STEMming the Gender Gap in the Applied Fields: Where are the Leaks in the Pipeline?

Shasha Wang*

April 16, 2024

Latest Version

Abstract

In pure-STEM fields, women are no longer a minority, but in applied-STEM fields like computer science and engineering, their representation remains persistently low for nearly half a century. To understand where the leaks are in the pipeline, I develop and estimate a dynamic model spanning from high school to early career to examine four sources of female under-representation: initial skill gaps, preference differences, wage disparities in STEM sectors, and aversion to male-dominated occupations. In the National Longitudinal Survey of Youth - 1979 Cohort, males show a higher interest in STEM coursework and better STEM skills by 10*th* grade, primarily in mechanical skills, leading to wider skill disparities. Simulation results show that mechanical skills are more important than math skills in explaining women's low participation in applied-STEM fields and have contrasting effects on college enrollment and the selection of applied-STEM majors and occupations. Closing gender skill gaps upon exiting high school reduces female under-representation by 67% in applied-stem majors and 31% in applied-STEM occupations. Removing the preference for female-dominated workplaces reduces female under-representation by 29% in applied-STEM majors and 85% in applied-STEM occupations. Equalizing wage offers in STEM sectors has a smaller effect (3% in majors and 10% in occupations). Mandating more high-school STEM courses increases overall STEM participation but doesn't address the gender gap.

*Shasha Wang, Department of Economics, University of Pennsylvania; Email: wss@sas.upenn.edu

1 Introduction

Are there leaks in the STEM pipeline? Yes, in the applied-STEM fields. Although women obtain more than half of U.S. undergraduate degrees in some pure-STEM fields such as biology, chemistry, and mathematics (Cheryan et al. 2017; Sloane et al. 2021), majors in applied-STEM fields, such as computer science and engineering, are among the highest paid and yet are also those in which the representation of women is persistently 20% or lower (Spertus 1991; Rosser and Lane 2002; Hill et al. 2010; Sloane et al. 2021; Todd and Wang 2022). In terms of the applied-STEM occupations, according to the Census Bureau's American Community Survey (ACS) spanning from 1980 to 2019, within the demographic of individuals aged between 22 and 35 years old who are employed in applied-STEM occupations, women consistently represent approximately 20%.[1] The persistent low participation of women in applied fields is striking when we consider that women have made great strides in catching up with men in both STEM-related skills and high-school coursework, including mathematics and science (Ceci et al. 2014, Todd and Wang 2022), and that jobs in the applied-STEM fields are generally less prone to gender discrimination (Kuhn and Shen 2013), with smaller gender wage gaps compared to other sectors (Beede et al. 2011).

Understanding the reasons for women's low participation in the applied-STEM fields and where the leaks are in the pipeline has important ramifications for enhancing gender equality. First, women may be foregoing career opportunities in lucrative and esteemed professions, resulting in a gap in economic status between men and women (Cortes and Pan 2018; Card and Payne 2021). Second, if women dislike working in male-dominated occupational sectors, then minority status can perpetuate itself and lead to a vicious cycle of female under-representation in these fields (Carrell et al. 2010; Mouganie and Wang 2020; Morando 2020; Shan 2021; Bostwick and Weinberg 2022). Addressing these concerns requires an in-depth understanding of the dynamics of educational and career choices that influence women's participation in STEM.

The under-representation of women in certain STEM fields is a multifaceted issue, as

[1] Following Todd and Wang (2022), I categorize STEM fields into applied-STEM and pure-STEM. Applied-STEM includes computer science, engineering, and engineering technologies ("T" and "E" in STEM), while pure-STEM encompasses biological and physical sciences, environmental science, mathematics, and some social sciences ("S" and "M" in STEM). Further details can be found in Todd and Wang (2022).

