(\theta))\,V_u(\mathbf{h})}_{\text{no meeting}} + \underbrace{q_u(\theta)\int\limits_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}V_e(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}}_{\text{successful match}} + \underbrace{q_u(\theta)V_u(\mathbf{h})\int\limits_{A_w^c(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}\mathrm{d}\mathbf{r}}_{\text{meet unacceptable firm}}\right] \tag{15}$$

There is a probability of $(1 - q_u(\theta))$ that a worker will not meet any firm in the current period, leading to continued unemployment in the subsequent period. Conversely, with a probability of $q_u(\theta)$, a meeting occurs. If the firm's type aligns with the worker's set of acceptable jobs, a successful match is established. The term $g_v(\mathbf{r})/V$ represents the probability of encountering a type-$\mathbf{r}$ firm, serving as a weighting factor on the value of employment, $V_e(\mathbf{h}, \mathbf{r})$. However, the firm may also be an unsuitable match, falling into A_w^c, the complement of the set of acceptable jobs. This complementary set includes all firms with which the worker is unwilling to form a match. Only successful matches result in employment.

The value of an employed worker with human capital $\mathbf{h}$ at a job with requirements $\mathbf{r}$ is given by:

$$V_e(\mathbf{h},\mathbf{r}) = \omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)[\,\underbrace{\delta V_u(\mathbf{h})}_{\text{separation}} + \underbrace{(1-\delta)V_e(\mathbf{h},\mathbf{r})}_{\text{continued employment}}\,] \tag{16}$$

where $\omega(\mathbf{h}, \mathbf{r})$ is the match-specific wage and $g(\mathbf{h}, \mathbf{r})$ is the disutility in the case of overqualification. The continuation value is contingent on the exogenous separation shock, δ.

Firms' values are also similar. The value of a vacant job is as follows

$$V_v(\mathbf{r}) = -c(g_v(\mathbf{r}))+\beta\left[\underbrace{(1\text{-}q_v(\theta))\,V_v(\mathbf{r})}_{\text{no meeting}} + \underbrace{q_v(\theta)\int\limits_{A_f(\mathbf{r})}\frac{g_u(\mathbf{h})}{U}V_p(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{h}}_{\text{successful match}} + \underbrace{q_v(\theta)V_v(\mathbf{r})\int\limits_{A_f^c(\mathbf{r})}\frac{g_u(\mathbf{h})}{U}\mathrm{d}\mathbf{h}}_{\text{meet unacceptable worker}}\right] \tag{17}$$

As is the case with the value of unemployed individuals, there are three contingencies: no meeting, a meeting with a worker in the acceptable set of the firm, and a meeting with a worker outside of the acceptable set. $g_u(\mathbf{h})/U$ represents the probability of encountering a specific worker type in the labor market.

Finally, the value of a producing job of type-$\mathbf{r}$ matched with a worker of human capital $\mathbf{h}$ is given by:

$$V_p(\mathbf{h},\mathbf{r}) = y(\mathbf{h},\mathbf{r}) - \omega(\mathbf{h},\mathbf{r}) + \beta[\underbrace{(\delta(1-\rho)+\rho)V_v(\mathbf{r})}_{\text{separation}} + \underbrace{(1-\rho)(1-\delta)V_p(\mathbf{h},\mathbf{r})}_{\text{continued employment}}] \tag{18}$$

The current period payoff is the match-specific output net of the bargained match-specific wage. In the subsequent period, there is a probability $(1-\rho)(1-\delta)$ that the match endures, generating the same value in the next period but discounted by β. Alternatively, the match terminates with complementary probability, giving rise to the option value of a vacancy.

2.7 Equilibrium

Before defining equilibrium, three key equations need to be introduced.

Surplus function The value functions can be expressed as follows:[14]

$$V_u(\mathbf{h}) = b\overline{\omega} + \beta(1-\rho)\left[V_u(\mathbf{h}) + \alpha q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}\right] \tag{19}$$

$$V_e(\mathbf{h},\mathbf{r}) = \omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1-\rho)[V_u(\mathbf{h}) + \alpha(1-\delta)S(\mathbf{h},\mathbf{r})] \tag{20}$$

$$V_v(\mathbf{r}) = -c(g_v(\mathbf{r})) + \beta\left[V_v(\mathbf{r}) + (1-\alpha)q_v(\theta)\int_{A_f(\mathbf{r})}\frac{g_u(\mathbf{h})}{U}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{h}\right] \tag{21}$$

$$V_p(\mathbf{h},\mathbf{r}) = y(\mathbf{h},\mathbf{r}) - \omega(\mathbf{h},\mathbf{r}) + \beta[V_v(\mathbf{r}) + (1-\alpha)(1-\delta)(1-\rho)S(\mathbf{h},\mathbf{r})] \tag{22}$$

The surplus function $S(\mathbf{h},\mathbf{r})$ is sufficient to characterize all labor market values—$V_u(\cdot)$, $V_e(\cdot)$, $V_v(\cdot)$, and $V_p(\cdot)$—each of which can be expressed in terms of the surplus $S(\cdot)$. Then the surplus function can be obtained by plugging equations 19-22 into equation 11 under the free entry condition, $V_v(\mathbf{r}) = 0$:

[14] Detailed derivations are provided in Appendix B.

$$S(\mathbf{h},\mathbf{r}) = y(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\delta)(1\text{-}\rho)S(\mathbf{h},\mathbf{r}) - \left(b + g(\mathbf{h},\mathbf{r}) + \beta\left[\alpha(1\text{-}\rho)q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}\right]\right) \tag{23}$$

The match surplus is composed of match-specific output and the discounted match surplus, net of the value of the worker's outside option.

Stationary match distribution In the stationary equilibrium, the number of matches destroyed equals the number of matches created for all combinations of worker and firm types in the matching set:

$$(\delta + \rho - \rho\delta)g_m(\mathbf{h},\mathbf{r}) = g_u(\mathbf{h})q_u(\theta)\frac{g_v(\mathbf{r})}{V}\mathbb{1}\{S(\mathbf{h},\mathbf{r}) \geq 0\} \tag{24}$$

The left-hand side represents the inflow to the unemployment pool, where $(\delta + \rho - \rho\delta)$ fraction of matches dissolve every period. The right-hand side is the outflow where the term $g_u(\mathbf{h})q_u(\theta)\frac{g_v(\mathbf{r})}{V}$ represents the probability of a type-$\mathbf{h}$ unemployed individual and a type-r vacant job meeting, with $\mathbb{1}S(\mathbf{h},\mathbf{r}) \geq 0$ indicating the acceptance decision, as both parties accept only when the joint surplus is positive.

Free-entry condition Each type of firm post vacancies as long as the value of vacancy creation is positive, i.e. $V_v(\boldsymbol{r}) \geq 0$. After simple algebra, this free entry condition can be formulalize as follows:

$$c(g_v(\mathbf{r})) = \beta(1-\alpha)q_v(\theta)\int_{A_f(\mathbf{r})}\frac{g_u(\mathbf{h})}{U}S(\mathbf{r},\mathbf{h})\mathrm{d}\mathbf{h} \tag{25}$$

The vacancy posting cost takes the form of $c(g_v(\mathbf{r})) = c_r g_v(\mathbf{r})^2$, where c_r is specific to firm type.[15]

Definition 1 *A stationary equilibrium for this economy consists of objects denoted as $V(.)$, $S(.)$, $A_w(.)$, $A_f(.)$, $g_m(.)$, and $g_v(.)$, satisfying the following conditions:*

1. *Individuals make optimal education decisions ($V(.)$ solves equation (5)).*

2. *The last firm entrant achieves zero expected profits ($g_v(.)$ solves equation (25)).*

3. *Individuals and firms make optimal acceptance or rejection decisions ($S(.)$, $A_w(.)$, $A_f(.)$ solves equation (23), (9), and (10)).*

[15] I target the density of types of producing job to back up c_r parameters.

4. *The joint distribution of matches is stationary ($g_m(.)$ solves equation (24)).*

2.8 The Source of Inefficiency

Why should we subsidize college education? Subsidies for college education have been justified on efficiency grounds by several researchers, addressing various types of market failures, including liquidity constraints, positive externalities of education, and individual inertia.[16]

Acemoglu and Shimer (1999) highlights a specific externality related to education decisions in markets with search frictions, known as the *hold-up* problem. This issue arises when one party makes an investment and bears the cost while both parties share the payoff. In such cases, the decentralized equilibrium is always inefficient, even if the standard Hosios (1990) condition is satisfied.[17] Put simply, individuals tend to underinvest in education when the cost is private because a portion of the return on education is shared with firms, and this portion is not internalized in the wage bargaining process.

I extend this result to college major choice in Appendix C. In a stylized version of the model outlined above, I show that individuals' college major choices are inefficient unless workers have all the bargaining power (see Proposition 1 in Appendix C). Intuitively, individuals differ in the private costs associated with pursuing different majors; for example, those with lower pre-college math skills may need to exert more effort to complete a STEM degree. However, these additional efforts are not reflected in the bargaining process, which only accounts for their post-college human capital. This mechanism serves as a source of inefficiency in college major decision.

I also show that the inefficiency in the decentralized economy can be mitigated through college major-specific subsidies. Specifically, it is possible to address the inefficiencies caused by the hold-up problem by adjusting subsidies across majors in a way that offsets private cost asymmetries (see Proposition 1 in Appendix C). Moreover, this correction can be achieved while keeping the social planner's total subsidy expenditure at the benchmark level.[18]

[16]For a more detailed discussion, see sections 2.1 to 2.4 in Dynarski et al. (2023).

[17]In fact, non-degenerate equilibria are always inefficient; the only efficient equilibrium occurs when workers have all the bargaining power. However, in this case, no vacancies will be posted if there are any positive costs associated with posting them.

[18]Proposition 1 (the inefficiency of college major choice) can be extended to the quantitative model presented above, though Proposition 2 cannot. Nevertheless, the quantitative results below suggest that major-specific subsidies can improve the inefficiencies of the decentralized economy and increase overall welfare, even if they do not fully achieve a first-best outcome.

Additionally, Charlot and Decreuse (2005) show that self-selection in education is inefficient in a search environment with frictions, as individuals fail to internalize the effect their education decisions have on the wages and employment prospects of others. In other words, an increase in the number of individuals within a specific education group worsens the job search outcomes for both firms and other workers. This concept of inefficiency can be extended to the college major context. The wage equation (equation 14) illustrates that as more workers become willing to accept jobs from specific firms, the wages offered by those firms decrease. Thus, by choosing a particular major or developing a specific set of skills, individuals impose a negative externality on others following the same educational path.

3 Data

To discipline the model, I rely on two primary data sources: the 1979 National Longitudinal Survey of Youth (NLSY79) and the Occupational Information Network (O*NET). The NLSY79 tracks a nationally representative sample of individuals who were aged 14 to 22 in 1979, offering comprehensive data on their education and work history. At the outset of the survey, all respondents completed a written test, which I use to approximate their abilities across various dimensions. O*NET is an occupation-level dataset that describes the characteristics and requirements of different occupations. By merging the NLSY79 and O*NET data, I constructed a matched worker-job dataset that includes multidimensional measures of both worker abilities and job requirements. In this section, I describe these data sources and the method used to construct the measures of worker abilities and job requirements. Additionally, I provide insights into how individuals sort into college majors and jobs based on these measures.

3.1 Data Sources

3.2 NLSY79

The NLSY79 provides detailed information on individuals' educational history, such as college majors if applicable, their work history, including their occupations, as well as a set of test scores. I exclude individuals with military service experience and those weakly attached to the labor market.[19]

[19]See Appendix D.1 for details.

Construction of ability measures In the first years of the sample period, all respondents took a set of standardized tests, including the Armed Services Vocational Aptitude Battery (ASVAB), the Rotter's Locus of Control Scale (RLCS), and the Rosenberg Self-Esteem Scale (RSES). The ASVAB, administered by the United States Department of Defense, evaluates individuals in various cognitive categories. The RSES and RISB are psychometric tests designed to measure non-cognitive abilities.[20] To construct skill bundles for individuals, I follow a similar procedure to Guvenen et al. (2020).[21] First, I normalize all ability scores to lie within the range of [0, 1] through a linear transformation, ensuring comparability across different ability dimensions. Then, I construct three ability dimensions—*math*, *verbal*, and *social*—by applying principal component analysis (PCA) as follows: (i) math ability is the first principal component of *arithmetic reasoning*, *mathematics knowledge*, *numeric operations*, and *general science* test scores from the ASVAB; (ii) *verbal* ability is the first principal component of *word knowledge* and *paragraph comprehension* scores from the ASVAB; (iii) social ability is the first principal component of scores from the RISB and the RSES.

The constructed measures of math and verbal abilities are highly correlated, with a correlation coefficient of 0.81, suggesting that individuals who excel in one of these areas tend to do well in the other. In contrast, social abilities exhibit much weaker correlations with both math and verbal abilities, with coefficients of 0.40 and 0.41, respectively. These lower correlations indicate that social abilities are more distinct from math and verbal abilities, likely reflecting different underlying skill sets or factors. Despite these differences, each component—math, verbal, and social abilities—conveys unique and independent information.

Ability versus human capital The ASVAB was administered to the NLSY sample in the summer of 1981, while the RLCS and RSES tests were conducted in 1980. This timing implies that individuals took these tests at different ages; some were still in high school, while others had already completed college. This age heterogeneity at the time of testing is critical for the theoretical framework described above. For those who took the tests before reaching college age, the measured abilities reflect pre-college abilities ($\mathbf{a}$ in the model). In contrast, for individuals who had already graduated from college, the measured abilities represent post-college human capital ($\mathbf{h}$ in the model). To account for this variation, I construct two subsamples by exploiting the age heterogeneity at the time the tests were

[20]See Appendix D.2 for a detailed description of these tests.

[21]Several other papers have used these tests to create skill bundles for individuals. Some examples: Huang and Qiu (2021), Lise and Postel-Vinay (2020), Lindenlaub and Postel-Vinay (2023), and Sanders (2012).

Table 2: Samples

	Pre-college takers			Post-college takers		
	College	Non-College	Total	College	Non-College	Total
Age	18.6	18.1	18.2	23.4	23.6	23.6
Sample Size						
Cross-Section	806	2,390	3,196	621	1,998	2,609
Panel	-	-	-	11,302	35,487	46,789

Notes: Source: National Longitudinal Survey of Youth (1979). This table presents the sample sizes and average ages for two samples: *pre-college takers* and *post-college takers*, as defined in the text.

taken.

The first sample, referred to as the **pre-college takers**, includes individuals who, despite eventually earning a 4-year college degree, had completed no more than 12th grade by 1982.[22] To create a comparison group for this sample consisting of individuals who did not pursue a college education, I only include individuals who were younger than 19 years old when taking the test and who did not commence a college degree at any point in their lives. I construct the second sample, **post-college takers**, applying a similar approach. I limit the college sample to individuals who had earned a college degree (or completed at least the 16th grade) in 1982. As for the non-college comparison, I restrict the sample to individuals who were older than 23 years old in 1981.

This approach allows me to differentiate between pre-college abilities and post-college human capital, as highlighted in the model section. While the ability measures of the *pre-college takers* sample correspond to the pre-college abilities in the model ($\boldsymbol{a}$), I use the *post-college takers* sample to characterize post-college human capital ($\boldsymbol{h}$) in the model.

Table 11 presents the sample sizes and average ages for two subsamples. The pre-college takers sample, used as a cross-section, consists of 3,196 individuals, of whom 806 eventually attended college. For the post-college takers, I leverage the panel dimension, which includes 2,609 individuals and 46,789 individual-year observations.

College majors: I categorize college majors into three broad categories: Science and Engineering (S&E), Business and Economics (B&E), and Humanities, and Social Sciences

[22]Since the surveys report the highest degree completed, an individual in their first year of college in 1982 would still indicate 12th grade as their highest level of education, meaning they had not yet started college by the summer of 1981 when the test was administered. Although the exact threshold for RLCS and RSES would typically be 1981 rather than 1982, I choose to use the 1982 threshold for ASVAB, as the accumulation of social skills in college tends to be slower than that of cognitive skills—a point I will emphasize in the next section.

(HSS). Table D.1 in Appendix D.4 shows the details of this classification.

3.3 O*NET

The Occupational Information Network (O*NET) is a comprehensive database developed and maintained by the United States Department of Labor. It provides detailed information on 974 occupations, including job duties, required skills and abilities, educational requirements, salary ranges, and projected job growth. I use O*NET data to construct job requirements for occupations across the *math*, *verbal*, and *social* dimensions.

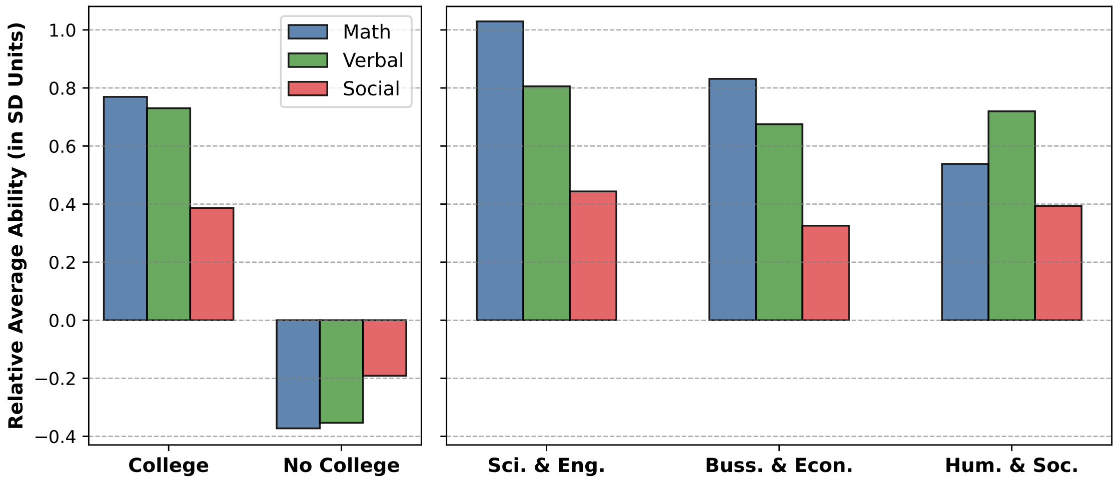
Construction of skill requirements I follow a very similar procedure to Guvenen et al. (2020) to crosswalk between abilities (or human capital) constructed from the NLSY79 sample and job requirements across the same dimensions. To broaden the ASVAB's overall appeal, the US Department of Defense introduced the ASVAB Career Exploration Program, aiming to provide career guidance to high school students. They established a connection between ASVAB test scores and O*NET occupation requirements through a tool called OCCU-Find. The Defense Manpower Data Center (DMDC) selected 26 O*NET descriptors deemed particularly relevant and assigned a relatedness score to each ASVAB category test. This crosswalk enables the translation of ASVAB test scores into corresponding skill requirements for various occupations, facilitating the analysis of how individuals' abilities align with job requirements ($\boldsymbol{r}$ in the model).

3.4 Insights from the Data

3.4.1 Ability Sorting into College Majors

I document ability selection not only at the college versus non-college level but also across different college majors. Figure 2 illustrates the average pre-college abilities of individuals based on their educational decisions, with the y-axis representing deviations from the overall mean ability in standard deviation units. Consistent with findings from several studies, individuals who pursue a college degree tend to have higher pre-college abilities than those who do not attend college, across all three skill dimensions. The most pronounced difference is in math ability, where the average for college attendees is more than one standard deviation higher. Furthermore, among those who attend college, individuals majoring in Science and Engineering (S&E) possess higher pre-college abilities across all skill dimensions, with math skills standing out in particular. The average pre-

Figure 2: Ability Sorting into College Majors



Notes: Source: NLSY79. This figure displays the average pre-college abilities across three dimensions for various education groups. The y-axis represents deviations from the overall mean ability, measured in standard deviation units. The left panel compares the pre-college abilities of college attendees versus non-attendees, while the right panel breaks down the abilities by each college major.