elucidated by several specific papers in the literature. Research, such as that conducted by Kahn and Ginther (2017), emphasizes that the roots of this disparity extend back to childhood, with gender differences emerging in STEM interest and performance early on. Ceci et al.'s comprehensive survey in 2014 further suggests that biological factors may contribute to these early disparities. Gender gaps in STEM expectations solidify by middle school, with boys being twice as likely as girls to anticipate careers in science or engineering, as Legewie and DiPrete (2012) finds. Later studies, like Todd and Wang (2022), underline the pivotal role of high-school coursework in shaping STEM majors and careers, emphasizing the significance of adolescent skill profiles and course choices. Moreover, Cheryan et al. (2017) underscores the importance of early exposure to applied-STEM fields, such as computer science and engineering, in reducing gender imbalances. Additionally, Cortes and Pan (2018) reviews the influence of psychological traits and preferences, shedding light on occupational segregation trends in STEM. This body of research collectively underscores the intricate and interconnected factors contributing to the under-representation of women in STEM, from childhood development to high-school choices, encompassing both biological and societal influences. The paper in question aims to delve deeper into these complex mechanisms, using a multi-period model to better understand and address these gender disparities.

To better understand the reasons for low rates of female entry into the applied-STEM fields and where the leaks are in the pipeline, I develop and estimate a comprehensive model of women and men's educational and career pipeline that starts with high-school coursework choices and extends to the phase of career decision-making. The model is a discrete choice dynamic programming (DCDP) model, which I use to analyze individuals' decision-making processes related to high-school course selection, college enrollment, choice of college major, and preferred occupational sector. To the best of my knowledge, this paper represents the first exploration using a dynamic model of the factors contributing to gender gaps along the high-school-to-career pipeline in the context of enhancing STEM participation. Within this model, I establish four channels that contribute to gender disparities: initial 10th-grade skill gaps, choice-specific preference differences, wage offer disparities, and aversions to working in male-dominated occupations. urthermore, I investigates how multi-dimensional skill profiles evolve and influence individual's decisions regarding education and career paths.

The model consists of multiple stages - the first and second concerning high-school

coursework and the later ones pertaining to post-secondary education, college major, and occupation choice. In the first two model time periods, based on their skill levels and gender-and-choice-specific preferences, the individual makes decisions about high-school coursework that have implications for their skill accumulation. In the third period, the individual decides whether to go to a two-year or a four-year college, again based on their skill levels and gender-and-choice-specific preferences. If they choose not to go to college, they make career sector choices right after finishing high school. Individuals choosing to obtain a two-year college degree make career decisions two years later. Those choosing to acquire a four-year college degree make college major decisions two years later and then, another two years later, make occupational sector choices. The model allows occupational choices to depend on wage offers, which may differ by gender and sector. Also, it allows preferences for occupational sectors to potentially depend on the percentage of the opposite sex working in that sector to capture any distaste for working in a male-dominated field.

The model is estimated on the National Longitudinal Survey of Youth - 1979 Cohort (NLSY 79) dataset and using Simulated Method of Moments (SMM).[2] According to the summary statistics of the NLSY79 data, women trail behind men in STEM-related skills (mathematics, science, and mechanical) in 10th grade. The gender gap in mechanical skills stands out as the most pronounced, with females lagging behind by nearly one standard deviation. One of the key findings of my estimation is that the mechanical skill gap is a primary contributor to the under-representation of females in the applied-STEM fields. Any policy aiming to enhance female representation in applied-STEM should prioritize the improvement of mechanical skills.

The estimation results reveal contrasting effects of mechanical skills on college enrollment and the selection of applied-STEM majors. To elaborate, higher pre-college skill levels in math and verbal positively influence the likelihood of pursuing a four-year college education. Higher mechanical skills actually lower the likelihood of attending either type of college, two-year and four-year; a one-standard-deviation unit increase in mechanical skills reduces an individual's odds of pursuing a two(four)-year college education compared to not pursuing any post-secondary education by a factor of 0.754 (0.494). However, if an indi-

[2]The NLSY 79 dataset has a more complete record of high-school courses and college majors than the NLSY 97 dataset. Please refer to Todd and Wang (2022) for the classifications of high-school courses, college majors, and occupations.

vidual chooses to pursue a four-year college degree, mechanical skills significantly increase the chances of attaining an applied-STEM major. A one-standard-deviation increase in mechanical skills multiplies one's odds of achieving an applied-STEM four-year college major against obtaining a non-STEM degree by a factor of 2.570. Additionally, verbal skills negatively impact one's odds of obtaining either type of STEM majors, pure-STEM and applied-STEM. More specifically, verbal skills diminish the odds of obtaining a pure (applied)-STEM degree compared to a non-STEM one, with a one-standard-deviation increase in verbal skills reducing the odds by a factor of 0.602 (0.260).