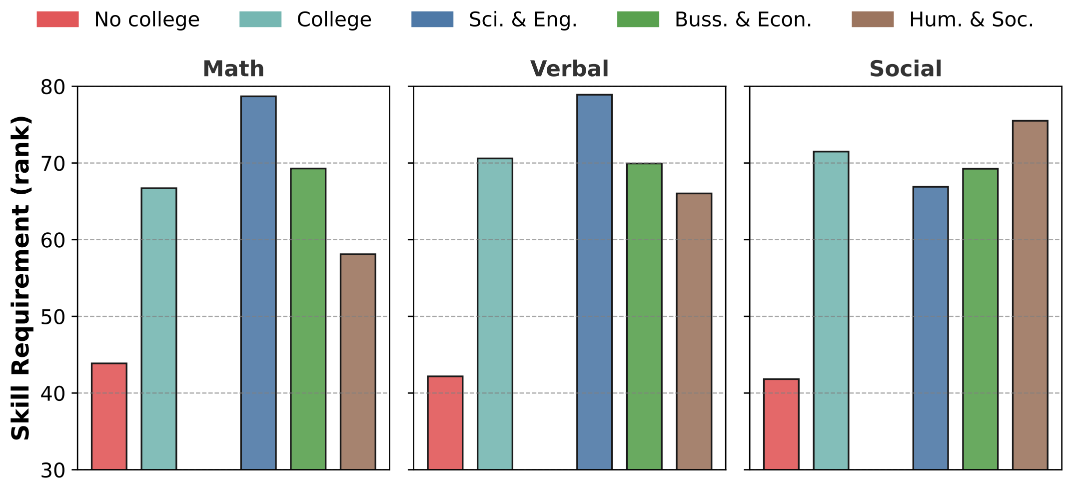
college math ability of S&E majors is approximately 0.5 standard deviations higher than that of Humanities and Social Sciences (HSS) majors. In contrast, the differences in social abilities across majors are minimal compared to the disparities in other abilities.

The observed patterns of ability selection into college majors suggest that evaluating the potential effects of higher education policies targeting specific majors should account for this dynamic. For example, using the superior labor market performance of S&E graduates as justification for increased subsidies may not necessarily lead to a higher number of S&E graduates or a more skilled workforce. Individuals may avoid choosing S&E majors initially because their abilities may not be sufficient to complete the demanding coursework. Even if increased subsidies shift preferences and result in more students opting for S&E majors, the composition of S&E graduates might change. Students who would typically choose other majors—perhaps those with lower pre-college abilities—might now select S&E majors, potentially undermining the intended impact of the subsidies, which is to enhance labor market outcomes by producing graduates with higher skills.

3.4.2 Sorting into Jobs

Figure 3 presents the average rank of job requirements for each education group across three skill dimensions. Using job histories from the post-college sample, I link individuals' occupations to job requirement data from O*NET as previously described. The figure

Figure 3: Sorting into Jobs



Notes: Source: NLSY79. This figure shows the average skill requirement rankings of individuals in their current occupations across three skill dimensions.

shows that S&E graduates tend to be sorted into jobs requiring, on average, higher levels of math and verbal skills. B&E majors follow, while HSS majors rank the lowest in these dimensions. However, HSS graduates tend to be sorted into jobs with higher social skill requirements compared to the other groups.

Why do differential sorting patterns among college majors matter? Guvenen et al. (2020) finds that, even after controlling for worker skill heterogeneity, job requirements remain a significant determinant of worker productivity. In other words, all else being equal, workers in more sophisticated occupations tend to earn higher wages. This implies that changes in the composition of college majors—resulting from targeted higher education subsidies—could also influence how individuals sort into jobs. For example, individuals who shift their preferences toward a Science and Engineering (S&E) major due to incentives might not be able to secure the same high-skill jobs as those who initially chose S&E majors. The quantitative model I use addresses this issue by allowing for (i) endogenous job creation and (ii) endogenous sorting of worker characteristics and job requirements.

3.4.3 College Majors and Skill Mismatch

In an ideal world with perfect sorting, where there is no mismatch between workers and jobs, individuals with higher skills would be matched to jobs requiring higher skill levels. Therefore, I define skill mismatch as the rank difference between an individual's post-college human capital and the requirements of their job. This can be formalized as follows:

$$m_{ij} = |q(h_{ij}) - q(r_{ij})| \tag{26}$$

where $q(h_{ij})$ represents the percentile rank of the human capital of individual i in skill dimension j, and $q(r_{ij})$ is the percentile rank of the requirement ranking of individual i's job in skill dimension j.

Based on this definition, if $q(h_{ij}) - q(r_{ij}) < 0$, it implies that worker i is underqualified in skill dimension j (negative skill mismatch). Therefore, the negative skill mismatch is:

$$m_{ij}^{-} = \min(q(h_{ij}) - q(r_{ij}), 0) \tag{27}$$

If $q(h_{ij}) - q(r_{ij}) > 0$, it implies overqualification, and the positive skill mismatch is:

$$m_{ij}^{+} = \max(q(h_{ij}) - q(r_{ij}), 0) \tag{28}$$

Finally, the total mismatch for individual i, denoted $\overline{m_i}$, across all skill dimensions $j = \{m, v, s\}$, is defined as the sum of the mismatches in each dimension:

$$\overline{m_i} = \sum_{j=\{m,v,s\}} m_{ij} \tag{29}$$

Table 3 presents the average skill mismatch, along with the average positive and negative mismatches, across three skill categories and in total. Individuals without a college education exhibit the highest degree of skill mismatch across all dimensions, with a total rank difference of 0.79. Graduates from Humanities and Social Sciences (HSS) experience a similar level of mismatch. However, while the mismatch for non-college individuals is primarily due to being underqualified (negative mismatch), the mismatch for HSS graduates is largely driven by overqualification (positive mismatch).

In contrast, graduates in Science and Engineering (S&E) and Business and Economics (B&E) experience significantly lower mismatches compared to HSS graduates. This pattern is consistent with reports that 26% of HSS graduates indicate their job is not related to their college degree, compared to 20% of BE graduates and just 12% of S&E graduates. For a detailed explanation of this subjective measure of skill mismatch, refer to Appendix D.7.

Table 3: Skill Mismatch by College Major

	Non-College	Humanities and Social Sciences	Business and Econ	Science and Engineering
Math	0.26	0.25	0.20	0.19
positive	0.10	0.18	0.16	0.13
negative	-0.16	-0.08	-0.04	-0.06
Verbal	0.26	0.24	0.21	0.21
positive	0.11	0.17	0.12	0.11
negative	-0.14	-0.07	-0.09	-0.11
Social	0.28	0.27	0.23	0.24
positive	0.14	0.08	0.11	0.13
negative	-0.14	-0.19	-0.12	-0.11
Total	0.79	0.76	0.64	0.63
positive	0.36	0.42	0.40	0.36
negative	-0.44	-0.34	-0.24	-0.27

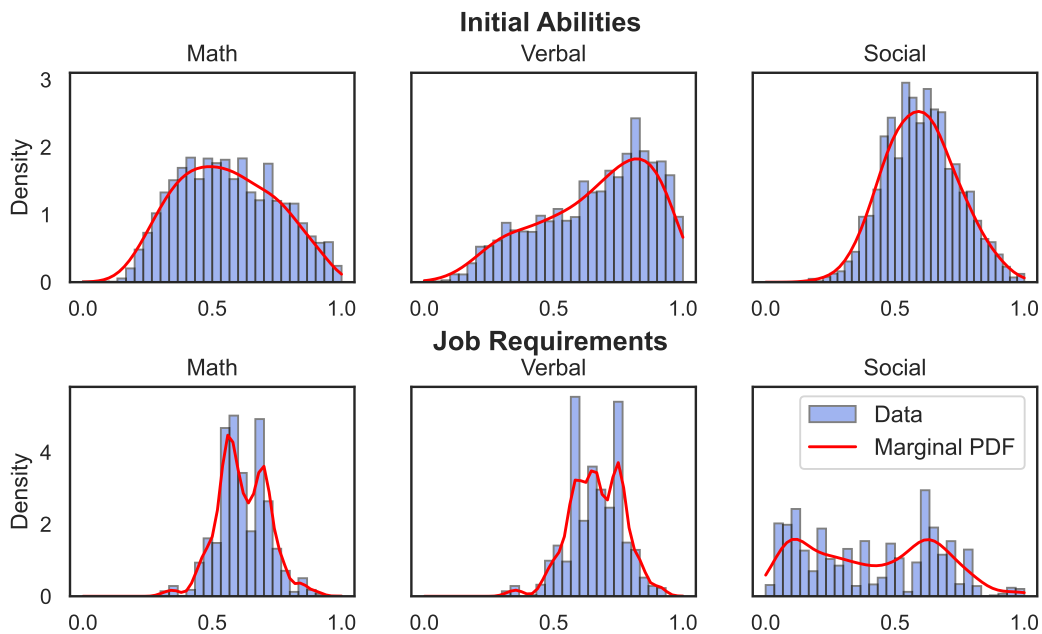
Notes: Source: NLSY79 and O*NET. Each entry in this table shows the average skill mismatch for the corresponding skill dimension among the respective education groups. The skill mismatches for an individual are described in equations (26), (27), and (28). The absolute values of negative and positive skill mismatches should sum up to the overall skill mismatch, except for rounding errors.

4 Calibration

I calibrate the benchmark model parameters to capture key features of college major choices and labor market sorting in the US economy. Some parameters are set to standard values from the literature, while others are internally calibrated within the model. In model computation I use 15 grid points for each ability dimension and 7 grid points for each job requirement dimension. With three dimensions—*math, verbal, and social*—there are 3,375 (15^3) possible initial ability bundles and 343 (7^3) possible job requirement bundles. Each individual chooses one of four education options: *Non-college, Science and Engineering (S&E), Business and Economics (B&E), or Humanities and Social Sciences (HSS).* Consequently, there are 13,500 (3,375 $\times$ 4) distinct human capital bundles.

Ability and job requirement distributions I extract the initial ability distribution of individuals, $g_a(\boldsymbol{a})$, and the job requirement distribution of producing firms, $g_p(\boldsymbol{r})$, directly from combined NLSY79 and O*NET data. First, I estimate the non-parametric multidimensional distribution of abilities using the abilities of individuals in the *pre-college takers* sample with the kernel density method. I then incorporate the estimated joint PDF of initial ability distributions into the model. Similarly, I use the type-specific vacancy posting cost parameters (c_rs) to back up the multidimensional job requirement distribution from the combined NLSY79 and O*NET data.[23] Figure 4 shows the data histograms and the estimated marginal PDFs.

Figure 4: Distribution of Pre-college Abilities and Job Requirements



Notes: Source: NLSY79 and O*NET. This figure presents the data histograms (blue bars) and the estimated marginal PDFs (red lines) for pre-college abilities and job requirements across three dimensions: math, verbal, and social. The joint probability distributions are estimated non-parametrically using the kernel density method, from which the marginal PDFs are calculated. Figure F.1 provides joint distribution of pre-college abilities

Externally calibrated parameters Table F.1 summarizes the externally calibrated parameter choices. The model period is set to one quarter. The quarterly discount factor is $\beta = 0.99$, consistent with a 4% annual risk-free interest rate.[24] The separation rate is $\delta = 0.1$, corresponding to an average employment spell of 5 quarters Shimer (2005). An average working life of 40 years implies $\rho = 0.00625$. Additionally, I set $\psi = 0.72$ and $b = 0.4$,

[23]Specifically, type-specific vacancy costs determine the densities for vacant jobs, $g_v(\boldsymbol{r})$ (equation 25), and the density of vacant jobs determines the distribution of active matches (equation 24). The density of producing jobs, $g_p(\mathbf{r})$, is calculated as $g_p(\mathbf{r}) = \int g_m(\mathbf{h}, \mathbf{r})$.

[24]Specifically, $\beta = (1 - r)^{0.25}$.

Table 4: Jointly Calibrated Parameter Values

Parameter	Definition	Value	Parameter	Definition	Value
π_{math}	Productivity	0.34	k_{math}	Skill Shortage Cost	27.4
π_{verbal}	Productivity	0.15	k_{verbal}	Skill Shortage Cost	22.0
π_{social}	Productivity	0.59	k_{social}	Skill Shortage Cost	9.13
γ_{math}	Complementarity	0.27	$c^{S\&E}_{math}$	Skill Content	0.75
γ_{verbal}	Complementarity	0.15	$c^{S\&E}_{verbal}$	Skill Content	0.18
γ_{social}	Complementarity	0.25	$c^{S\&E}_{social}$	Skill Content	0.13
η_{math}	Underskill Cost	2.27	$c^{B\&E}_{math}$	Skill Content	0.67
η_{verbal}	Underskill Cost	2.26	$c^{B\&E}_{verbal}$	Skill Content	0.25
η_{social}	Underskill Cost	8.15	$c^{B\&E}_{social}$	Skill Content	0.02
ϕ_{math}	Overskill Cost	0.01	c^{HSS}_{math}	Skill Content	0.10
ϕ_{verbal}	Overskill Cost	0.00	c^{HSS}_{verbal}	Skill Content	0.74
ϕ_{social}	Overskill Cost	0.05	c^{HSS}_{social}	Skill Content	0.09
A	Matching Efficiency	0.75	ρ	Preference param.	6.66

Notes: This table lists the jointly calibrated parameters and their values.

following Shimer (2005). I also assume the Hosios (1990) condition holds, implying that workers' bargaining power α equals the matching elasticity ψ.[25]

The other two externally calibrated parameters are the college subsidy rates (τ) and the pecuniary cost of college (p). The average financial cost of a four-year college in the U.S. between 2007 and 2022 is approximately $150,000, while the average annual wage income during the same period is about $61,538 (College Board (2022), Bureau of Labor Statistics (2023)). Therefore, I set the pecuniary college cost to be 9.75 times the average wage, i.e., $p = 9.75 \times \overline{\omega}$. Additionally, the average financial aid per student amounted to 43% of the average published cost (College Board (2022)), which I set τ as 0.43.

Internally calibrated parameters There are 26 remaining parameters to be calibrated internally: 3 skill productivity parameters (π_n's), 3 skill-college major complementarity parameters (γ_i's), 3 parameters for the cost of underqualification (η_n's), 3 parameters for the cost of overqualification (ϕ_n's), 3 parameters for the cost of skill shortage in college (k_n's), 9 skill content parameters (c^i_n's), the matching efficiency parameter (A), and the scale parameter of the college major preference distribution (ρ). I normalized the location parameter of the distribution to 0. These parameters are jointly calibrated to match the following 26 moments: shares of college majors (3), college premiums (3), unemployment

[25]The standard Hosios condition in this model does not guarantee efficiency due to ex-ante investment in skills not being fully internalized in the labor market, as discussed in section 2.8.

Table 5: Targeted Moments: Model vs. Data

Moment	Data	Model	Moment	Data	Model
Share			**Rank Correlation**		
S&E	0.12	0.12	Math	0.31	0.32
B&E	0.09	0.09	Verbal	0.36	0.36
HSS	0.16	0.15	Social	0.21	0.22
College premium			**Math abilities**		
S&E	0.71	0.78	S&E	0.75	0.75
B&E	0.61	0.63	B&E	0.71	0.71
HSS	0.43	0.54	HSS	0.65	0.60
Unemp. (%)			**Verbal abilities**		
S&E	2.6	3.0	S&E	0.84	0.82
B&E	3.2	3.3	B&E	0.81	0.79
HSS	3.5	3.6	HSS	0.81	0.79
Non-College	7.2	7.2	**Social abilities**		
Mismatch (neg)			S&E	0.67	0.63
Math	-0.12	-0.11	B&E	0.66	0.62
Verbal	-0.09	-0.08	HSS	0.66	0.62
Social	-0.02	-0.02	s.d. of NC math	0.18	0.18

Notes: This table shows the moments targeted in the model and their corresponding counterparts in the data. *s.d. of NC math* stands for the standard deviation of pre-college math abilities for the non-college group. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively.

rates among each college major (3), the negative mismatch between individual skills and job requirements across each ability dimension (3), rank correlation between individual abilities and job requirements across each dimension (3), average pre-college abilities conditional on college major choices (9), the standard deviation of pre-college math abilities for the non-college group (1), and the unemployment rate of non-college individuals (1). Table 4 lists the values of the parameters calibrated internally, while Table 5 compares the moments targeted in the model and their corresponding counterparts in the data.

I use data from the NLSY79 and O*NET to compute the negative mismatch, rank correlation, and pre-college ability moments. Specifically, the pre-college abilities are measured using the pre-college takers sample, while the mismatch and rank correlation moments are derived from the post-college takers sample. This distinction is crucial because, in the model, pre-college abilities reflect individuals' skills before any human capital accumulation through higher education. In contrast, the labor market interaction moments are based on the human capital accumulated after educational decisions have been made.

For college major shares college premiums and unemployment rates, I draw on data from the American Community Survey (ACS). The detailed methodologies for calculating these moments are outlined in Appendix D.6.

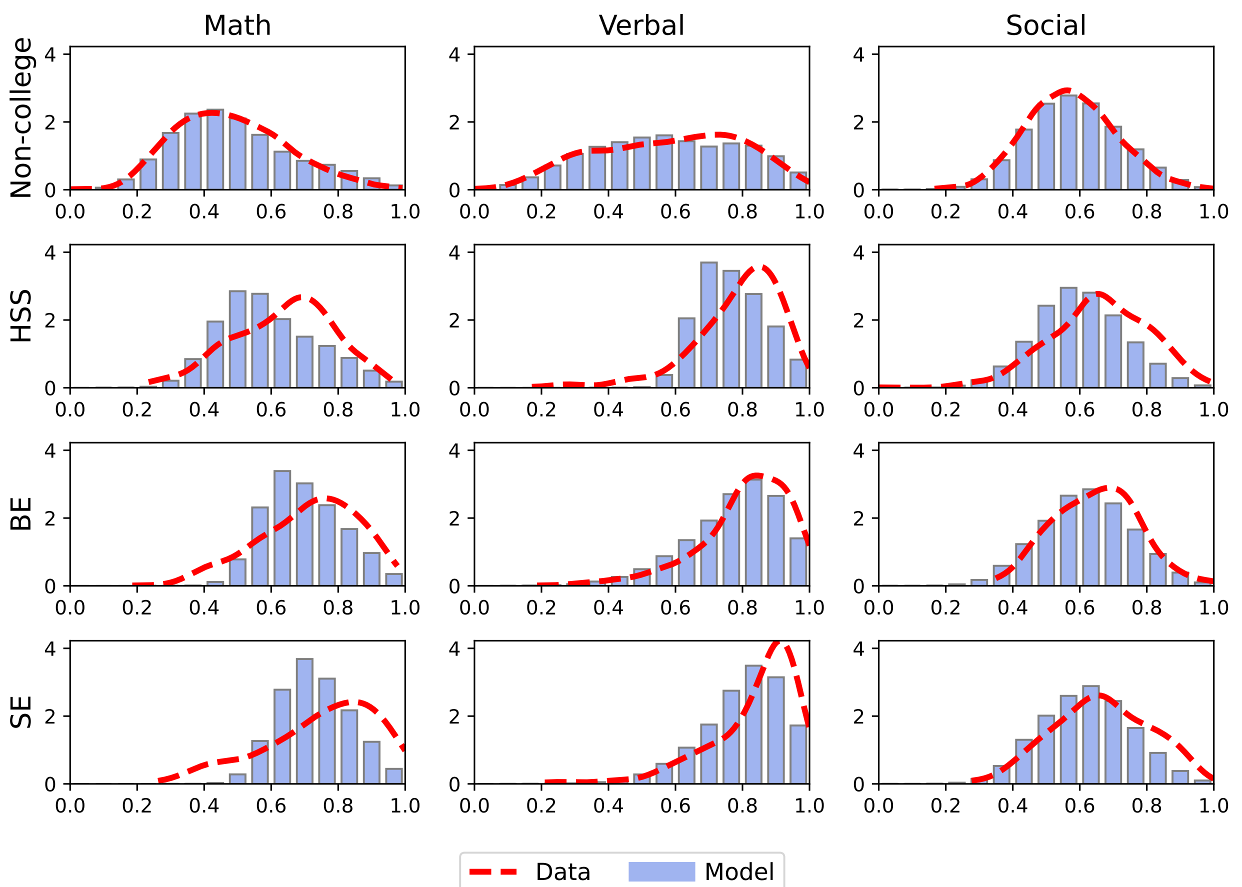
4.1 Discussion and Model Validation

Although all moments are determined jointly and many parameters influence multiple moments, specific relationships between parameters and moments can be identified. The matching efficiency parameter A primarily determines the non-college unemployment rate. The preference parameters affect the variability in pre-college math abilities among non-college individuals. If preferences had no effect, i.e., $\rho = 0$, college major choices would be driven entirely by innate abilities. In this scenario, individuals with the same innate abilities would uniformly select the same educational path, with those having lower abilities more likely to choose non-college education. However, a positive value of ρ introduces variability in educational choices among individuals with identical initial abilities. As ρ increases, individuals with higher math abilities might choose the non-college option due to their preferences, even if pecuniary and non-pecuniary costs favor other educational options.

The cost of skill shortage in college (k_n) and the skill content parameters (c_n^i) primarily determine the distribution of individuals across different educational options and the moments related to ability selection into college majors. In a hypothetical scenario where the cost of skill shortage is zero ($k_n = 0$), equation 3 suggests that the non-pecuniary cost of college majors would also be zero and independent of initial abilities. In this situation, individuals would select their college major solely based on the complementarities between their skills and the skill content of the chosen major (as described in Equation 2).

The underskill (η_n) and overskill (ϕ_n) cost parameters primarily influence labor market matching moments, such as the rank correlation between individual human capital and job requirements, as well as the extent of negative skill mismatch. Specifically, as η_n and ϕ_n approach infinity, the rank correlation converges to one, and the negative mismatch converges to zero. This indicates that individuals only accept job offers from firms that perfectly match their human capital, leading to a narrower set of acceptable jobs and workers. Consequently, these parameters also influence the variations in unemployment rates among graduates of different college majors. Finally, the productivity parameters (π_n) determine the differences in skill premiums across college majors by assigning differential weights to each ability in match output. Since match output is a key determinant of match-specific wages (see Equation 14), variations in π_n directly impact the skill premiums associated with different college majors.

Figure 5: Selection into Collge Majors: Model vs. Data



Notes: This figure compares the distribution of pre-college abilities across each skill dimension for participants in different education groups, as generated by the model (blue bars) and as observed in the data (red lines). SE, BE, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively.

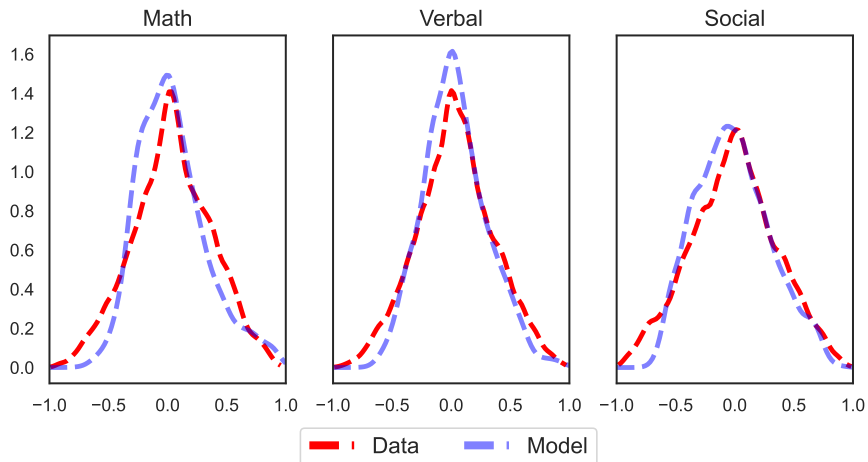
Ability sorting into college majors The primary aim of this paper is to evaluate the effects of college major-targeted subsidies. This quantitative exercise is motivated by the observation that graduates from different college majors exhibit varying labor market performances. For instance, graduates of Science and Engineering (S&E) majors typically have higher college premiums and lower unemployment rates compared to graduates from Humanities and Social Sciences (HSS) majors, who often face the opposite outcomes. The model used in this paper successfully replicates these features. However, the superior performance of S&E graduates is not solely attributable to their major; it is also influenced by the fact that students with higher abilities are more likely to select these majors.This implies that, in counterfactual exercises where cross-major subsidies are adjusted to encourage individuals to choose specific majors, students carry their existing pre-college abilities with them and accumulate human capital accordingly. They do not acquire the same level of human capital as the current incumbents of those majors. Consequently, the calibration procedure specifically targets the ability sorting of students by college major, focusing on the average pre-college abilities based on their major choices. Figure 5 compares the distributions generated by the benchmark model with those observed in the data. Although the model primarily targets average abilities, it also successfully replicates the distribution of pre-college abilities conditional on college major choice.

Sorting into jobs In addition to the selection of pre-college abilities, job characteristics play a significant role in determining individuals' labor market performance. Specifically, graduates from certain majors may secure more productive jobs that align well with their skills, resulting in higher earnings. However, it is not guaranteed that individuals who prefer these majors after college major-targeted subsidies will be able to find such well-suited jobs. To address this issue, the model endogenizes worker-job matches and targets moments related to worker-job match characteristics. This includes the magnitude of negative skill mismatch and the rank correlation between job requirements and worker characteristics. Figure 5 illustrates the distribution of skill mismatch in each skill dimension, where skill mismatch is measured as the difference between individuals' ranks of human capital and their occupations' ranks in the related dimension. Besides the targeted first-degree moments of worker-job matching, the model also closely approximates the overall distribution of matches across each ability dimension.

4.2 Lessons from Calibrated Parameter Values

The values of the skill content parameters (c_n^i) highlight the distinct attributes of each college major in terms of skill accumulation. A comparison between S&E and HSS majors shows that S&E majors are highly math-intensive, while HSS majors focus more on verbal ability development, with limited math skill growth. B&E majors fall between these two extremes—accumulating more math skills than HSS majors but less than S&E majors, and exhibiting the opposite pattern for verbal skills. Social skill accumulation, however, is slower across all majors, as indicated by the low and relatively similar values of the social skill content parameters. This suggests that social skills are not developed to the same extent as math and verbal skills in college.

Figure 6: Distribution of Skill Mismatch: Model vs. Data



Notes: This figure compares the distribution of skill mismatch between individuals' human capital and the requirements of their jobs. Skill mismatch is measured as the rank difference between each skill and the job requirements, as defined in Section 3.4.3.

Math abilities play a crucial role in determining college major choices, as a shortage in math abilities incurs a significantly higher cost compared to other skill shortages. Quantitatively, the utility cost of having less math ability than required by the chosen major is approximately three times higher than the cost associated with a shortage of social abilities. The relatively low cost linked to a shortage of social skills (k_{social}) suggests that a lack of social skills does not deter individuals from pursuing college education.

These findings suggest that social skills may not be as critical in determining college major choices. However, they play a crucial role in the labor market, particularly in worker-job matching. Social skills are more productive than verbal and math skills—evidenced by the larger π_{social} value—and the cost of skill mismatch in social skills is significantly higher compared to other skills (reflected by the larger η and ϕ values). This indicates that social skills are an important determinant of labor market performance, aligning with recent literature emphasizing the significance of non-cognitive abilities in wages and earnings (see Heckman et al. (2019) for a survey of this literature).

Among cognitive skill types, math skills are particularly noteworthy due to their productivity, as the productivity parameter for math is twice that of verbal skills. The greater accumulation of math skills in S&E and B&E majors, compared to HSS majors, is linked to the higher college premiums associated with these majors.

5 The Role of Subsidies

In this section, I begin by analyzing the effects of uniform changes in college subsidy rates on the distribution of college graduates and overall enrollment rates. Next, I introduce college major-specific subsidies by adjusting the subsidy rate for each major individually while maintaining the others at the benchmark level. This analysis helps to understand the responsiveness of each major's share to subsidies and explores the potential effects of major-specific subsidies on skill composition, earnings, and match output.

5.1 Uniform College Subsidies

Table 6 presents the share and composition of college graduates under various subsidy regimes, ranging from no subsidy to full coverage. Intermediate scenarios consider the effects of an additional $1,000 subsidy, corresponding to a subsidy rate of 0.456, and a $5,000 subsidy, corresponding to a rate of 0.563. Column 2 presents the results for the benchmark economy, where the subsidy rate is set at 0.43. In all cases, subsidy rates are uniformly applied across all college majors. Completely eliminating higher education subsidies would decrease the share of college graduates by approximately 6.4 percentage points, from 35.3% to 28.9%. Conversely, a fully subsidized higher education would increase the share of graduates by 8.4 percentage points, raising it to 43.67%—an approximately 23.5% increase in college enrollment.[26] The enrollment effect of a fully subsidized college education has not been extensively studied, given the lack of large-scale implementations. However, some studies provide comparable estimates. For example, Shapiro and Yoder (2021) estimate that Biden's free tuition plan would lead to a 17.7% increase in enrollment at four-year public colleges and universities.[27]

The effects of additional $1,000 and $5,000 subsidies are relatively modest, leading to a 0.5 percent and 1.8 percent increase in college enrollment rates, respectively. The empirical literature typically finds that a $1,000 reduction in tuition or equivalent subsidy results in a 2-3 percentage point increase in enrollment, although results are mixed.[28] There may be several reasons why my model may understate the impact of subsidies compared to empirical findings. First, many studies focus on public colleges and universities, where

[26]Notably, a fully subsidized higher education system increases the social planner's subsidy costs by a substantial 188%

[27]Biden's free tuition plan proposes covering all in-state students' tuition, whether full-time or part-time, for those with a household income of up to $125,000 who enroll in a four-year public college or university. In my model, "free college" encompasses not just tuition and fees but also boarding, food, books, and other associated costs, which may explain the higher enrollment increase predicted by the model.

[28]Deming and Dynarski (2009) for a survey of the literature.

Table 6: Uniform College Subsidies and Education Choices

Shares (%)	No Subsidies	Benchmark	\$1000 Subsidy	\$5000 Subsidy	Free College
	$\tau = 0$	$\tau = 0.43$	$\tau = 0.456$	$\tau = 0.563$	$\tau = 1$
Non-College	71.1	64.7	64.2	62.7	56.3
College	28.9	35.3	35.8	37.3	43.7
S&E	34.2	33.6	33.6	33.5	32.9
B&E	25.2	25.2	25.2	25.1	25.1
HSS	40.6	41.2	41.3	41.4	42.0

Notes: This table presents the college enrollment rates and the composition of college graduates by major under various subsidy scenarios, ranging from no subsidy to full tuition coverage. The corresponding τ values for the \$1,000 and \$5,000 subsidies are calculated by scaling each amount to the average annual published cost of college, which is \$37,500. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively. The entries in the bottom panel display the share of each major among college graduates.

tuition is lower, and marginal increases in subsidies cover a larger portion of tuition costs. Additionally, these studies often target specific groups of students, such as those eligible for particular grant programs, who are more likely to be on the margin of the college decision. Similarly, most of the grand aids as evaluated in empirical literature targets students from lower income families which are more likely to be more responsive to subsidies.[29][30]

While uniformly adjusting the college subsidy rate effectively influences overall college enrollment, it has a minimal impact on the composition of college graduates. Moving from no subsidy to free college results in only a modest shift, with the share of Humanities and Social Sciences (HSS) graduates increasing by just 1.4 percentage points, while the shares of Science and Engineering (S&E) and Business and Economics (B&E) graduates experience a slight decline.

5.2 Major-specific Subsidies

Given the finding that uniform changes in higher education subsidies have little effect on the college major composition of graduates, I next analyze the impact of college major-targeted subsidies. In this analysis, I adjust the subsidy rate for each major individually,

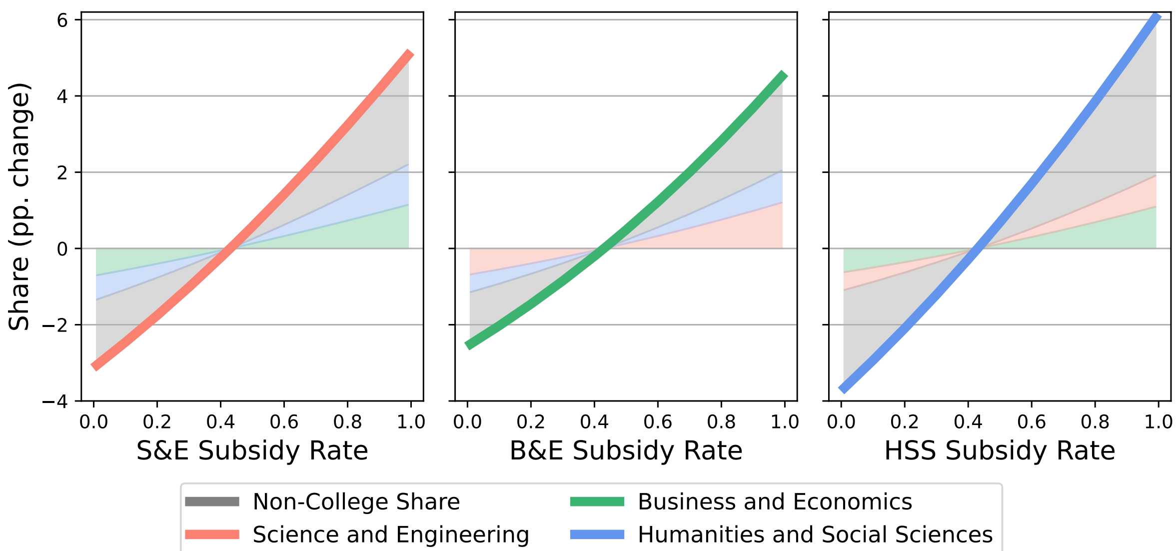
[29]For example, a recent analysis by Avery et al. (2019) found that increased tuition support significantly raised enrollment among students from households earning less than \$60,000. Specifically, they reported that each \$1,000 reduction in tuition at four-year public colleges and universities increased enrollments by an average of 2.0 to 3.0 percentage points.

[30]Another reason might be that emprical studies make the distinction between enrollment and completetion while in my model enrollment and completion means the same given abstraction of college drop out.

while keeping the subsidy rates for other majors fixed at their benchmark levels, to assess how these targeted subsidies influence the educational composition of the labor force.

Major composition of college graduates Figure 7 presents the impact of targeted subsidies, with each subplot highlighting the changes in enrollment for the subsidized major. The shaded areas in each subfigure show where students are switching from or to in response to the subsidy changes.

Figure 7: Targeted Subsidies and College Major Choice



Notes: This figure illustrates the effect of targeted subsidies on the educational composition of the labor force. Each subplot examines the impact of adjusting the subsidy rate for a specific major while keeping the subsidies for other majors fixed at the benchmark level of 0.43. For example, in the first subplot from the left, the line shows the change in the share of Science and Engineering (S&E) graduates as the subsidy rate (τ) varies. The shaded areas indicate the source of changes, showing the education options from which individuals are shifting into or out of the S&E major. The y-axis is percentage point deviation of the corresponding major's share from the benchmark. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively.

The first key finding is that the responsiveness of enrollment in targeted majors to changes in subsidy rates varies significantly among different college majors. Humanities and Social Sciences (HSS) exhibit the highest sensitivity to subsidy adjustments. Increasing the subsidy rate for HSS to 100 percent results in a 6 percentage point increase in the share of graduates from this major. Conversely, Business and Economics (B&E) majors see an increase of approximately 4 percentage points, while Science and Engineering (S&E) majors experience a 5 percentage point rise.[31]

[31] In other words, with a 100 percent increase in HSS subsidies, 7.6 percent of individuals who would have

Moreover, the profile of individuals attracted by higher subsidies differs across college majors. Subsidies for each major tend to draw in a substantial number of individuals who would not have pursued higher education under the benchmark subsidy rates, as indicated by the gray shaded areas in each figure. This pattern is particularly pronounced for HSS major, highlighting their higher sensitivity to subsidies. Higher responsiveness of individuals who would typically not attend college to HSS subsidies is attributed to the distribution of pre-college abilities. Specifically, inadequate math abilities are a more significant barrier for S&E and B&E majors due to their high math intensity. [32] Therefore, individuals with lower math abilities are less likely to enroll in S&E and B&E majors even with substantial subsidies, making them more responsive to subsidies for HSS majors, which are less math-intensive.

The transition between S&E and B&E majors is more common compared to the transition from HSS to these majors, as indicated by the blue shaded areas in the first and second figures. Mathematical abilities again play a crucial role in this pattern because both S&E and B&E majors are math-intensive. Students with sufficient math abilities in one of these majors are likely to have the skills needed for the other, as these fields share more similar math skill contents compared to HSS. Consequently, changes in subsidies can more effectively influence transitions between S&E and B&E majors. In contrast, individuals in HSS majors, who generally have lower math abilities, face greater private costs when attempting to transition to S&E or B&E majors due to their higher math skill content.

Skill composition of workforce Table 7 documents the changes in average post-college skills relative to the benchmark economy resulting from a 50% increase in the subsidy rate for each college major individually, as shown in each column.[33]

When increasing the subsidy rate for S&E, as discussed earlier, a fraction of individuals who would normally prefer other majors or opt not to attend college choose S&E instead. These individuals tend to have the highest comparative advantage in S&E, meaning those with the strongest math abilities from other educational groups are drawn to S&E. Consequently, the average math skills of graduates from other majors drop significantly

otherwise chosen different educational options, including non-college pathways, now prefer HSS majors. In comparison, the corresponding figures for SE and BE majors are 6.1 percent and 5.2 percent, respectively.

[32] In terms of model parameters, the cost of skill shortages is greater for math abilities than for verbal abilities, with calibrated parameters indicating $k_{\text{math}} > k_{\text{verbal}}$. Additionally, S&E and B&E majors require more math abilities compared to HSS, as shown by the higher values of $c^{S\&E}_{\text{math}}$ and $c^{B\&E}_{\text{math}}$ compared to c^{HSS}_{math} (see Table 4).

[33] A 50% increase in subsidy rates is selected to illustrate the effect of targeted subsidies on skill composition. For the effects of different subsidy rates, refer to Table F.2

Table 7: Targeted Subsidies and Skill Composition

	$\tau_{S\&E} = 0.645$	$\tau_{B\&E} = 0.645$	$\tau_{HSS} = 0.645$
Non-college			
math	-0.80	-0.54	-0.41
verbal	-0.63	-0.43	-0.76
social	-0.14	-0.11	-0.10
Science & Engineering			
math	-0.10	-0.01	+0.02
verbal	-0.05	+0.02	-0.06
social	+0.09	+0.03	+0.07
Business & Economics			
math	-0.17	-0.04	+0.02
verbal	-0.11	+0.01	-0.11
social	+0.05	+0.02	+0.07
Humanities and Social Sciences			
math	-0.52	-0.30	+0.45
verbal	-0.10	-0.06	+0.04
social	-0.02	-0.02	+0.15
All			
math	+1.36	+1.02	-0.20
verbal	+0.02	+0.09	+1.79
social	+0.43	-0.07	+0.36

Notes: This table shows the percentage change in average skills across each skill dimension for each educational group relative to the benchmark. In each column, the subsidy rate (τ) for the corresponding college major is increased by 50% (from 0.43 to 0.645), while the subsidy rates for other majors are kept fixed at the benchmark level.

because they lose individuals with the strongest math abilities. This effect is particularly pronounced among those who would otherwise choose the non-college option.

Although S&E attracts individuals with high math abilities from other majors, these individuals' math skills are still lower compared to those who originally chose S&E in the benchmark. As a result, the average post-college math skill of S&E graduates is lower than the benchmark.

Increasing the subsidy rate for S&E by 50% results in a 1.36% increase in the average math skills in the economy, attributable to changes in the student composition. In other words, while individuals switching to S&E after the subsidy increase have lower math abilities compared to those who initially chose S&E under the benchmark rates, they still accumulate more math human capital than they would have in their previous non-S&E education option, due to the high math skill content of S&E majors.

Since initial math and verbal abilities are highly correlated, increasing the subsidy for S&E not only attracts individuals with higher math abilities but also those with higher verbal abilities. Consequently, as more students with higher math and verbal abilites are drawn to S&E, the average verbal human capital among graduates from other majors and non-college individuals decreases. However, the overall verbal human capital in the economy remains relatively unchanged. This is because individuals who transition from the non-college pool to S&E tend to accumulate more verbal skills, while those switching from other college majors to S&E accumulate fewer verbal skills due to the lower verbal skill content in S&E majors. As a result, these opposing effects balance each other out, maintaining a stable level of verbal human capital in the economy.