Considering the interplay among the four skills that influence educational choices and each gender's comparative advantage, it can be inferred that, primarily due to lower mechanical skills, (1) women are more likely than men to enroll in four-year colleges, and (2) female college students are less inclined than men to choose applied-STEM majors. Coupled with the fact that women face the most significant deficit in mechanical skills, my results suggest that addressing the gender gap in mechanical skills could be an effective way to increase participation in applied-STEM fields.

I use the estimated model to perform a series of model simulations aimed at identifying the primary drivers of gender disparities at various stages. Firstly, I conduct simulations to evaluate different sources of gender gaps by (1) equalizing the female choice-specific preference parameters concerning high-school coursework, college enrollment, and college majors with those of males, and (2) aligning females' skill levels with those of males. Next, to examine female under-representation within the context of workplace dynamics, I perform simulations to delve into the non-pecuniary factors by (3) eliminating females' workplace gender composition preferences, and to explore how women react to different monetary incentives, I perform counterfactual simulations by (4) experimenting with different wage policies, including equalizing or enhancing wage offers in specific sectors or across all sectors for males and females. In addition, I explore the effectiveness of coursework requirement policies by (5) implementing coursework requirements mandating high-school students to enroll in more math and science courses to better equip them for entry into STEM fields upon high-school graduation.

The main results from the simulations reveal that closing the initial gender skill gaps in 10th grade reduces female under-representation in applied-STEM majors by 38% and in applied-STEM occupations by 15%. Letting the females have the male choice-specific pref-

erence closes the gender representation gap in applied-STEM majors by 27% and in applied-STEM occupations by 13%. Equalizing wage offers across gender in STEM sectors reduces female under-representation in applied-STEM majors only by 3% and in applied-STEM occupations by 10%. Eliminating females' workplace gender composition preferences decreases the gender participation gap in applied-STEM majors by 29% and in applied-STEM occupations by 85%. Mandating more high-school STEM coursework is effective in increasing overall STEM participation but has little effect on the gender gap. In alignment with the estimation results, which revealed opposing effects of mechanical skills on college enrollment and the pursuit of a four-year college applied-STEM degree, the simulation results underscore a significant trade-off. When using the closure of the mechanical skill gap as a strategy to enhance female participation in applied-STEM fields, there is a noticeable trade-off between the college completion rate and the rate of holding four-year college applied-STEM majors. This dynamic relationship underscores the complexity of addressing gender disparities in these fields.

This paper builds on four strands of literature. The first is the literature that structurally models course selection and skill development in high school, college major choice, and occupation choice. Notable works employing structural models to investigate human capital investment and career specialization include Keane and Wolpin (1997), Arcidiacono et al. (2012), Kinsler and Pavan (2015), and Chuan and Zhang (2021). Other papers that focus on structural modeling of high-school course selection or college major selection, without distinguishing among job types, include Arcidiacono (2004), Harvill (2011), Stinebrickner and Stinebrickner (2012), Stinebrickner and Stinebrickner (2014a), and Stinebrickner and Stinebrickner (2014b). Recall that the recent closure of the gender gaps in high-school coursework and skill levels (Todd and Wang 2022) contrasts with the persistent gender disparities in applied-STEM majors and occupations. This underscores the need to comprehend how high-school coursework contributes to later-stage gender segregation and the dynamics among relevant factors shaping the STEM journey from high school to the workplace. Hence, my model integrates high-school course-taking decisions and skill developments with subsequent college and early career choices within the unified framework of non/pure/applied-STEM taxonomy. In a broader context, my analysis could be considered a dynamic extension of the classic Roy (1951) model, where individuals initially invest in skills through high-school coursework, years of education, and college major selection be-

fore choosing a job type based on expected wage returns given their skills. To the best of my knowledge, this is the first paper to integrate individuals' high-school coursework selection, college major choice, and occupation specialization into a sequence of dynamic decisions within the context of STEM participation.