The same mechanism operates when the subsidy for B&E is increased, but with a notable difference. Although larger subsidies for B&E can attract individuals with higher math abilities from the non-college and HSS pools, they tend to draw only average individuals from the S&E major. This occurs because S&E has a comparative advantage in math skill accumulation. Consequently, the overall math human capital in the economy increases by 1.02 %, which is lower than the 1.36 % increase observed when subsidizing S&E.

Subsidizing HSS benefits verbal abilities, leading to a 1.79 % increase, while causing a moderate 0.20 % decrease in math abilities. The average human capital of non-college individuals declines across all skill dimensions compared to the benchmark, as the HSS major attracts individuals with higher abilities from the non-college pool. In both S&E and B&E, the average math human capital slightly increases by 0.02 percent, while the average verbal human capital decreases by 0.06 % and 0.011 %, respectively. The individuals transitioning from S&E and B&E to HSS can be characterized by higher-than-average verbal abilities and slightly lower-than-average math abilities. Although these marginal individuals have slightly lower math human capital compared to the benchmark case, their math abilities are still significantly higher than those of the benchmark HSS graduates. This results in an overall increase of 0.45 % in the average math ability of HSS graduates.

The changes in social human capital resulting from increases in subsidy rates are moderate. This is because the social skill component of these majors is relatively lower and similar across the board, indicating that social skill accumulation in college majors is less pronounced and more uniform compared to other skills. Consequently, the impact of subsidies on social human capital is less significant.

Table 8: Targeted Subsidies and Output

	$\tau_{S\&E} = 0.645$	$\tau_{B\&E} = 0.645$	$\tau_{HSS} = 0.645$
Total Output(Δ %)	0.48	0.19	0.37
Shares (Δ pp.)			
S&E	1.82	-0.42	-0.38
B&E	-0.41	1.57	-0.29
HSS	-0.38	-0.29	2.17
Non-college	-1.03	-0.86	-1.50
Average Skills (Δ %)			
Math	1.36	1.02	-0.20
Verbal	0.02	0.09	1.79
Social	0.43	-0.07	0.36

Notes: This table shows the percentage change in output, percentage point changes in education groups, and percentage change in average skills across each skill dimension. All numbers are relative to the benchmark. In each column, the subsidy rate (τ) for the corresponding college major is increased by 50% (from 0.43 to 0.645), while the subsidy rates for other majors are kept fixed at the benchmark level. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively.

Match output I apply the same exercise described above to evaluate the effects of major-targeted subsidies on match output.[34] The results are shown in Table 8. Increasing the subsidy rate for each major at the same rate leads to differential increases in output: a 0.48 % increase for S&E, followed by a 0.38 % increase for HSS, and a more modest 0.19 % increase for B&E.

Subsidizing S&E and HSS results in similar output gains, while BE subsidies yield significantly lower gains. There are two main reasons for this. First, HSS majors attract a substantial number of individuals who would not have otherwise pursued higher education. Specifically, increasing HSS subsidies raises HSS enrollment by 2.17 percentage points, with 1.5 percentage points (approximately 70 percent) coming from the non-college pool. The marginal impact of individuals from the non-college pool on the economy's skill composition is greater than that of individuals transitioning from other majors, as non-college individuals contribute more to human capital accumulation due to the absence of previous college-level skill development. Consequently, subsidizing HSS leads to a 1.79 % increase in verbal abilities, which surpasses the average skill increase observed from subsidies for any other major.

Second, subsidizing S&E majors attracts fewer individuals from the non-college pool compared to HSS subsidies. Specifically, only 1.03 percentage points of the additional 1.82 percentage points increase in S&E graduates come from non-college backgrounds.

[34] The total match output can be defined as $\int \int y(\boldsymbol{h}, \boldsymbol{r}) g_m(\boldsymbol{h}, \boldsymbol{r}) d\boldsymbol{h} d\boldsymbol{r}$.

Table 9: Targeted Subsidies and Earnings

	$\tau_{S\&E} = 0.645$	$\tau_{B\&E} = 0.645$	$\tau_{HSS} = 0.645$
Earnings (Δ %)	0.48	0.18	0.36
S&E	0.00	0.00	0.00
B&E	-0.03	-0.01	0.01
HSS	-0.12	-0.09	0.13
Non-college	-0.55	-0.39	-0.48
College Premium(Δ pp.)			
S&E	0.96	0.70	0.85
B&E	1.37	1.02	1.30
HSS	1.23	0.87	1.78

Notes: This table shows the percentage change in earnings and percentage point changes in college premium. All numbers are relative to the benchmark. In each column, the subsidy rate (τ) for the corresponding college major is increased by 50% (from 0.43 to 0.645), while the subsidy rates for other majors are kept fixed at the benchmark level. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively. College premium is calculated relative to non-college earnings.

However, S&E subsidies lead to a significant 1.36 percent increase in math human capital, which is more productive compared to verbal skills. This substantial increase in a more productive type of human capital explains the higher output gain associated with S&E subsidies.

Among these effects, when B&E majors are subsidized, they attract fewer individuals from the non-college pool, drawing only 54 percent compared to 70 percent for HSS. Additionally, the increase in the most productive skills is limited for BE subsidies, with only a 1.02 percent rise in math human capital compared to a 1.36 percent increase with S&E subsidies. Consequently, the output gain from subsidizing BE majors is more limited compared to other majors. For a more detailed analysis of output gains, see the discussion in Section 6.2.

Earnings and college premium Table 9 presents the effects of college major-specific subsidies on the earnings of each education group and the college premium for each major. In general, the earnings trends mirror those observed in match output. Specifically, a 50% increase in the subsidy for S&E results in a 0.48% rise in mean earnings. For B&E and HSS, the increases are 0.18% and 0.36%, respectively. In all cases, the non-college group experiences the largest decline in mean earnings. This occurs because subsidies attract higher-ability individuals to the targeted college majors, thereby lowering the average ability level in the non-college group

Subsidies to S&E also attract individuals from other majors who possess higher abilities, as noted earlier. Consequently, the mean earnings in other college majors decrease, although the decline is not as pronounced as in the non-college group. The mean earnings of non-college do not change significantly because, while the individuals switching from other majors have, on average, lower abilities compared to those in S&E at the benchmark, they now develop human capital that is more highly valued in the labor market. These two effects effectively cancel each other out. In contrast, subsidizing HSS leads to a modest increase in the earnings of this group because the individuals drawn from other majors tend to have higher pre-college abilities than the original HSS graduates.

Even though college major-specific subsidies may not directly increase the mean earnings within subsidized majors, the college premiums for all majors rise. This is because the college premium is calculated as the average wage of each college major relative to the non-college group. As subsidies attract higher-ability individuals into college majors, the earnings of the base group of non-college individuals shrinks. Consequently, the college premium for each major increases. Among the college majors, HSS experience the largest increase in its college premium, rising by 1.78% when subsidized.

6 Welfare-Maximizing Subsidies

My findings to date suggest that while uniform subsidies are effective in increasing overall college enrollment, they have limited influence on the distribution of graduates across majors. In contrast, subsidies targeted at specific fields of study can significantly alter the composition of college graduates. The effectiveness of these targeted subsidies varies depending on the major being subsidized. For instance, increasing subsidies for S&E leads to a greater overall output gain compared to B&E and HSS. Conversely, subsidizing HSS proves more effective in attracting individuals who might otherwise not pursue higher education.

Motivated by this, I compute the welfare-maximizing college major-specific subsidy schemes. I define welfare as the ex-ante value of individuals:

$$V(\boldsymbol{a}) = \max_{i \in \mathcal{I}} \{-p^i(1-\tau^i) - d(\boldsymbol{a}, i) + \beta_i V_u(h(\boldsymbol{a}, i))\} \tag{30}$$

where welfare consists of three components: the pecuniary cost of the chosen education option (i.e., monetary cost), the non-pecuniary cost of the chosen education option (i.e., study effort), and the value of entering the labor market with the skill bundle resulting from the chosen education path.[35] The total welfare in the economy is then expressed as $\int V(\boldsymbol{a})g_a(\boldsymbol{a})da$.

Table 10: Welfare Maximization

	Benchmark	Welfare Max.	10% more subsidy	10 % more +Welfare Max.	20% more subsidy	20 % more +Welfare Max.
$\tau_{S\&E}$	0.43	0.78	0.47	0.83	0.50	0.87
$\tau_{B\&E}$	0.43	0.00	0.47	0.00	0.50	0.04
τ_{HSS}	0.43	0.21	0.47	0.25	0.50	0.27
Tot. Cost	100	100	110	110	120	120
Shares (%)						
S&E	11.9	16.2	12.0	16.5	12.2	16.8
B&E	8.9	6.0	9.0	5.9	9.1	6.0
HSS	14.5	12.3	14.8	12.6	15.0	12.7
Non-College	64.7	65.5	64.2	65.0	63.6	64.5
Output (Δ %)	0.00	0.23	0.17	0.40	0.33	0.57
Welfare (Δ %)	0	0.47	0.13	0.73	0.26	0.97
Welf. Gain Cont. (%)						
Pecuniary Cost	0	32.6	129.0	51.7	128.3	64.6
Study Effort	0	-15.5	-40.0	-19.1	-39.2	-23.2
Future Value	0	82.9	11.0	67.3	10.9	58.6

Notes: This table shows the changes in aggregate welfare under various subsidy schemes. The first column presents the benchmark results. The second column shows the results for welfare-maximizing college major-specific subsidies, with the overall subsidy cost fixed at the benchmark level. The third column displays the effects of uniformly increasing subsidies, with the total subsidy cost rising by 10%. The fourth column presents the results for welfare-maximizing subsidy schemes with an additional 10% increase in subsidy expenditure. The last two columns provide the results for 20% increases in subsidy expenditure. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively. $\tau_{\text{S\&E}}$, $\tau_{\text{B\&E}}$, and τ_{HSS} represent the subsidy rates for S&E, B&E, and HSS, respectively.

Welfare-maximizing subsidy scheme First, I compute the welfare-maximizing subsidy rates, ensuring that the total subsidy cost remains fixed at the benchmark level. The results of this exercise are presented in the second column of the Table 10. Introducing differential subsidy rates across college majors leads to a 0.47% increase in welfare, achieved without allocating any additional resources, as the total cost of subsidies is held constant. Under this welfare-maximizing scheme, the subsidy rate for S&E increases significantly to 78%, while the subsidy for HSS is reduced to 21%, and no subsidy is allocated to the B&E major.

The absence of subsidies for B&E is due to the fact that educational choices and labor market outcomes are least sensitive to changes in B&E subsidy rates, as discussed in

[35]This equation corresponds to equation 4, substituted into equation 5 from the model section. I exclude the preference shock since the mean of the distribution remains constant and does not vary across different specifications, thereby not affecting welfare comparisons.

previous sections. In contrast, subsidies for S&E and HSS remain positive, although the rates are adjusted in favor of S&E. This adjustment reflects the tradeoff between subsidizing S&E and HSS: while human capital from S&E is highly valued in the labor market, college enrollment is more responsive to HSS subsidies due to the higher transition rate from HSS to non-college options. Thus, even though welfare maximization requires increasing subsidies for S&E to attract individuals to this major, it does not eliminate subsidies for HSS. This is because removing HSS subsidies could lead too many individuals to opt out of college altogether, which would reduce overall welfare more than the gains from increasing S&E subsidies.

Major composition of college graduates Under this welfare-maximizing subsidy scheme, the share of students majoring in S&E increases from 11.9% to 16.2%, representing an approximately 35% rise in the number of S&E graduates. Most of this increase results from students switching from other majors, as the lower subsidies for those majors lead them to choose S&E instead. Notably, the overall college enrollment rate decreases by only 0.8 percentage points due to the reduced subsidy rates for other majors.

Decomposing welfare gains I decompose the welfare gain into three components, as presented in equation 30. While changes in the future value of human capital and monetary costs contribute positively to welfare, the study effort works in the opposite direction. The gain in monetary cost is a mechanical effect due to the decrease in college enrollment. The increase in value is the most significant component of the welfare gain, contributing 82.9% of the total gain. This is because human capital accumulated in the S&E major is highly valued in the labor market, leading to a rise in the total discounted value for individuals. In contrast, the study effort negatively impacts welfare, contributing -15.5% to the overall gain, as individuals are shifting to more challenging majors.

Increasing the total college subsidy cost Next, I analyze the impact of increasing the total college subsidy cost by 10%, both with fixed subsidy rates and by allowing differential subsidies across college majors. The results are summarized in columns 3 and 4 of Table 10. First, I examine the effects of a uniform increase in subsidy rates. With a 10% increase in the total subsidy expenditure, welfare gains are relatively modest, amounting to just 0.13%. This minimal gain is largely attributable to the additional resources allocated by the uniform subsidy increase. Despite the increase in subsidy expenditure, the college enrollment rate only rises by 0.5 percentage points. Specifically, the non-college share decreases slightly from 64.7% to 64.2%, indicating only a marginal shift in the overall college enrollment. Furthermore, the major composition of college graduates does not change significantly under uniform subsidies.

In contrast, allowing for differential subsidies across college majors with an additional 10% subsidy expenditure results in a more substantial welfare gain of 0.73% relative to the benchmark. Moreover, if the total subsidy cost were increased by 20%, as shown in the last column of Table 10, the welfare gain could rise to approximately 1%. This suggests that college-major-targeted subsidy policies may be even more effective when accompanied by an increase in overall higher education subsidy expenditures.

7 Discussion

7.1 Decomposing Output Gain

In Section 5.2, I discuss the differential output gains resulting from targeted college major subsidies, depending on which major is subsidized. The key insight is that while human capital accumulated in the S&E major is more valued in the labor market—leading to higher output gains—the subsidy for HSS attracts more individuals who would otherwise not attend college. This higher sensitivity is responsible for the significant output gains observed when subsidizing HSS. In this section, I further decompose these output gains.

The match output, defined in equation 7, consists of three components: the characteristics of the job, the human capital of the individual, and the degree of mismatch between these two. The mismatch represents the underqualification of workers relative to the requirements of their jobs. Table 11 shows the contribution of each component to the output gain resulting from increased subsidies for the relevant major as well as welfare-maximizing subsidy scheme described above.

Improvements in human capital are the most significant contributors to output gains across all majors. Specifically, human capital improvements account for 88.9% of the output gain when subsidizing B&E, 83% when subsidizing S&E, and 68.1% when subsidizing HSS. In each case, the primary driver of these gains is the enhancement of human capital, with improvements in math skills contributing to gains for B&E and S&E, and verbal skills driving the gains for HSS.

The contributions of improvements in job characteristics and reductions in skill mismatch are relatively limited for S&E and B&E. However, in the case of HSS, these contributions are more significant. Specifically, improvements in job characteristics account for

Table 11: Decomposing Output Gain

	Welfare Max.	$\tau_{S\&E} = 0.645$	$\tau_{B\&E} = 0.645$	$\tau_{HSS} = 0.645$
Output (Δ % Benchmark)	0.23	0.48	0.19	0.37
Skill Requirements	**11.3**	**11.9**	**3.2**	**19.0**
math	4.8	3.4	4.6	8.3
verbal	-4.4	2	1.8	4.6
social	10.9	6.5	-3.2	6.1
Human Capital	**65.8**	**83.6**	**88.9**	**68.1**
math	87.1	70.7	89.9	-8.3
verbal	-60.8	0.5	4.2	64.8
social	39.5	12.4	-5.2	11.6
Skill Mismatch	**22.8**	**4.5**	**7.9**	**12.9**
math	21.8	1.5	5.1	10.8
verbal	3.1	1.3	4.9	0.6
social	-2.1	1.6	-2.1	1.4

Notes: This table shows the percentage contributions of skill improvement, human capital, and skill mismatch to the output gain resulting from a 50% increase in the subsidy rate for each major, while keeping the other subsidy rates at the benchmark level as well as welfare-maximizing subsidy scheme. The entries in the table are expressed as percentages (%), and the sum of the bold entries in each column should total 100%. There may be rounding errors.

19% of the total output gain, while reductions in skill mismatch contribute nearly 13%.

This difference is due to the profile of individuals attracted to the targeted major: HSS subsidies draw more people from the non-college pool who are likely to experience underqualification-related output losses in the benchmark economy. As a result, the skill mismatch in HSS improves more significantly, reducing the overall underqualification of workers in the economy. Additionally, firms respond by posting jobs with higher requirements, as the likelihood of matching workers who meet job requirements increases.

The welfare-maximizing subsidy scheme results in a 0.23 percent output gain. The largest contributor is the improvement in human capital, accounting for 65.8 percent of the gain. Specifically, improvements in math skills play a crucial role, while a decline in verbal skills works in the opposite direction. This occurs because the welfare-maximizing subsidy scheme directs individuals away from more verbally intensive majors towards Science and Engineering, which are more math-intensive. Additionally, the effects of job improvements and reductions in skill mismatch are significant, contributing 11.3 percent and 22.8 percent, respectively. Nearly one-third of the output gain comes from channels other than human capital improvement, underscoring the importance of general equilibrium effects when evaluating the potential impacts of college major-specific subsidies.

7.2 The Role of Pre-college Abilities

College major subsidies play a crucial role in influencing both the decision to pursue higher education and the choice of college major. However, subsidies alone do not capture the entire picture. For instance, even under a free tuition policy, there is only a 8.4 percentage point increase in the proportion of individuals choosing to pursue college education compared to the benchmark. Furthermore, when only S&E majors are made free while other majors remain at the benchmark subsidy level, there is an additional 4.5 percentage point increase in the proportion of individuals selecting S&E majors.

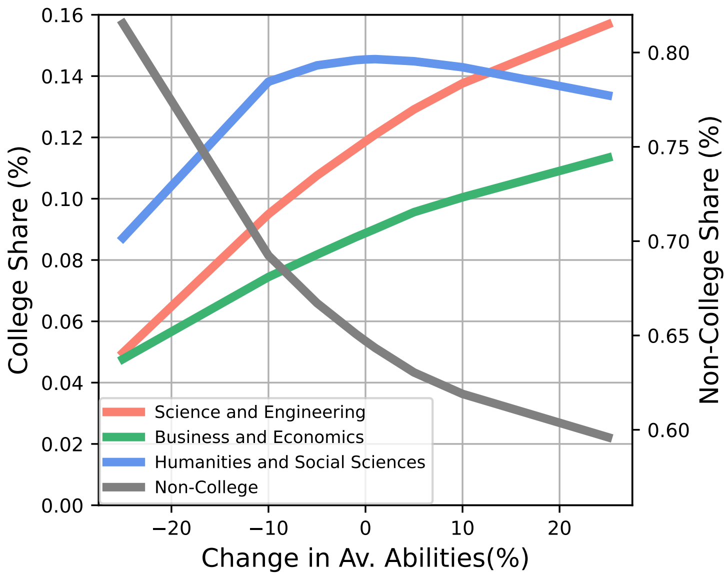
The model emphasizes the critical role of non-pecuniary costs in shaping individuals' educational decisions, particularly in relation to their multidimensional pre-college abilities and preferences. The alignment between an individual's pre-college ability profile and the skill content of their chosen major significantly influences their study effort. For instance, a student with limited algebra skills may struggle in an engineering major, even if they excel in writing. Conversely, a student with exceptional algebra skills but weak verbal communication may find it challenging to succeed in a humanities major. Moreover, the private benefits of different college majors are closely tied to initial abilities. The complementarity between an individual's initial abilities and the skill content required by their chosen major affects the development of human capital for the labor market.

Figure 8 illustrates the impact of initial abilities on education choices. I shift the initial ability distribution by applying various multipliers and display the corresponding education choices in each case. A value of 10 on the x-axis indicates that everyone's pre-college abilities in each dimension are 10 percent higher than their benchmark levels. This exercise highlights that as pre-college abilities increase relative to the benchmark, more individuals opt for S&E and B&E majors. Although many of these individuals would have chosen not to attend college under the benchmark scenario (as evidenced by the significant decrease in the non-college share), a notable finding is the slight decline in the share of HSS graduates. This occurs because individuals with insufficient math abilities who initially chose HSS majors due to the high skill demands of S&E and B&E majors now opt for higher-paying S&E and B&E majors, thanks to their improved math abilities.