Another contribution of this paper is its exploration of the evolution of multi-dimensional skill profiles within a dynamic model. Previous studies mostly focus on math ability or a combination of math and verbal skills in previous studies, as seen in the works of Arcidiacono (2004) and Kinsler and Pavan (2015). Although Chuan and Zhang (2021) introduces mechanical skills alongside cognitive skills, they do not examine the relevance for STEM participation. Building on Todd and Wang (2022)'s observation that gender representation varies between pure-STEM and applied-STEM fields, I refine non-math, non-verbal abilities to accommodate the finer classification of college majors and occupations. Specifically, I adopt Todd and Wang (2022)'s framework to establish a multi-dimensional skill profile, which includes four dimensions: three types of STEM-related skills (math, science, mechanical[3]) and verbal skills.[4] Correspondingly, high-school coursework is similarly categorized to better assess the value-added knowledge production function. STEM-related high-school coursework is divided into three types: math courses, non-math pure-STEM courses (referred to as "pure-STEM courses" hereafter), and applied-STEM courses. This richer approach to characterizing skill profiles and coursework enables a more in-depth exploration of the dynamic interplay among these distinct skill categories, the dynamics of educational and career pathways, and the emergence of gender disparities.

A third contribution of this paper is its examination of skill development during high school, specifically through specialized course selection. The concept that a portion of an individual's acquired human skills is not universally transferable to all fields has been integrated into education and career models for years, such as Bartolj and Polanec (2012) and Kinsler and Pavan (2015). However, in most of these models, the primary means by which workers accumulate specific human skills is through work experience and the level of education, as exemplified by Keane and Wolpin (1997) and Sullivan (2010). What sets this paper

[3]Mechanical knowledge entails the ability to comprehend the functioning of mechanical equipment and tools and apply this understanding in practical scenarios. Further details can be found in Todd and Wang (2022).

[4]Administrative skills are excluded from this profile, as the primary focus here pertains to participation in STEM fields, which is closely linked to cognitive and practical skills rather than administrative capabilities.

apart is the model setup that individuals actively build specific human skills through their choice of educational path, namely, through the selection of specific coursework.

A fourth contribution of this paper is to allow occupational sector choices to depend on gender composition and to quantify the importance of this channel.[5] Evidence from Delfino (2021) and Larson-Koester (2017) suggests that women tend to favor occupations with higher proportions of females. However, it is unknown whether these preferences are significant enough to account for aggregate patterns of occupational segregation in STEM fields. Although Schuh (2022) quantifies preferences for workplace gender makeup, her work primarily focuses on occupational gender sorting and does not consider college-major gender sorting, which is closely linked to occupational gender segregation. Drawing from Schuh (2022)'s findings, which indicate that such preferences are non-linear and concave, my model operates within the framework of the non/pure/applied-STEM taxonomy and adopts a quadratic form of the utility function concerning the proportion of the opposite sex in the occupation. This approach allows me to quantify the wage-equivalent necessary to counteract females' inclination to avoid applied-STEM jobs to be an extra 81% of the wage offer.

In summary, this study provides a comprehensive examination of gender gaps in STEM participation, uncovering the intricate relationships between skills, preferences, and policy interventions. Three main implications can be drawn from the analysis. First, subject-specific preferences are the primary driving force in making high-school coursework decisions. Recent studies showing the closure of gender coursework gaps (Todd and Wang 2022; Ceci et al. 2014) speak to the recent closure of math and science skills and of the underlying subject-specific preference gaps. Second, since mechanical skills are the most important factor in deciding one's choosing a four-year applied-STEM major, it is mainly due to the limited shrinking of the gender gap in mechanical skills in comparison to the closure of that in math skills and science skills, that the percentage of women choosing applied-STEM majors has remained consistently low among the overall population, and become even lower among the college population. Third, it is primarily due to women's distaste for working in male-dominated occupations that female under-representation in applied-STEM occupa-