7.3 The Role of Skill Mismatch

The model highlights the importance of labor market matching, as underqualification leads to output loss (captured by η), while overqualification results in utility losses (captured by ϕ). The calibrated values of all skill mismatch parameters are positive. To understand the

Figure 8: Pre-college abilities and college major choice



Notes: This figure displays the composition of education groups for different values of initial abilities. The y-axis on the left represents the shares of college majors, while the y-axis on the right represents the share of non-college. The x-axis shows the percentage change in initial abilities across all dimensions. A value of 0 represents the benchmark economy, while a value of 10 represents a scenario where initial abilities across all dimensions for all individuals increase by 10 percent.

role of skill mismatch in model dynamics, I conduct the same analysis without accounting for the cost of skill mismatch—specifically, by setting all skill mismatch parameters to zero. Table 12 presents the results for selected variables.

When skill mismatch is not costly, the share of the non-college group increases by 3 percentage points. The decline in college enrollment primarily stems from a drop in Humanities and Social Sciences (HSS) graduates, falling from 14.5% to 11.6%. One key reason individuals enroll in college is to reduce the likelihood of being underqualified for jobs. By eliminating the phenomenon of skill mismatch, this incentive weakens, leading fewer people to pursue college. The cost of skill mismatch is higher for those attending HSS majors, as discussed in Section 7.1. As a result, without costly skill mismatch, attending HSS becomes less attractive.

College premiums in the model without skill mismatch are significantly lower. In the model that includes skill mismatch, college graduates with higher human capital are less likely to experience output loss due to underqualification. Removing the cost of skill mismatch narrows the wage gap between college and non-college individuals. Furthermore, eliminating skill mismatch costs leads to a 9.71% increase in output, underscoring the

Table 12: The Role of Skill Mismatch

	Mismatch				No Mismatch			
	Benchmark	S&E subs.	B&E subs.	HSS subs.	Benchmark	S&E subs.	B&E subs.	HSS subs.
Share(%)								
S&E	11.9	13.7	11.4	11.5	11.4	13.2	11.0	11.1
B&E	8.9	8.5	10.4	8.6	9.2	8.8	10.8	8.9
HSS	14.5	14.2	14.3	16.7	11.6	11.3	11.3	13.6
Non-col.	64.7	63.7	63.9	63.2	67.7	66.7	66.9	66.4
Col. Prem.(%)								
S&E	76.3	77.2	77.0	77.1	40.5	40.6	40.6	40.6
B&E	61.0	61.8	61.6	61.7	32.3	32.4	32.4	32.4
HSS	53.0	53.6	53.4	53.9	24.5	24.5	24.5	24.7
Output	100.00	100.48	100.19	100.37	109.71	110.05	109.87	109.87
Col. Prem.(Δ %)								
S&E	0.0	1.3	0.9	1.1	0.0	0.4	0.4	0.4
B&E	0.0	1.4	1.0	1.3	0.0	0.4	0.4	0.4
HSS	0.0	1.2	0.9	1.8	0.0	0.1	0.2	0.7
Output (Δ %)	0.00	0.48	0.19	0.37	0.00	0.31	0.15	0.15

Notes: This table compares selected moments for economies with and without skill mismatch. The economy with mismatch corresponds to the model discussed so far. The economy without mismatch uses the same parametrization, except that all skill mismatch parameters are set to zero, i.e., $\eta_{math} = \eta_{verbal} = \eta_{social} = \phi_{math} = \phi_{verbal} = \phi_{social} = 0$. Each column shows the results when a specific major is subsidized by 50% more. For example, *S&E subs.* implies that $\tau_{S\&E} = 0.645$, while the subsidies for other majors remain at their benchmark values. S&E, B&E, and HSS refer to Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively

importance of labor market matching in determining output and wages.

In the model without mismatch costs, college major-specific subsidies are less effective in altering college premiums and output. This is because improvements in skill mismatch serve as an important channel through which subsidies impact output by increasing human capital in specific dimensions.

8 Conclusion

In this paper, I examine the aggregate effects of higher major-targeted higher education subsidies by considering college major heterogeneity. I build an equilibrium labor market search model with two-sided multidimensional heterogeneity and endogenous college major decisions by workers. I use NLSY79 and O*NET datasets to discipline model parameters. The model can succesfully capture the college major decisions of individuals conditional on their multidimensional abilities and observed differences between graduates of different college majors.

Several key findings are presented. First, the model shows that college majors differ in their patterns of skill accumulation. Science and Engineering (S&E) is math-intensive, requiring strong pre-college math abilities and leading to significant math human capital gains. Humanities and Social Sciences (HSS) are verbal-intensive, while Business and Economics (B&E) falls between the two. The model also highlights differences across ability types: social skills develop more slowly and uniformly across all majors, playing a smaller role in major choice but proving crucial in the labor market, particularly for job matching. Math skills, offering higher returns than verbal skills, help explain the superior labor market outcomes for S&E graduates.

Second, while uniform changes in college subsidy rates have minimal effect on the composition of college graduates by major, targeted subsidies can effectively shift individuals toward specific majors. Targeting HSS is particularly effective because a larger proportion of individuals who would otherwise choose non-college paths respond to higher subsidies in HSS majors.

Third, the output gains from targeted subsidies vary by major. S&E and HSS yield higher gains compared to B&E, but for different reasons. In S&E, individuals acquire more highly valued math skills, driving the increase in output. In HSS, the gains come from attracting more individuals from the non-college pool, leading to an overall increase in human capital and, consequently, higher output.

Lastly, the results reveal that allowing college major-specific subsidy rates, without increasing the total subsidy cost, leads to a 0.5% welfare gain. The welfare-maximizing subsidy scheme results in a 4.3 percentage point increase in S&E graduates. It is important to note that this welfare gain is achieved without the need for additional resources and is solely the result of allowing differential subsidy rates across college majors.

References

Abbott, B., G. Gallipoli, C. Meghir, and G. L. Violante (2019). Education policy and intergenerational transfers in equilibrium. *Journal of Political Economy 127*(6), 2569 – 2624.

Acemoglu, D. and R. Shimer (1999). Holdups and efficiency with search frictions. *40*(4), 827–849.

Ahn, T., P. Arcidiacono, A. Hopson, and J. R. Thomas (2019, December). Equilibrium Grade Inflation with Implications for Female Interest in STEM Majors. NBER Working Papers 26556, National Bureau of Economic Research, Inc.

Altonji, J. G., P. Arcidiacono, and A. Maurel (2015, October). The Analysis of Field Choice in College and Graduate School: Determinants and Wage Effects. NBER Working Papers 21655, National Bureau of Economic Research, Inc.

Altonji, J. G., E. Blom, and C. Meghir (2012). Heterogeneity in human capital investments: High school curriculum, college major, and careers. *4*(1), 185–223.

Altonji, J. G., L. B. Kahn, and J. D. Speer. Trends in earnings differentials across college majors and the changing task composition of jobs. *104*(5), 387–393.

Arcidiacono, P. (2004). Ability sorting and the returns to college major. *121*(1), 343–375.

Arcidiacono, P., V. J. Hotz, and S. Kang (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *166*(1), 3–16.

Avery, C., J. Howell, M. Pender, and B. Sacerdote (2019). Policies and payoffs to addressing america's college graduation deficit. *Policy Brief*.

Ben-Porath, Y. (1967). The production of human capital and the life cycle of earnings. *Journal of Political Economy 75*(4), 352–365.

Bureau of Labor Statistics (2023). Occupational employment and wage statistics. [2023].

Charlot, O. and B. Decreuse (2005, April). Self-selection in education with matching frictions. *Labour Economics 12*(2), 251–267.

Cohen, P. (2016, February 22). A rising call to promote stem education and cut liberal arts funding. *The New York Times*.

College Board (2022). *Trends in College Pricing and Student Aid 2022*. New York.

College Board (2023). *Trends in College Pricing and Student Aid 2023*. New York.

Deming, D. and S. Dynarski (2009, September). Into College, Out of Poverty? Policies to Increase the Postsecondary Attainment of the Poor. NBER Working Papers 15387, National Bureau of Economic Research, Inc.

Dynarski, S., L. Page, and J. Scott-Clayton (2023). Chapter 4 - college costs, financial aid, and student decisions. Volume 7 of *Handbook of the Economics of Education*, pp. 227–285. Elsevier.

Fredriksson, P., L. Hensvik, and O. N. Skans (2018, November). Mismatch of Talent: Evidence on Match Quality, Entry Wages, and Job Mobility. *American Economic Review 108*(11), 3303–3338.

Frey, M. C. and D. K. Detterman (2004). Scholastic assessment or g? the relationship between the scholastic assessment test and general cognitive ability. *Psychological Science 15*(6), 373–378.

Guvenen, F., B. Kuruscu, S. Tanaka, and D. Wiczer (2020, January). Multidimensional skill mismatch. *American Economic Journal: Macroeconomics 12*(1), 210–44.

Hagedorn, M., T. H. Law, and I. Manovskii (2012). Identifying equilibrium models of labor market sorting. NBER Working Papers 18661, National Bureau of Economic Research, Inc.

Heckman, J., L. Lochner, and C. Taber (1998). General-equilibrium treatment effects: A study of tuition policy. *American Economic Review 88*(2), 381–86.

Heckman, J. J., T. Jagelka, and T. D. Kautz (2019, November). Some Contributions of Economics to the Study of Personality. NBER Working Papers 26459, National Bureau of Economic Research, Inc.

Hosios, A. (1990). On the efficiency of matching and related models of search and unemployment. *Review of Economic Studies 57*(2), 279–298.

Huang, J. E. and X. Qiu (2021). Precautionary mismatch. *Working Paper*.

Lee, D. (2005, February). An Estimable Dynamic General Equilibrium Model Of Work, Schooling, And Occupational Choice. *International Economic Review 46*(1), 1–34.

Lindenlaub, I. and F. Postel-Vinay (2023). Multi-dimensional sorting under random search. pp. 725362.

Lise, J. and F. Postel-Vinay (2020, August). Multidimensional skills, sorting, and human capital accumulation. *American Economic Review 110*(8), 2328–76.

National Center for Education Statistics (2022). Table 307.10: Number and percentage of undergraduate students receiving financial aid, by type and source of aid: Selected years, 1999-2000 through 2020-21. `https://nces.ed.gov/programs/digest/d22/tables/dt22_307.10.asp`. Accessed: 2024-09-26.

Patnaik, A., M. J. Wiswall, and B. Zafar (2020, August). College Majors. NBER Working Papers 27645, National Bureau of Economic Research, Inc.

Perry, A., S. Wiederhold, and D. Ackermann-Piek (2014). How can skill mismatch be measured? new approaches with piaac. *Methods, Data, Analyses 8*(2), 137–174.

Rosenberg, M. (1965). *Society and the Adolescent Self-Image*. Princeton, NJ: Princeton University Press.

Rotter, J. B. (1966). Generalized expectancies for internal versus external control of reinforcement. *Psychological Monographs 80*, 1–28.

Sanders, C. (2012). Skill Uncertainty, Skill Accumulation, and Occupational Choice. 2012 Meeting Papers 633, Society for Economic Dynamics.

Shapiro, R. and I. Yoder (2021). The impact of a national program of free tuition at public community colleges and free tuition for most students at public four-year colleges and universities on college enrollments, graduations, and the economy.

Shephard, A. and M. Sidibe (2019, July). Schooling Investment, Mismatch,and Wage Inequality. PIER Working Paper Archive 19-013, Penn Institute for Economic Research, Department of Economics, University of Pennsylvania.

Shimer, R. (2005). The cyclical behavior of equilibrium unemployment and vacancies. *95*(1), 25–49.

Shimer, R. and L. Smith (2000). Assortative matching and search. *68*(2), 343–369.

U.S. Bureau of Economic Analysis (2024, January 23). Government current expenditures: Education: Higher [g160311a027nbea]. Retrieved from FRED, Federal Reserve Bank of St. Louis. https://fred.stlouisfed.org/series/G160311A027NBEA.

Valletta, R. G. (2018, July). Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both? In *Education, Skills, and Technical Change:*

Implications for Future US GDP Growth, NBER Chapters, pp. 313–342. National Bureau of Economic Research, Inc.

Webber, D. (2014). The lifetime earnings premia of different majors: Correcting for selection based on cognitive, noncognitive, and unobserved factors. *Labour Economics 28*(C), 14–23.

Wiswall, M. and B. Zafar (2015). Determinants of College Major Choice: Identification using an Information Experiment. *Review of Economic Studies 82*(2), 791–824.

Xia, X. (2016). Forming wage expectations through learning: Evidence from college major choices. *Journal of Economic Behavior & Organization 132*(PA), 176–196.

A An Overview of Higher Education Subsidies

A.1 Price of College

I calculate the cost of attending a 4-year college between 2003 and 2023 using data provided by College Board (2023), focusing only on 4-year institutions to maintain consistency with my quantitative framework. Private for-profit institutions are excluded from this analysis, as their enrollment share was relatively low, around 5

Table A.1 shows the average published prices for private non-profit and public colleges during selected years. To align with my quantitative framework, the "published price" refers to the sticker price charged by institutions. FT represents tuition and fees, while FTHF includes housing and food costs for students residing on campus. The cost of attendance (COA) further incorporates allowances for books, course materials, supplies, and transportation. In this study, I refer to COA as the "college price" because it encompasses all relevant expenses faced by an average student during their college years.

The average price of college is calculated by weighting private non-profit and public 4-year college prices according to their full-time equivalent (FTE) enrollment. FTE enrollment is defined as the number of full-time students, plus the full-time equivalent of part-time students.

The average annual cost of attending a 4-year college between 2003 and 2023 was \$39,075, amounting to \$156,300 for a four-year period. More than half of this cost is attributed to tuition and fees. In 2023, tuition and fees made up 53% of the total cost of attendance, while housing and food constituted 34%, with other expenses accounting for the remaining 12%.

During the same period, the average quarterly wage earnings were \$16,030. (U.S. Bureau of Economic Analysis (2024) Therefore, in my model, I set the price of college at 9.75 times the average wage, i.e., $p = 9.75 * \varpi$, where \$156,300 is divided by the average wage of \$16,030.

A.2 Types of Financial Aid

Table A.2 provides an overview of the budgets for current financial aid programs in the U.S. for selected years. The total amount of federal, state, institutional, and other financial aid sums to \$189.4 billion in 2023. Below is a brief description of selected programs. For a more detailed discussion, see Dynarski et al. (2023).

Table A.1: Published Cost of College

	07-08	12-13	17-18	22-23
Private Non-Profit				
Published TF	$34,540	$38,620	$43,310	$41,740
Published TFHF	$47,180	$52,550	$58,640	$56,400
Published COA	$51,710	$57,570	$63,570	$60,720
Public				
Published TF	$9,130	$11,520	$12,450	$11,480
Published TFHF	$20,000	$23,740	$25,940	$24,350
Published COA	$25,530	$29,600	$31,580	$29,250
Average				
Published TF	$17,595	$20,423	$22,318	$21,223
Published TFHF	$29,055	$33,204	$36,396	$34,669
Published COA	$34,252	$38,788	$41,809	$39,383
FTE Enrollment				
Private Non-Profit	2,993,901	3,309,242	3,435,813	3,465,781
Public	5,992,611	6,764,184	7,309,343	7,298,227

Source: The source for FTE enrollment counts is National Center for Education Statistics (2022), available at https://nces.ed.gov/programs/digest/d22/tables/dt22-307.10.asp. The source for cost of college is College Board (2023) https://research.collegeboard.org/trends/college-pricing *Notes:* This table presents the average annual published cost of a 4-year college in the U.S. between 2007-2008 and 2022-2023 for every five-year period. TF stands for Tuition and Fees, TFHF stands for Tuition, Fees, Housing, and Food, and COA represents the Cost of Attendance. FTE stands for Full-Time Enrollment. All prices are 2023 prices.

Federal aid programs: The federal government plays a significant role in subsidizing higher education through grants, loans, and tax benefits, providing a total of $96.5 billion in 2023—more than 50 percent of all financial aid. The federal share of financial aid has fluctuated between 50 and 75 percent over the years. Federal financial aid is distributed through various grant programs, loans, and tax benefits. Grants accounted for 38 percent of total federal aid in 2023, totaling over $37 billion, while loans made up 50 percent, and tax benefits contributed the remaining 12 percent.

Table A.2: Financial Aid by Source

	06-07	11-12	16-17	21-22
Federal aid	$92.31	$174.10	$135.29	$96.52
Total Federal Grants	$24.07	$56.33	$45.99	$37.17
Pell Grants	$18.61	$43.68	$32.79	$27.93
FSEOG	$1.12	$0.96	$0.89	$0.94
Veterans and Military	$3.61	$11.69	$12.31	$8.30
Federal Loans	$57.87	$92.76	$70.48	$47.07
Perkins Loans	$1.81	$0.98	$0.84	$0.00
Subsidized Stafford	$24.45	$37.70	$26.40	$16.96
Unsubsidized Stafford	$19.80	$39.67	$27.91	$18.83
ParentPLUS	$11.80	$14.41	$15.32	$11.28
Federal Work-Study	$1.25	$1.13	$1.07	$1.10
Education Tax Benefits	$9.11	$23.88	$17.75	$11.18
State Grants	$10.81	$12.04	$13.16	$13.75
Institutional Grants	$29.43	$42.60	$56.93	$63.14
Private Grants	$9.70	$12.37	$14.22	$12.81
Total Financial Aid	$142.25	$241.11	$219.61	$186.22

Source: College Board (2023) https://research.collegeboard.org/trends/college-pricing
Notes: This table presents financial aid awarded to undergraduate students by source for selected years from 2006 to 2022. All amounts are in billions of dollars, adjusted to 2022 prices.

The primary federal aid programs include:

- **Pell Grants:** Pell Grants are need-based awards for undergraduate students from low-income families that do not require repayment. They are a cornerstone of federal grant aid, accounting for over 70 percent of such funding in 2023. The amount awarded is based on the student's financial need, cost of attendance, and enrollment status (full-time or part-time).

- **Federal Supplemental Educational Opportunity Grant (FSEOG):** The FSEOG is an additional need-based grant program aimed at students with the highest financial need. Unlike Pell Grants, FSEOG funding is limited, and not all eligible students receive it, as awards depend on availability at each participating institution.

- **Perkins Loans:** Perkins Loans were designed for students with exceptional financial need, offering a fixed interest rate of 5%. These subsidized loans covered interest while students were enrolled and were administered directly by participating schools. The program was discontinued in 2017, and no new loans have been issued since.
- **Subsidized Stafford Loans:** Subsidized Stafford Loans are available to undergraduate students with demonstrated financial need. The government pays the interest while students are in school, during a six-month grace period, and during deferment. This makes them an appealing option for those with limited financial resources. While they have lower borrowing limits than unsubsidized loans, they come with fixed, low-interest rates determined by the federal government each year.
- **Unsubsidized Stafford Loans:** Unsubsidized Stafford Loans are available to both undergraduate and graduate students, regardless of financial need. Interest begins accruing immediately upon disbursement, and borrowers are responsible for paying it throughout their time in school. While these loans have higher borrowing limits than subsidized loans, the early interest accumulation can lead to larger balances upon graduation. Nevertheless, they remain popular due to their relatively low fixed interest rates and flexible repayment options compared to private loans.
- **ParentPLUS Loans:** ParentPLUS Loans are federal loans for parents of dependent undergraduate students, allowing them to borrow up to the full cost of attendance minus any financial aid received. Unlike Stafford Loans, these require a credit check but offer flexible repayment plans and fixed interest rates. While parents can defer payments while the student is in school, interest accrues during this period. ParentPLUS Loans can help cover funding gaps but come with higher interest rates than other federal loans, necessitating careful consideration of repayment obligations.
- **Education Tax Benefits:** Education tax benefits in the U.S. help alleviate the cost of higher education through tax relief. The American Opportunity Tax Credit (AOTC) provides up to $2,500 per student for the first four years of college, with partial refunds available. The Lifetime Learning Credit (LLC) offers up to $2,000 annually for education at any stage. Additionally, the Tuition and Fees Deduction allows for a deduction of up to $4,000 from taxable income. These benefits lower education costs and promote investment in higher education.