[5]Such preferences could arise for a variety of reasons possibly related to discrimination and harassment, gender-related behavioral disparities, access to gender-specific accommodations such as nursing facilities, and turnover rates attributed to disappointment.

tions has remained stubbornly low. Closing the gender skill gap in mechanical skills would reduce much of the female under-representation in applied-STEM majors, but not as much of that in applied-STEM occupations. Thus it is a policy imperative to address women's concern for working in male-dominated occupations. By illuminating the evolving roles of these factors, this research offers insights for policymakers, educators, and stakeholders aiming to create a more inclusive and diverse STEM landscape. The urgency of rectifying gender disparities in STEM demands evidence-based strategies, and my study is a crucial step in guiding these efforts toward a more equitable future.

The subsequent sections are organized as follows: Section 2 outlines the specifics of the model setup, Section 3 presents the estimation results, followed by the counterfactual analysis in Section 4, and the concluding remarks in Section 5.

2 Model specification

The analysis is restricted to high-school graduates starting from before the beginning of the $11th$ grade till the end of college when they make decisions about careers. There are male and female high-school students. Each student i is characterized by $\{X_i, \mathbf{\Theta}_{it}\}$, where $X_i = \{F_i, B_i, PG_i\}$. Note that $F_i = 1$ if i is female and 0 if otherwise, $B_i = 1$ if i is black and 0 if otherwise, $PG_i \in \{1,2,3\}$ where $PG_i = 1$ if i has no parent completing more than 12 years of education, $PG_i = 2$ if i has at least one parent completing more than 12 and less than 16 years of education and neither completing 16 years and above, and $PG_i = 3$ if i has at least one parent completing more than 16 years of education. Each individual i is also characterized by a time-evolving vector of four-dimensional skill profile $\mathbf{\Theta}_{it} = [\theta_{it}(math), \theta_{it}(sci), \theta_{it}(mech), \theta_{it}(verb)]$.

The remainder of this section presents the details of the model. First, I characterize the skill accumulation phase, i.e., the very first two periods of the model, since this phase binds the high-school course-taking and later education and career choices together. I then discuss individual's choices and outcomes in a recursive fashion, starting with outcomes in the labor market and moving backward to the decision regarding college major and post-secondary education choice. The complete timeline of the model is drawn in Figure 9 in the appendix.

2.1 Skill accumulation phase $(t = 1, 2)$: 11-12th grade's coursework

The decisions are made annually during high school from the very beginning of the $11th$ grade and bi-annually afterward. For $t \in \{1, 2\}$, each individual makes decisions about their high-school coursework. There are three types of courses $h \in \{math, ps, as\}$ - math, pure-STEM courses consisting of physical sciences (e.g., physics and chemistry), life sciences (e.g., biology, health), and courses on social sciences, and applied-STEM courses including vocational training.

High-school Course-taking Preferences Let n_{it}^h denote the number of courses in the course category h the individual i chooses to take at period t, and let $\boldsymbol{N}_{it} = [n_{it}^{math}, n_{it}^{ps}, n_{it}^{as}]$ denote the course-taking vector of individual i at period $t \in \{1, 2\}$. More specifically, $n_{it}^{math} \in \{0, 1, 2\}$, $n_{it}^{ps} \in \{0, 1, 2, 3\}$, $n_{it}^{as} \in \{0, 1, 2\}$, so $\boldsymbol{N}_{it} \in \{[0, 0, 0], [1, 0, 0], [2, 0, 0], [0, 1, 0], [1, 1, 0], [2, 1, 0], ..., [2, 3, 2]\}$.