State Grants State grants constitute around 8 percent of financial aid for college students. In the U.S., state grants provide financial assistance to students, typically based on need or merit, to support their college education. Each state administers its own grant programs, often prioritizing residents attending in-state public colleges, although some grants are also available for private institutions. The largest state programs, such as Cal Grants in California and the Tuition Assistance Program (TAP) in New York, focus on helping low- and middle-income families cover tuition costs. Eligibility criteria and award amounts vary by state, but these grants serve as a crucial supplement to federal aid for many students.

Institutional Grants Institutional grants are financial aid awards provided directly by colleges or universities to help students cover tuition and other educational expenses. These grants are typically funded by the institution's own resources, such as endowments or donations, and are often awarded based on need, merit, or a combination of both. Unlike loans, institutional grants do not need to be repaid. Many private universities and selective public institutions use institutional grants as part of their financial aid packages to attract talented students and ensure access for low-income students. These grants play a significant role in reducing the net cost of attending college, accounting for nearly one-third of all financial aid in 2023.

A.3 College Subsidy Rate

Table A.3 illustrates the financial aid per student, calculated based on the Full-Time Equivalent (FTE) enrollment figures.[36] In the 2022-2023 academic year, the average financial aid awarded per student amounted to $15,480, with approximately 70 percent of this aid originating from federal, state, and institutional grants.

In 2023, the ratio of average financial aid to the average Tuition and Fees (TF) at colleges was 73 percent when federal loans were included. In contrast, when considering only grants—excluding federal loans—this ratio dropped to 55 percent. When examining the Cost of Attendance (COA), which encompasses not only tuition and fees but also housing, food, books, and supplies, the ratios were significantly lower. Without accounting for loans, the aid-to-COA ratio was 30 percent, whereas, when federal loans were included, the ratio rose to 39 percent.

I use the ratio of total financial aid per student to the average published cost of attendance (COA) to parameterize the college subsidy rate τ in my model. For the sampling

[36]FTE enrollment is defined as the number of full-time students, plus the full-time equivalent of part-time students.

Table A.3: College Subsidies per Student

	07-08	12-13	17-18	22-23
Fianncial aid per Student	\$12,820	\$17,620	\$17,410	\$15,480
Grant Aid	\$6,680	\$9,420	\$10,730	\$10,680
Federal Loans	\$5,260	\$6,510	\$5,330	\$3,860
Other	\$880	\$1,690	\$1,350	\$940
Average TF	\$17,595	\$20,423	\$22,318	\$21,223
Average COA	\$34,252	\$38,788	\$41,809	\$39,383
Subsidy Rate (for COA)				
with federal loans	0.37	0.45	0.42	0.39
without federal loans	0.22	0.29	0.29	0.30
Subsidy Rate (For TF)				
with federal loans	0.73	0.86	0.78	0.73
without federal loans	0.43	0.54	0.54	0.55

Source: College Board (2023) https://research.collegeboard.org/trends/college-pricing
Notes: This table shows financial aid per Full-time-equivalent enrollment. TF stands for Tuition and Fees, COA represents the Cost of Attendance. All prices are 2023 prices.

period from 2003 to 2023, the subsidy rate—defined as the ratio of financial aid per student to the average published cost of attending a four-year college—is 43 percent. Consequently, I set $\tau = 0.43$.

B Model Derivations

The value function of an unemployed individual with human capital $\mathbf{h}$ is recursively given in equation:

$$V_u(\mathbf{h}) = b\overline{\omega}+\beta(1\text{-}\rho)\left[(1\text{-}q_u(\theta))\,V_u(\mathbf{h}) + q_u(\theta)\int\limits_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}V_e(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r} + q_u(\theta)V_u(\mathbf{h})\int\limits_{A_w^c(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}\mathrm{d}\mathbf{r}\right]$$

we can rewrite it as:

$$\begin{aligned}
V_u(\mathbf{h}) &= b\overline{\omega} + \beta(1\text{-}\rho)V_u(\mathbf{h}) - \beta(1\text{-}\rho)q_u(\theta)V_u(\mathbf{h}) + \beta(1\text{-}\rho)q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}V_e(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r} \\
&\quad + \beta(1\text{-}\rho)q_u(\theta)V_u(\mathbf{h})\int_{A_w^c(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}\mathrm{d}\mathbf{r} \\
&= b\overline{\omega} + \beta(1\text{-}\rho)V_u(\mathbf{h}) - \beta(1\text{-}\rho)q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}V_u(\mathbf{h}) - \beta(1\text{-}\rho)q_u(\theta)\int_{A_w^c(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}V_u(\mathbf{h}) \\
&\quad + \beta(1\text{-}\rho)q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}V_e(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r} + \beta(1\text{-}\rho)q_u(\theta)V_u(\mathbf{h})\int_{A_w^c(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}\mathrm{d}\mathbf{r} \\
&= b\overline{\omega} + \beta(1\text{-}\rho)V_u(\mathbf{h}) + \beta(1\text{-}\rho)q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}[V_e(\mathbf{h},\mathbf{r}) - V_u(\mathbf{h})]\mathrm{d}\mathbf{r}
\end{aligned}$$

by using nash bargaining result in equation 12

$$V_u(\mathbf{h}) = b\overline{\omega} + \beta(1\text{-}\rho)\left[V_u(\mathbf{h}) + \alpha q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}\right] \tag{B.1}$$

Recall that the value of an employed worker with human capital $\boldsymbol{h}$ at a job with requirements $\boldsymbol{r}$ is the following:

$$V_e(\mathbf{h},\mathbf{r}) = \omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)[\delta V_u(\mathbf{h}) + (1-\delta)V_e(\mathbf{h},\mathbf{r})]$$

by adding and subtracting $\beta(1\text{-}\rho)V_u(\mathbf{h})$, we can rearrange the equation:

$$\begin{aligned}
V_e(\mathbf{h},\mathbf{r}) &= \omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)\delta V_u(\mathbf{h}) + \beta(1\text{-}\rho)(1-\delta)V_e(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)V_u(\mathbf{h}) - \beta(1\text{-}\rho)V_u(\mathbf{h}) \\
&= \omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)V_u(\mathbf{h}) - \beta(1\text{-}\rho)(1-\delta)V_u(\mathbf{h}) + \beta(1\text{-}\rho)(1-\delta)V_e(\mathbf{h},\mathbf{r}) \\
&= \omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)V_u(\mathbf{h}) + \beta(1\text{-}\rho)(1-\delta)[V_e(\mathbf{h},\mathbf{r}) - V_u(\mathbf{h})]
\end{aligned}$$

by plugging surplus sharing rule in equation 12:

$$V_e(\mathbf{h},\mathbf{r}) = \omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)[V_u(\mathbf{h}) + \alpha(1-\delta)S(\mathbf{h},\mathbf{r})] \tag{B.2}$$

Similarly, by using surplus sharing rules, we can rearrange the value of vacant jobs $V_v(\mathbf{r})$ and the value of producing jobs, $V_p(\mathbf{h},\mathbf{r})$ as follows:

$$V_v(\mathbf{r}) = -c(g_v(\mathbf{r})) + \beta\left[V_v(\mathbf{r}) + (1-\alpha)q_v(\theta)\int_{A_f(\mathbf{r})}\frac{g_u(\mathbf{h})}{U}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{h}\right] \tag{B.3}$$

$$V_p(\mathbf{h},\mathbf{r}) = y(\mathbf{h},\mathbf{r}) - \omega(\mathbf{h},\mathbf{r}) + \beta[V_v(\mathbf{r}) + (1-\alpha)(1-\delta)(1-\rho)S(\mathbf{h},\mathbf{r})] \tag{B.4}$$

Note that the total surplus of a match is given by $S(\mathbf{h},\mathbf{r}) \equiv V_p(\mathbf{h},\mathbf{r}) - V_v(\mathbf{r}) + V_e(\mathbf{h},\mathbf{r}) - V_u(\mathbf{h})$ in equation 11. After applying the free entry condition $V_v(\mathbf{r}) = 0$, and using equations B.1 through B.4, the surplus function simplifies to:

$$S(\mathbf{h},\mathbf{r}) = y(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\delta)(1\text{-}\rho)S(\mathbf{h},\mathbf{r}) - \left(b + g(\mathbf{h},\mathbf{r}) + \beta\left[\alpha(1\text{-}\rho)q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}\right]\right) \tag{B.5}$$

Wage equation By rearranging equations 12 and 13:

$$V_e(\mathbf{h},\mathbf{r}) - V_u(\mathbf{h}) = \frac{\alpha}{1-\alpha}V_p(\mathbf{h},\mathbf{r}) - V_v(\mathbf{r})$$

Also, by the free entry condition $V_v(\mathbf{r}) = 0$, then:

$$\begin{aligned}
&\omega(\mathbf{h},\mathbf{r}) - g(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)(1\text{-}\delta)\alpha S(\mathbf{h},\mathbf{r}) - b\overline{\omega} - \beta(1\text{-}\rho)\alpha q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r} \\
&= \frac{\alpha}{1-\alpha}\left[y(\mathbf{h},\mathbf{r}) - \omega(\mathbf{h},\mathbf{r}) + \beta(1\text{-}\rho)(1-\alpha)(1-\delta)S(\mathbf{h},\mathbf{r})\right] \\
&\omega(\mathbf{h},\mathbf{r}) = \alpha y(\mathbf{h},\mathbf{r}) + (1-\alpha)[g(\mathbf{h},\mathbf{r}) + b\overline{\omega}] + \beta\alpha(1-\alpha)(1-\rho)q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}
\end{aligned}$$

After plugging free entry condition in equation 25 and rearrange

$$\omega(\mathbf{h},\mathbf{r}) = \alpha y(\mathbf{h},\mathbf{r}) + (1-\alpha)[g(\mathbf{h},\mathbf{r}) + b\overline{\omega}] + \alpha(1-\rho)c(g_v(r))\frac{q_u(\theta)\int_{A_w(\mathbf{h})}\frac{g_v(\mathbf{r})}{V}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}}{q_v(\theta)\int_{A_f(\mathbf{r})}\frac{g_u(\mathbf{h})}{U}S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{h}}$$

$$\omega(\mathbf{h},\mathbf{r}) = \alpha y(\mathbf{h},\mathbf{r}) + (1-\alpha)[g(\mathbf{h},\mathbf{r}) + b\overline{\omega}] + \alpha(1-\rho)c(g_v(r))\frac{\int_{A_w(\mathbf{h})} g_v(\mathbf{r})S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}}{\int_{A_f(\mathbf{r})} g_u(\mathbf{h})S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{h}}$$

Thus, match specific wage is:

$$\omega(\mathbf{h},\mathbf{r}) = \alpha\left(y(\mathbf{h},\mathbf{r}) + (1\text{-}\rho)c(g_v(\mathbf{r}))\left[\frac{\int_{A_w(\mathbf{h})} g_v(\mathbf{r})S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{r}}{\int_{A_f(\mathbf{r})} g_u(\mathbf{h})S(\mathbf{h},\mathbf{r})\mathrm{d}\mathbf{h}}\right]\right) + (1-\alpha)[g(\mathbf{h},\mathbf{r}) + b\overline{\omega}] \quad \text{(B.6)}$$

C The Source of Inefficiency - A Simplified Model

In this section, I present a simplified and stylized version of the quantitative model to showcase a particular source of inefficiency inherent in college major choice within this model environment. The process involves describing the decentralized economy, outlining the corresponding social planner problem, and ultimately identifying the sources of inefficiency in college major choices. Proposition 1 characterizes the inefficiency of college major choice in a decentralized economy by comparing it with the solution of the social planner's problem. Proposition 2 establishes a cross-college major subsidy scheme that restores the efficiency of the decentralized economy.

C.1 Decentralized Economy

This section introduces a one-shot static model that integrates random search and college major choice within a decentralized economic framework. Firms initiate the process by posting vacancies, while workers make decisions regarding one of two college majors: Science (S) and Humanities (H). The non-pecuniary cost associated with selecting a college major, denoted as (c_s, c_h), follows a uniform distribution across $[0,1] \times [0,1]$. Specifically, an individual facing non-pecuniary costs of (c_s, c_h) incurs a cost of c_s if inclined toward the science major and c_h if leaning toward the humanities major. Further, a pecuniary cost of college is introduced as t, with each college major being subsidized by τ_s and τ_h. For the purpose of the benchmark, it is assumed that $\tau_s = \tau_h = \tau$.

The job landscape encompasses two distinct types: Science jobs and Humanities jobs. Notably, firms lack prior knowledge (*ex ante*) of their job type, drawing the type of vacancy from a distribution where both jobs are equally likely[37]. Workers engage in a search process, and production occurs contingent upon a successful match with the appropriate job, resulting in output values denoted as y_s and y_h. The share of science and humanities workers is captured by σ_s and σ_h, adhering to the constraint $\sigma_s + \sigma_h = 1$[38].

The matching technology in this model is expressed by $M(u,v) = u^{\alpha}v^{1-\alpha}$. In the specific scenario of this one-shot model, where everyone begins in an unemployed state ($u = 1$), the matching function simplifies to $M(v) = v^{1-\alpha}$. This simplification offers clarity

[37] For the sake of simplicity in the equations, equally likely probabilities are assumed. It's important to note that these probabilities do not significantly influence the primary outcomes of the analysis in this context. The equal likelihood assumption is made to streamline the mathematical expressions and facilitate a more straightforward analysis without substantially altering the key results.

[38] It's important to note that these σ's represent equilibrium outcomes derived from individual college major choices.

in understanding the initial stages of the matching process. Subsequently, the probabilities associated with worker-firm interactions are defined. Workers' probability of meeting a firm, denoted as $p(v)$, is expressed as $v^{1-\alpha}$, while firms' probability of meeting a worker, represented by $f(v)$, takes the form $v^{-\alpha}$.

The value of filled vacancies for Science (S) and Humanities (H) jobs is denoted as $J_s = y_s - \omega_s$ and $J_h = y_h - \omega_h$, respectively. Additionally, the value of vacancy posting, represented by V, is determined by an entrance cost (κ) and the expected values of filled vacancies:

$$V = -\kappa + f(v)\left[\frac{1}{2}\sigma_s J_s + \frac{1}{2}\sigma_h J_h\right] \tag{C.1}$$

Furthermore, the values associated with workers are described by the equations $E_s = \omega_s$, $E_h = \omega_h$, and $U = 0$. In the Nash bargaining framework, introducing workers' bargaining power (ϕ), the relationship between job values and bargaining power is established as follows: $\omega_s = \phi y_s$ and $\omega_h = \phi y_h$.

Then, the free entry condition $(V = 0)$ implies the following equilibrium condition to solve market v:

$$\frac{2\kappa}{f(v)} = (1-\phi)\big[\sigma_s y_s + \sigma_h y_h\big] \tag{C.2}$$

The decision-making process for college major choices involves evaluating the value of majors and making decisions based on relative costs. The value of choosing a Science major (W_s) and a Humanities major (W_h) is expressed as:

$$W_s = \frac{1}{2}p(v)\phi y_s - c_s - P(1-\tau) \tag{C.3}$$

$$W_h = \frac{1}{2}p(v)\phi y_h - c_h - P(1-\tau) \tag{C.4}$$

Individuals choose the Science major if $W_s > W_h$, and the Humanities major otherwise, leading to the following break-even condition:

$$\frac{1}{2}p(v)\phi(y_s - y_h) = \overline{c} \tag{C.5}$$

where $\overline{c} = c_s - c_h$. The figure illustrates the college major decision by multidimensional cost (c_s, c_h). Notably, only the relative cost of college majors, denoted as $\overline{c}$, matters in this

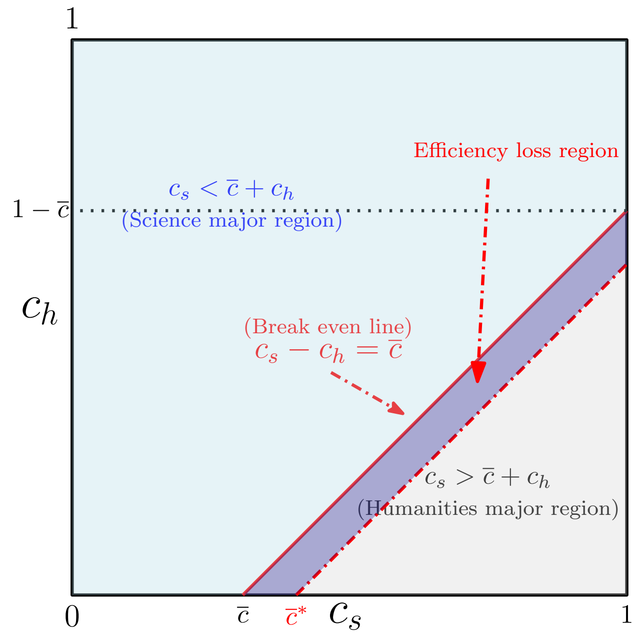


Figure C.1: College major choice

decision-making process [39].

C.1.1 Social Planner's Problem

I formulated the problem of a social planner with the objective of maximizing the total output produced by working science and humanities graduates, net of the total vacancy posting cost and the cost of college education. The latter is characterized by the summation of non-pecuniary costs for all individuals in the economy. The planner determines the number of vacancies, the number of working individuals, and the criteria for college major decisions. The overall decision of the planner is subject to labor market frictions and the distribution of relative costs. The optimization problem is formulated as follows:

[39] Equilibrium college major shares, as depicted in the figure below, are given by $\sigma_h = \frac{(1-\overline{c})^2}{2}$ and $\sigma_s = 1 - \frac{(1-\overline{c})^2}{2}$

$$
\begin{aligned}
&\max_{v,n_s,n_h,\bar{c}} \Bigg\{ y_s n_s + y_h n_h - \kappa v - \int_{\bar{c}}^{1} \int_{0}^{c_s-\bar{c}} c_h \, dc_h \, dc_s \\
&\qquad - \left[\int_{0}^{1-\bar{c}} \int_{0}^{\bar{c}+c_h} c_s \, dc_s \, dc_h + \int_{1-\bar{c}}^{1} \int_{0}^{1} c_s \, dc_s \, dc_h \right] \Bigg\} \\
&\text{s.t.} \quad \begin{cases} n_s = \frac{1}{2}\sigma_s f(v) v \\ n_h = \frac{1}{2}\sigma_h f(v) v \\ \sigma_h = \frac{(1-\bar{c})^2}{2} \\ \sigma_s + \sigma_h = 1 \end{cases}
\end{aligned} \tag{C.6}
$$

The following equations characterize the solutions for v^* and $\bar{c}^*$. The detailed derivations can be found in appendix.

$$
\frac{2\kappa}{f(v^*)} = (1-\alpha) \left[\left(1 - \frac{(1-\bar{c}^*)^2}{2} \right) y_s + \frac{(1-\bar{c}^*)^2}{2} y_h \right] \tag{C.7}
$$

$$
\frac{1}{2} p(v^*)(y_s - y_h) = \bar{c}^* \tag{C.8}
$$

C.1.2 (In)efficiency of College Major Choice

Proposition 1 *Suppose $y_s \neq y_h$. There exist no values $(\phi, \alpha) \in \mathbb{R}^2$ such that the decentralized equilibrium $(v^d, \bar{c}^d)$ is efficient, meaning $(v^d, \bar{c}^d) = (v^*, \bar{c}^*)$ and $v^d > 0$ simultaneously.*

Proof. Suppose, for the sake of contradiction, that there exist $(\phi, \alpha) \in \mathbb{R}^2$ such that $(v^d, \bar{c}^d) = (v^*, \bar{c}^*)$ and $v^d > 0$. Then, from equations (24) and (29), it follows that $\phi = \alpha$. Furthermore, from equations (27) and (30), we have $\phi = 1$. Substituting into equation (24), we obtain $\frac{2\kappa}{f(v^d)} = 0$. Since $\kappa > 0$ by assumption, it must be that $f(v^d) = (v^d)^{-\alpha} \to 0$. This implies $v \to 0$ for $\alpha = 1$, which contradicts the assumption that $v^d > 0$. ■

The idea of the proof is the comparison of the equations solving decentralized economy (equation 17 and 20) the equations solving Social planner's problem

The Proposition 1 asserts that the decentralized equilibrium with college major choice is always inefficient. The standard Hosios condition (Hosios (1990)) ($\alpha = \phi$), where the marginal benefit to the firm from an additional vacancy should be equal to the marginal cost of the worker, alone is insufficient to ensure the efficiency of a decentralized equilibrium. The only scenario in which efficiency is achieved is when $\alpha = \phi = 1$, indicating that workers possess all the bargaining power. Yet, this situation results in a degenerate equilibrium where $v = 0$ [40].

The underlying intuition is that the college major choice made by workers is not entirely internalized when the bargaining power is less than one, as a portion of the major's benefits flows to the firms. This outcome aligns with the well-established concept of a hold-up problem, wherein a party in the market bears the cost while others share in the payoff (Acemoglu and Shimer (1999)).

One fundamental assumption in this simplified model is the equal subsidy across college majors, denoted as $\tau_s = \tau_h = \tau$. This assumption implies that monetary costs do not influence individuals' college major decisions, as they cancel out in the decision-making process. The next proposition establishes the concept that in the case of a deviation from constant subsidies across majors, the efficiency of the decentralized economy can be restored through differential college major subsidies.

Proposition 2 *Suppose $y_s \neq y_h$ and $\alpha = \phi < 1$. There exists $(\tau_s, \tau_h) \in (0,1)^2$ such that $\sigma_s \tau_s + \sigma_h \tau_h = \tau$ and the implied decentralized equilibrium $(v^d, \overline{c}^d)$ is efficient, meaning $(v^d, \overline{c}^d) = (v^*, \overline{c}^*)$.*

Proof. Suppose $\tau_h = \dfrac{\sigma_h^*}{2P} p(v^*)(y_s - y_h)(1-\phi) + \tau$ and $\tau_s = \tau - \dfrac{\sigma_h^* \tau_h}{\sigma_s^*}$. In decentralized economy, the break even condition implies

$$\frac{1}{2} p(v^d)\phi(y_s - y_h) + P(\tau_s - \tau_h) = \overline{c^d}$$

We need to show for given values of τ_s and τ_h, the solution of social planner problem $(v^*, \overline{c}^*)$ also solves decentralized economy, i.e. equations 24 and 27. Also, we need to show that these values satisfy $\sigma_s \tau_s + \sigma_h \tau_h = \tau$.

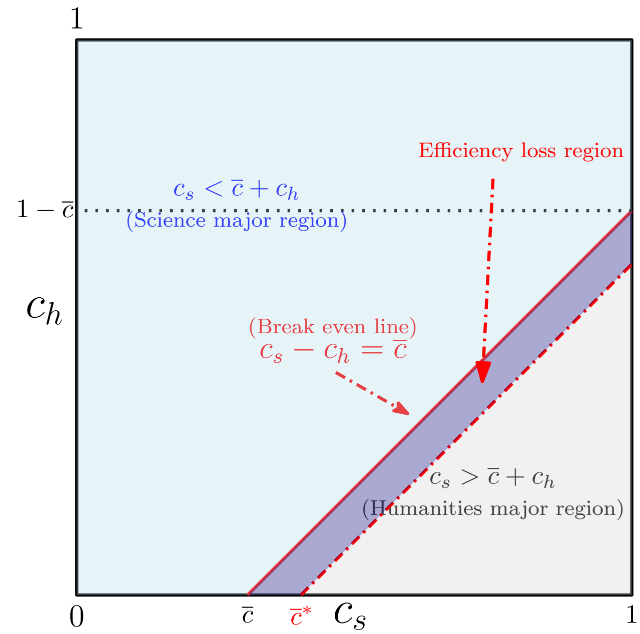
$\alpha = \phi$ by assumption, then $(v^*, \overline{c}^*)$ satisfies equation 24 trivially, since they solve equation 29. Also, the break even condition implies that

$$\frac{1}{2} p(v^d)\phi(y_s - y_h) + P(\tau_s - \tau_h) = \overline{c^d} \Rightarrow$$

$$\frac{1}{2} p(v^d)\phi(y_s - y_h) + \frac{1}{2} p(v^*)\phi(y_s - y_h)(1-\phi) = \overline{c^d} \Rightarrow$$

$$\frac{1}{2} p(v^*)(y_s - y_h) = \overline{c}^*$$

[40] The condition $y_s \neq y_h$ represents any asymmetry between college majors. In the full model, asymmetries across other dimensions of college majors, such as different skill contents, distinct distributions of individual skills, or varying productivities of skills, can play a similar role.



■

Proposition 2 suggests that it is possible to rectify the inefficiency resulting from the hold-up problem by adjusting cross-major subsidies in the opposite direction of private cost asymmetries. Furthermore, this restoration can be achieved while maintaining the overall subsidy cost of the social planner at the benchmark level, i.e., $\sigma_s \tau_s + \sigma_h \tau_h = \tau$.

D Data Details

D.1 Sample Selection

I exclude individuals who are college dropouts or have only completed two years of college, so the sample consists solely of individuals with education levels below high school or those who have earned a four-year college degree. Additionally, those who have served in the military for more than two years during the survey period are excluded from the sample. The sample is also limited to individuals with valid occupation codes, complete ASVAB scores, and accurate college major information. For the labor market analysis, I adopted a sample selection approach similar to that used by Guvenen et al. (2020). I focus on individuals who work over 1,200 hours for at least two consecutive years within the survey period. I exclude those who are only marginally attached to the labor force—specifically, those who have had multiple periods out of the labor force before accumulating at least 10 years of employment after starting their careers. I retain observations if an individual was out of the labor force for just one year or if they worked for over 10 years before their first period of being out of the labor force.

D.2 Test Scores

ASVAB (Armed Services Vocational Aptitude Battery) ASVAB is a comprehensive test used by the U.S. military to assess the mental aptitude and vocational strengths of individuals interested in enlisting. The ASVAB is divided into ten subtests, each designed to evaluate specific skills and knowledge areas. The main subtests are:

- **General Science (GS)**: Knowledge of physical and biological sciences
- **Arithmetic Reasoning (AR)**: Ability to solve arithmetic word problems
- **Word Knowledge (WK)**: Ability to select the correct meaning of words presented in context and to identify the best synonym for a given word
- **Paragraph Comprehension (PC)**: Ability to obtain information from written passages
- **Mathematics Knowledge (MK)**: Knowledge of high school mathematics principles
- **Electronics Information (EI)**: Knowledge of electricity and electronics
- **Auto Information (AI)**: Knowledge of automobile technology

- **Shop Information (SI)**: Knowledge of tools and shop terminology and practices
- **Mechanical Comprehension (MC)**: Knowledge of mechanical and physical principles
- **Assembling Objects (AO)**: Ability to determine how an object will look when its parts are put together

The ASVAB scores have been widely used as a proxy for various cognitive abilities. For instance, Frey and Detterman (2004) report a correlation of 0.82 between SAT scores and an IQ scale derived from ASVAB results, demonstrating the strong relationship between ASVAB scores and alternative measures of cognitive ability.

The Rotter Internal-External Locus of Control Scale is a condensed four-item version of the original 23-item questionnaire derived from Rotter's(Rotter (1966)) 60-item scale developed in 1966. This scale assesses the degree to which individuals perceive they have control over their lives, distinguishing between self-motivation and self-determination (internal control) versus external factors like chance, fate, or luck (external control). Scoring on the scale is oriented towards internal control: higher scores indicate a greater belief in personal control. Respondents are presented with four pairs of statements and are asked to choose which statement better reflects their views. They then rate how closely each selected statement aligns with their opinion—whether "much closer" or "slightly closer." These responses are used to create four-point scales for each paired item, which are averaged to produce a single Rotter Scale score for each individual.

The Rosenberg Self-Esteem Scale was used during the 1980 interviews. This 10-item scale, developed by Rosenberg in 1965 (Rosenberg (1965)), is designed to assess self-approval or disapproval in both adolescents and adults. It is a concise and widely used tool with substantial evidence supporting its validity and reliability. The scale consists of 10 statements related to self-esteem, and respondents indicate their level of agreement by choosing from "strongly agree," "agree," "disagree," or "strongly disagree."

D.3 O*NET Occupation Codes

The NLSY dataset uses the Census 1970 three-digit occupation codes before 2000 and the Census 2000 three-digit codes afterward. I convert all of these to the Census 1990 three-digit codes for consistency. In the NLSY, individuals report their occupation title for up to five jobs. For analysis, I use the most frequently observed occupation title reported.

D.4 College Major Classification

Table D.1: College Major Classification

Majors	Code
Science and Engineering	
Agriculture and Natural Resources	0100
Architecture and Environmental Design	0200
Biological Sciences	0400
Computer and Information Sciences	0700
Engineering	0900
Health Professions	1200
Mathematics	1700
Military Sciences	1800
Physical Sciences	1900
Biological and Physical Sci.	4902
Engineering and Other Disciplines	4904
Business and Economics	
Business and Management	0500
Home Economics	1300
Economics	2204
Humanities and Social Sciences	
Area Studies	0300
Communications	0600
Education	0800
Fine and Applied Arts	1000
Foreign Languages	1100
Law	1400
Letters	1500
Psychology	2000
Public Affairs and Services	2100
Social Sciences	2200
Theology	2300
General Liberal Arts and Sciences	4901
Humanities and Social Sciences	4903

Notes: This table shows the classification of college majors in NLSY79 data into three broad categories, including field of study codes used in the NLSY79 codebook. See the link: `https://www.nlsinfo.org/content/cohorts/nlsy79/other-documentation/codebook-supplement/nlsy79-attachment-4-fields-study`

D.5 Selection into Major: Regression Analysis

I estimate the following regression for the sample of college graduates:

$$CM_{ij} = \alpha + \beta_m \text{mathskill}_j + \beta_v \text{verbalskill}_j + \beta_s \text{socialskill}_j + \delta \text{controls}_j + \epsilon_j \qquad \text{(D.1)}$$

where CM_{ij} is a dummy variable indicating whether individual j attended college major i (1 if yes, 0 if not). The variables mathskill_j, verbalskill_j, and socialskill_j represent the math, verbal, and social pre-college abilities of individual j, respectively. The control variables controls_j include race, sex, and family income.

The results, shown in Table D.2, indicate that ability selection into college majors persists even after accounting for individual abilities and demographic characteristics. The first three columns present regression coefficients without demographic controls, while the next three columns include these controls. The preferred specifications are the last three columns.

The analysis suggests that, given math and social abilities, a one-unit increase in math skills increases the probability of choosing an S&E major over other majors by 58.3 percent. Column 6 reveals that a one-unit increase in verbal abilities raises the probability of choosing an HSS major by 60 percent, whereas a one-unit increase in math abilities decreases this probability by 79.6 percent. Social abilities do not appear to be significant for college major choice in any specification.

Table D.2: Regression Results

Dependent Var.	S&E	B&E	HSS	S&E	B&E	HSS
Math	0.569***	0.258	-0.827***	0.583***	0.214	-0.796***
	(0.171)	(0.170)	(0.178)	(0.153)	(0.158)	(0.166)
Verbal	-0.213	-0.421*	0.634**	-0.130	-0.471*	0.600**
	(0.179)	(0.200)	(0.209)	(0.181)	(0.187)	(0.196)
Social	0.018	-0.095	0.077	0.004	-0.082	0.079
	(0.115)	(0.113)	(0.123)	(0.100)	(0.104)	(0.109)
Demographic Controls	No	No	No	Yes	Yes	Yes
Observations	767	767	767	767	767	767

Notes: This table displays the results of the regression in Equation D.1 with and without controls. Each column uses a dummy variable for the stated major as the dependent variable. Standard errors are shown in parentheses. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively.

D.6 College Major Premium and Shares

I use pooled data from the American Community Survey (ACS) between 2009 and 2019 to calculate college premiums by major and the shares of college graduates across different majors. The analysis focuses on individuals of working age, specifically between 25 and 64 years old, and only includes wage workers who are employed full-time, defined as working more than 30 hours per week for at least 40 weeks in the year. The final sample consists of approximately 11 million observations. All hourly wages are converted to 2019 prices using the CPI index.

To calculate the college major premium, I adopt a regression framework similar to that used in the literature for estimating the overall college premium (see for example (Valletta (2018)):

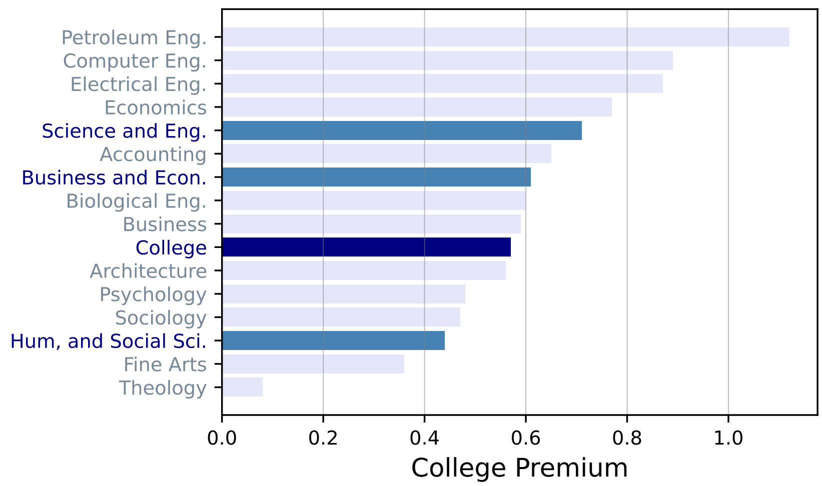
$$\log(\omega_i) = X_i\beta + S_i\theta + \epsilon_i$$

In this equation, X_i represents a set of demographic controls, including dummy variables for seven age groups, three racial/ethnic groups (White, Black, and Others), marital status, and geographic location. The term S_i denotes the college major, where a value of 0 is assigned to non-college individuals, and ω_i represents the hourly wage.

The figure shows the college premium for overall college graduates, three broad college major categories, and selected individual majors. The overall college premium is 0.56, consistent with previous findings in the literature. The college premium for the three broad major groups used in this paper is as follows: 0.71 for Science and Engineering, 0.61 for Business and Economics, and 0.43 for Humanities and Social Sciences. All premiums are calculated relative to the earnings of non-college individuals. For example, Science and Engineering graduates earn 71% higher hourly wages than non-college individuals, after controlling for demographic factors.

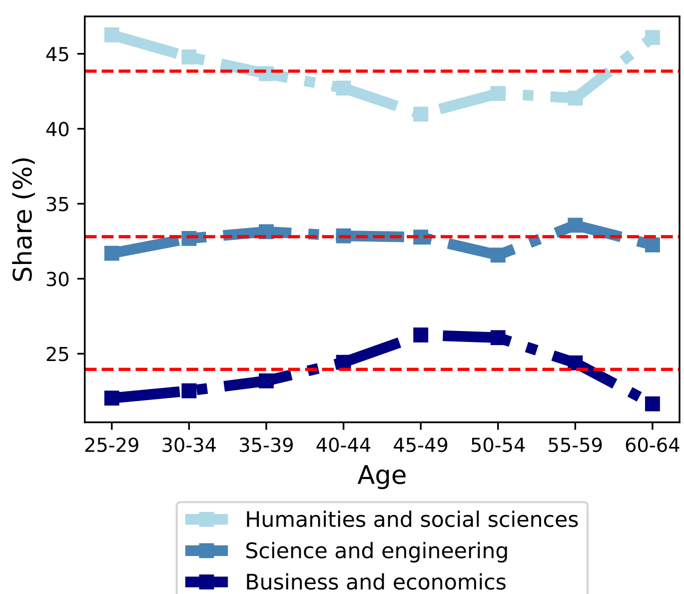
The figure also presents the share of each major's graduates among college graduates across different age groups. The main observation is that the major composition of college graduates remains stable across cohorts over time. Humanities and Social Sciences have the largest share, at almost 43% (16% of the total population). Science and Engineering graduates follow with 33% (12% of the population), and Business and Economics graduates account for 24% (9% of the population).

Figure D.1: College Premium



Notes: This figure displays the college premium for overall college graduates (dark blue bars), three broad college major categories (blue bars), and selected individual majors (light blue bars). The x-axis represents the log-point difference in earnings relative to the non-college group.

Figure D.2: Major Composition of College Graduates



Notes: The figure shows the share of each major's graduates among college graduates across different age groups.

D.7 A Subjective Measure of Skill Mismatch

The National Survey of College Graduates (NSCG) is a survey conducted by the U.S. Census Bureau that provides detailed information about the educational and employment experiences of college graduates in the United States. For this analysis, I pooled data from the 2013, 2015, 2017, and 2019 waves, resulting in a sample size of over 90,000 respondents. The NSCG offers very granular information on college majors. To align the NSCG college major categories with those used in the American Community Survey (ACS), I utilized the crosswalk described in Altonji, Kahn, and Speer (Altonji et al.).

All respondents were asked the following question: "To what extent was your work on your principal job related to your highest degree?" They could choose from three options: *Closely related*, *Somewhat related*, and *Not related*. The table presents the share of responses in each broad college major category, as well as for some selected individual college majors.

Only 53% of college graduates report that their job is closely related to their college degree. This figure rises to 65% for Science and Engineering graduates and drops to 43% for Humanities and Social Sciences graduates, aligning with the differential exposure to skill mismatch in the job market, as discussed in Section 4.3.

Table D.3: College major - Job Mismatch

	Closely Related (%)	**Somewhat Related (%)**	**Not Related (%)**
Science and Engineering	**65**	**23**	**12**
Nursing	87	9	4
Electrical Engineering	59	32	9
Business and Economics	**51**	**29**	**20**
Economics	39	39	22
Business Management	43	37	20
Humanities and Social Sciences	**43**	**31**	**26**
History	39	24	37
Fine Arts	51	22	27
All Sample	**53**	**27**	**20**

*Notes:*The table shows the distribution of responses to the question "To what extent was your work on your principal job related to your highest degree?" in the pooled NSCG sample for the years 2013, 2015, 2017, and 2019.

E Algorithm

Let n and m denote the number of grid points for each dimension of ability and each dimension of job requirements, respectively. The matrix $i(a)$ is a $n^3 \times 4$ matrix representing the share of individuals in each education option for their given ability a. Assume that $g_w(h) = h(g(a), i(a))$ is the $4n^3$-dimensional vector of the distribution of human capital, which is a by-product of the distribution of initial abilities and the education choice policy function. Additionally, I denote the distribution of active matches as $g_m(h, r)$ and the match-specific surplus as $S(h, r)$, both of which are $4n^3 \times m^3$-dimensional. The acceptance rule $A(h, r)$ can then be defined as:

$$A(h,r) = \begin{cases} 1 & if \quad S(h,r) \geq 0 \\ 0 & if \quad S(h,r) < 0 \end{cases} \tag{E.1}$$

The sketch of the solution algorithm is as follows:

1. Set the parameter values and the initial distribution of abilities $g(a)$. Compute the output value $y(h, r)$ and the disutility of overqualification $g(h, r)$. Assign random initial values for $i(a)$, $g_m(h, r)$, $g_v(r)$, and $S(h, r)$, denoting them as i_0, $g_{m,0}$, $g_{v,0}$, and S_0, respectively.

2. Calculate $g_{u,0}$ and U_0 based on the definitions in Table 1. Then, given S_0, $g_{v,0}$, and $g_{u,0}$, update $g_{m,0}$ to $g_{m,1}$ using Equation 25.

3. Find A_0 (equation E.1). For given densities, update S_0 as S_1 by using equation 23.

4. Solve Equation 25 to update g_v and V.

5. Check the convergence conditions for A and g_m:

$$\sum_{i=1}^{4n^3} \sum_{i=1}^{m^3} |A_1^{ij} - A_0^{ij}| \Bigg/ \sum_{i=1}^{4n^3} \sum_{i=1}^{m^3} A_0^{ij} < tol_1$$

$$\sum_{i=1}^{4n^3} \sum_{i=1}^{m^3} |g_{m,1}^{ij} - g_{m,0}^{ij}| \Bigg/ \sum_{i=1}^{4n^3} \sum_{i=1}^{m^3} g_{m,0}^{ij} < tol_1$$

6. If the conditions above are satisfied, proceed to the next step; otherwise, return to step 2 and update the initial values as $g_{m,0} = g_{m,1}$, $S_0 = S_1$, and $g_{v,1} = g_{v,0}$.

7. Calculate V_u for the given $g_{m,1}$ and S_1 using equation 19. Then, solve equation 5 to update the education choice policy function i_1.

8. Check the convergence condition for i.

$$\sum_{i=1}^{n^3} |i_1^i - i_0^i| * g_a^i < pol\_exit$$

9. If the above condition is satisfied, save the results; otherwise, return to step 1 and update the initial value as $i_0 = i_1$.

The algorithm finds the fixed point for matched jobs, the surplus function, and the policy function for education choice. The steps for finding the fixed point of g_m and S closely follow Hagedorn et al. (2012).

F Additional Tables and Figures

Table F.1: Externally Calibrated Parameters

Parameter	Definition	Value	Source
β	Discount factor	0.99	4% anual interest rate
δ	Seperation rate	0.1	Shimer (2005)
ρ	probability of retirement	0.00625	Av. 40 years of working life
ψ	Worker's bargaining power	0.72	Shimer (2005)
α	Matching elasticity	0.72	Hosios condition
b	Flow utility of unemp.	0.4	Shimer(2005)
p	Pecuniary cost of college	$9.75 * \overline{\omega}$	College Board (2022)
τ	College subsidy rate	0.43	College Board (2022)

Notes: This table presents a list of externally calibrated parameters, along with their corresponding values and sources.

Table F.2: Targeted Subsidies and Skill Composition-Additional Results

	100% Drop			75% Drop			25% Drop		
Targeted	S&E	B&E	HSS	S&E	B&E	HSS	S&E	B&E	HSS
Subsidies									
$\tau_{S\&E}$	0	0.43	0.43	0.1075	0.43	0.43	0.215	0.43	0.43
$\tau_{B\&E}$	0.43	0	0.43	0.43	0.1075	0.43	0.43	0.215	0.43
τ_{HSS}	0.43	0.43	0	0.43	0.43	0.1075	0.43	0.43	0.215
Non-college									
math	1.33	0.86	0.62	1.03	0.67	0.49	0.71	0.46	0.34
verbal	1.04	0.67	1.25	0.80	0.52	0.97	0.55	0.36	0.67
social	0.23	0.17	0.14	0.18	0.13	0.11	0.12	0.09	0.08
S&E									
math	0.16	0.02	-0.05	0.13	0.01	-0.04	0.09	0.01	-0.02
verbal	0.07	-0.05	0.09	0.06	-0.04	0.07	0.04	-0.02	0.05
social	-0.16	-0.05	-0.12	-0.13	-0.04	-0.09	-0.09	-0.03	-0.06
B&E									
math	0.27	0.06	-0.05	0.21	0.05	-0.04	0.15	0.03	-0.02
verbal	0.17	-0.02	0.17	0.13	-0.02	0.13	0.09	-0.01	0.09
social	-0.11	-0.04	-0.12	-0.08	-0.03	-0.09	-0.05	-0.02	-0.06
HSS									
math	0.85	0.47	-0.81	0.66	0.36	-0.62	0.45	0.25	-0.42
verbal	0.17	0.09	-0.07	0.13	0.07	-0.06	0.09	0.05	-0.04
social	0.02	0.03	-0.26	0.02	0.03	-0.20	0.01	0.02	-0.14
All									
math	-2.33	-1.66	0.30	-1.79	-1.29	0.24	-1.23	-0.89	0.17
verbal	-0.04	-0.14	-3.09	-0.03	-0.11	-2.38	-0.02	-0.07	-1.63
social	-0.73	0.12	-0.63	-0.56	0.09	-0.49	-0.39	0.06	-0.33

	25% Increase			50% Increase			100% Increase		
Targeted	S&E	B&E	HSS	S&E	B&E	HSS	S&E	B&E	HSS
Subsidies									
$\tau_{S\&E}$	0.5375	0.43	0.43	0.7525	0.43	0.43	0.86	0.43	0.43
$\tau_{B\&E}$	0.43	0.5375	0.43	0.43	0.7525	0.43	0.43	0.86	0.43
τ_{HSS}	0.43	0.43	0.5375	0.43	0.43	0.7525	0.43	0.43	0.86
Non-college									
math	-0.39	-0.26	-0.20	-1.23	-0.84	-0.65	-1.69	-1.17	-0.91
verbal	-0.30	-0.20	-0.37	-0.97	-0.67	-1.18	-1.33	-0.92	-1.63
social	-0.07	-0.05	-0.05	-0.22	-0.17	-0.15	-0.31	-0.23	-0.22
S&E									
math	-0.05	-0.01	0.01	-0.16	-0.02	0.03	-0.22	-0.03	0.04
verbal	-0.02	0.01	-0.03	-0.08	0.04	-0.10	-0.11	0.05	-0.14
social	0.04	0.01	0.03	0.14	0.05	0.11	0.18	0.06	0.14
B&E									
math	-0.08	-0.02	0.01	-0.26	-0.06	0.03	-0.36	-0.09	0.04
verbal	-0.05	0.00	-0.05	-0.17	0.01	-0.18	-0.24	0.01	-0.25
social	0.03	0.01	0.03	0.08	0.04	0.10	0.10	0.05	0.14
HSS									
math	-0.25	-0.14	0.22	-0.80	-0.47	0.68	-1.10	-0.65	0.92
verbal	-0.05	-0.03	0.02	-0.16	-0.09	0.06	-0.22	-0.12	0.08
social	-0.01	-0.01	0.07	-0.03	-0.04	0.22	-0.05	-0.05	0.30
All									
math	0.66	0.49	-0.10	2.08	1.58	-0.31	2.84	2.18	-0.43
verbal	0.01	0.04	0.87	0.04	0.13	2.74	0.05	0.18	3.73
social	0.21	-0.04	0.18	0.65	-0.11	0.55	0.89	-0.15	0.75

Notes: This table is analogous to Table 7 in the main text. It shows the percentage change in average skills across each skill dimension for each educational group relative to the benchmark. The columns in the top panel show the results for a 100%, 75%, and 25% drop in the targeted major's subsidy rate, while the subsidy rates for other majors are kept fixed at the benchmark level. The columns in the bottom panel show the results for a 100%, 75%, and 25% increase in the targeted major's subsidy rate. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively.

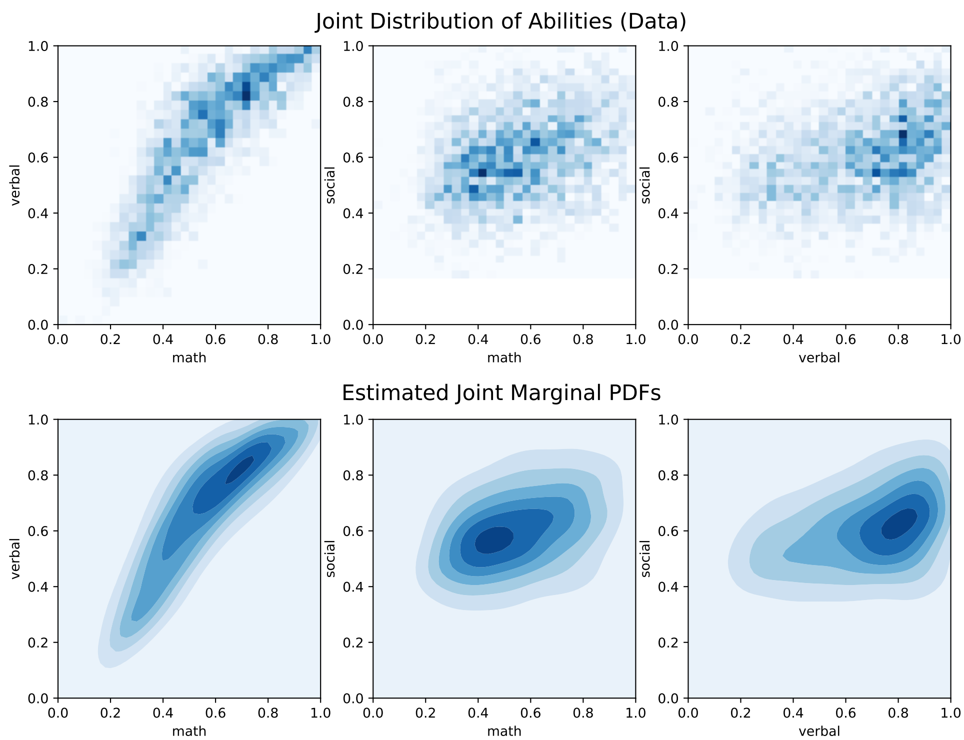
Table F.3: Targeted Subsidies and Earnings-Additional Results

	100% Drop			75% Drop			25% Drop		
Targeted	S&E	B&E	HSS	S&E	B&E	HSS	S&E	B&E	HSS
Subsidies									
S&E	0	0.43	0.43	0.1075	0.43	0.43	0.3225	0.43	0.43
B&E	0.43	0	0.43	0.43	0.1075	0.43	0.43	0.3225	0.43
HSS	0.43	0.43	0	0.43	0.43	0.1075	0.43	0.43	0.3225
Earnings(Δ %)	-0.82	-0.30	-0.65	-0.63	-0.23	-0.50	-0.22	-0.08	-0.17
S&E	0.89	0.62	0.73	0.69	0.48	0.57	0.25	0.17	0.20
B&E	-0.01	-0.01	-0.02	-0.01	-0.01	-0.02	0.00	0.00	-0.01
HSS	0.04	0.01	-0.03	0.03	0.01	-0.03	0.01	0.00	-0.01
Non-college	0.20	0.14	-0.25	0.16	0.11	-0.19	0.05	0.04	-0.07
Col. prem.(Δ pp.)									
S&E	-1.58	-1.10	-1.32	-1.22	-0.86	-1.03	-0.44	-0.31	-0.38
B&E	-1.36	-0.97	-1.22	-1.05	-0.76	-0.95	-0.37	-0.27	-0.34
HSS	-1.05	-0.72	-1.49	-0.82	-0.57	-1.16	-0.29	-0.20	-0.42

	25% Increase			75% Increase			100% Increase		
Targeted	S&E	B&E	HSS	S&E	B&E	HSS	S&E	B&E	HSS
Subsidies									
S&E	0.5375	0.43	0.43	0.7525	0.43	0.43	0.86	0.43	0.43
B&E	0.43	0.5375	0.43	0.43	0.7525	0.43	0.43	0.86	0.43
HSS	0.43	0.43	0.5375	0.43	0.43	0.7525	0.43	0.43	0.86
Earnings(Δ %)	0.23	0.09	0.18	0.73	0.28	0.54	0.98	0.38	0.74
S&E	-0.26	-0.19	-0.23	-0.84	-0.61	-0.75	-1.17	-0.86	-1.03
B&E	0.00	0.00	0.00	-0.01	0.00	0.00	-0.02	0.00	0.01
HSS	-0.01	0.00	0.01	-0.05	-0.01	0.01	-0.08	-0.02	0.01
Non-college	-0.06	-0.05	0.07	-0.19	-0.14	0.20	-0.28	-0.21	0.28
Col. prem.(Δ pp.)									
S&E	0.46	0.33	0.41	1.49	1.09	1.33	2.06	1.52	1.85
B&E	0.40	0.30	0.38	1.28	0.97	1.23	1.77	1.35	1.69
HSS	0.31	0.22	0.45	1.00	0.72	1.46	1.38	1.00	2.02

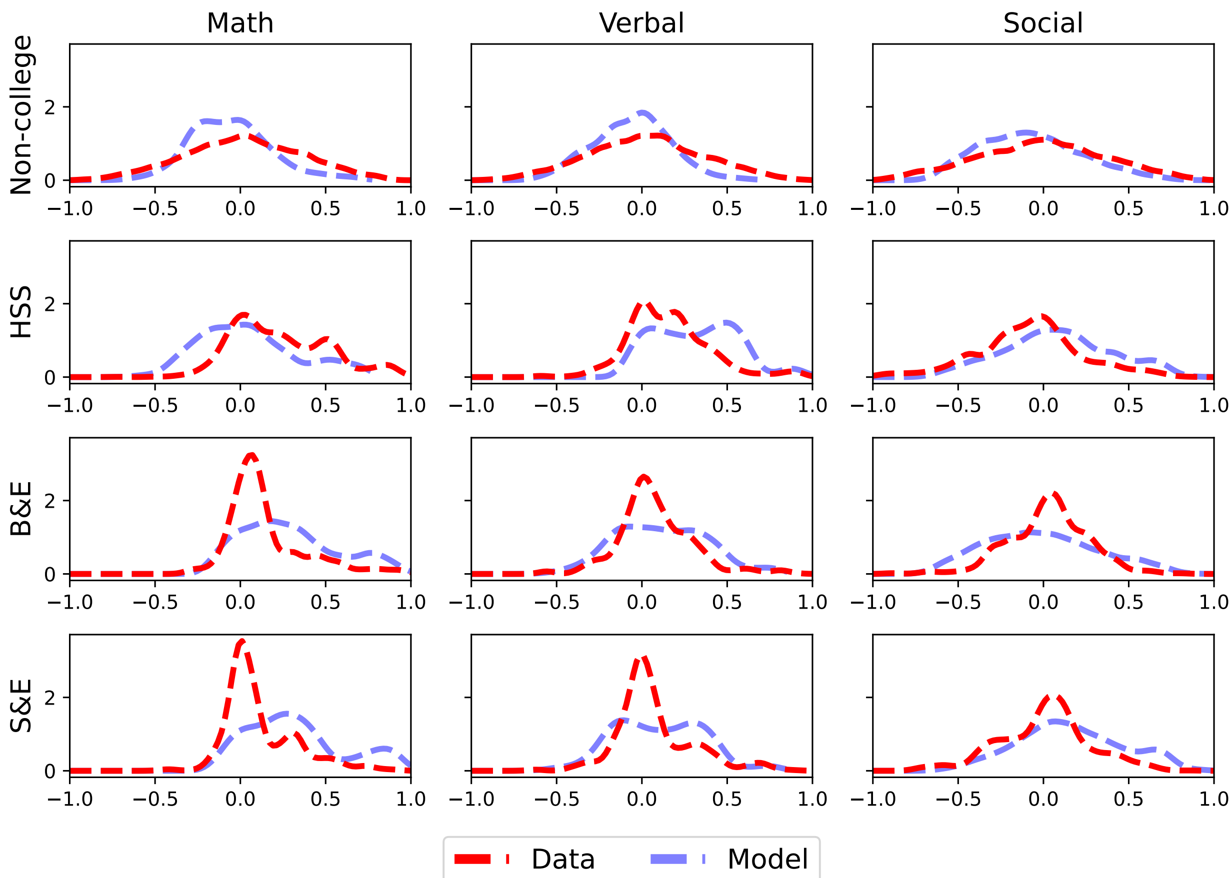
Notes: This table is analogous to Table 8 in the main text. the percentage change in earnings and percentage point changes in college premium. The columns in the top panel show the results for a 100%, 75%, and 25% drop in the targeted major's subsidy rate, while the subsidy rates for other majors are kept fixed at the benchmark level. The columns in the bottom panel show the results for a 100%, 75%, and 25% increase in the targeted major's subsidy rate. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively. College premium is calculated relative to non-college earnings.

Figure F.1: Joint Distribution of Pre-college Abilities



Notes: Source: NLSY79 and O*NET. This figure presents the data histograms (top) and the estimated marginal joint PDFs (bottom) for pre-college abilities across each pair: verbal-math, social-math, and social-verbal. The joint probability distributions are estimated non-parametrically using the kernel density method, from which the cross-joint marginal PDFs are calculated.

Figure F.2: Distribution of Skill Mismatch by Education Group: Model vs. Data



Notes: This figure compares the distribution of skill mismatch between individuals' human capital and the requirements of their jobs for each educaton group and skill dimension. Skill mismatch is measured as the rank difference between each skill and the job requirements, as defined in Section 3.4.3. S&E, B&E, and HSS denote Science and Engineering, Business and Economics, and Humanities and Social Sciences, respectively.

Figure F.3: The Role of Pre-college Abilities - Additional Results

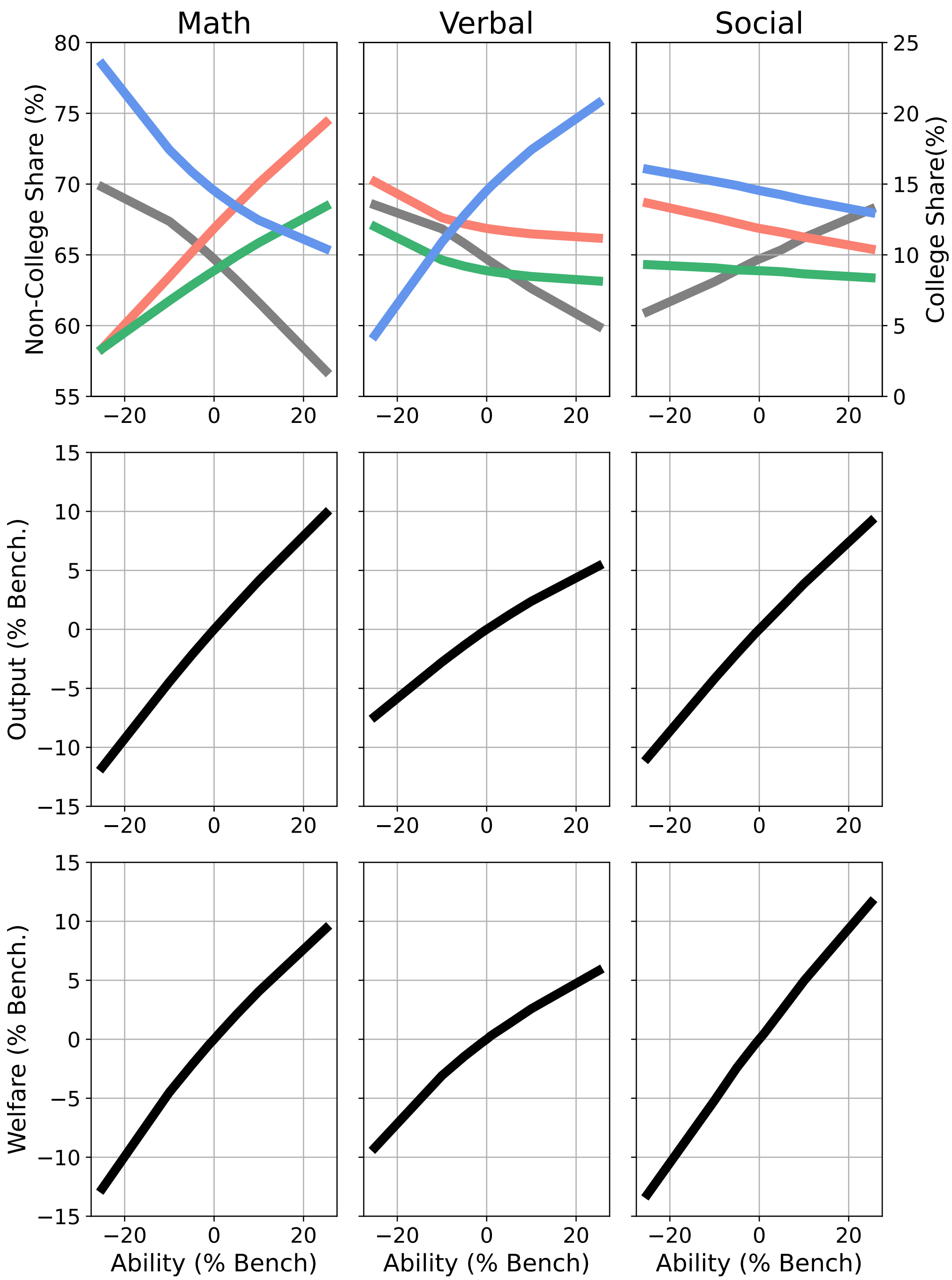


Figure F.4: This figure shows the effect of pre-college abilities on the composition of college graduates, output, and welfare. The x-axis shows the percentage change in initial abilities across relevant skill dimension in each columns. A value of 0 represents the benchmark economy, while a value of 10 represents a scenario where initial abilities of the relevant dimension for all individuals increase by 10 percent.