Let $\boldsymbol{\kappa}_{\theta H}^h = [\kappa_{\theta H}^h(math), \kappa_{\theta H}^h(sci), \kappa_{\theta H}^h(mech), \kappa_{\theta H}^h(verb)]'$ denote the interaction effect of skill and the number of the h course taken, where $h \in \{math, ps, as\}$ - math, pure-STEM, and applied-STEM courses. For instance, $\kappa_{\theta H}^{ps}(math)$ denotes the interaction effect of the math skill and the number of pure-STEM courses taken on one's utility. Similarly, let κ_{fH}^h denote the female gender effect on taking the course h, and κ_{bH}^h denote the black racial effect on talking the course h. Let κ_{pH2}^h and κ_{pH3}^h denote the parents' education effect on talking the course h.

The individual i's period utility function from taking course $h \in \{math, ps, as\}$, for $t \in \{1, 2\}$, i.e., $U_{it}^h(\overline{S}_{it}, n_{it}^h)$ depends on the state variables $\overline{S}_{it} = (\boldsymbol{\Theta}_{it}, F_i, B_i, PG_i)$, and part of the choice variable $n_{it}^h \in d_{it} = \boldsymbol{N}_{it}$, i.e.,

$$U_{it}^h(\overline{S}_{it}, n_{it}^h) = n_{it}^h \times \Bigg(\underbrace{\boldsymbol{\Theta}_{it}\boldsymbol{\kappa}_{\theta H}^h}_{\text{skill effect}} + \underbrace{\mathbb{1}\{F_i = 1\} \times \kappa_{fH}^h}_{\text{female gender effect}} + \underbrace{\mathbb{1}\{B_i = 1\} \times \kappa_{bH}^h}_{\text{race effect}} + \underbrace{\mathbb{1}\{PG_i = 2\} \times \kappa_{pH2}^h + \mathbb{1}\{PG_i = 3\} \times \kappa_{pH3}^h}_{\text{parents' education effect}} \Bigg), \quad t \in \{1, 2\}. \tag{1}$$

Assume the utility obtained from taking high-school courses is additive. The total utility gotten from taking all three types of courses $U_{it}(\overline{S}_{it}, d_{it})$ depends on the state variables $\overline{S}_{it} =$

$(\boldsymbol{\Theta}_{it}, F_i, B_i, PG_i)$ and the choice variable $d_{it} = \boldsymbol{N}_{it} = [n_{it}^{math}, n_{it}^{ps}, n_{it}^{as}]$ and can be written as below,

$$U_{it}(\overline{S}_{it}, d_{it} = [n_{it}^{math}, n_{it}^{ps}, n_{it}^{as}]) = \sum_{h \in math,ps,as} U_{it}^{h}(\overline{S}_{it}, n_{it}^{h}), \quad t \in \{1,2\}. \tag{2}$$

Let $\epsilon_{it}^{d}(hs) \in \boldsymbol{\epsilon}_{it}(hs)$ denote the alternative-specific utility shocks that follow the standard extreme value type one distribution with a standard deviation of $\sigma_{hs} = 1$, and let $\boldsymbol{\epsilon}_{it}^{\theta} = [\epsilon_{it}^{\theta}(math), \epsilon_{it}^{\theta}(sci), \epsilon_{it}^{\theta}(mech), \epsilon_{it}^{\theta}(verb)]'$ drawn from a multivariate normal distribution $MVN(0, \boldsymbol{\Sigma}_{hs})$ denote the skill evolution error shock.

The dynamics are that this period's coursework decision affects the next period's skill profiles, shown in the Bellman equation for choosing a particular choice d, i.e.,

$$V_t^d(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}, \epsilon_{it}^{d}(hs)) = U_{it}(\overline{S}_{it}, d_{it} = d) + \delta \mathbb{E}\left(V_{t+1}(\overline{S}_{i,t+1}, \boldsymbol{\epsilon}_{i,t+1}^{\theta} \mid \overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}, d_{it} = d) \right) + \epsilon_{it}^{d}(hs), \quad t \in \{1,2\}. \tag{3}$$

The value function is as below,

$$V_t(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}, \boldsymbol{\epsilon}_{it}(hs)) = \max_{d_{it}} \sum_{d} \left(V_t^d(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}, \epsilon_{it}^{d}(hs)) \times \mathbb{1}\{d_{it} = d\} \right), \quad t \in \{1,2\}. \tag{4}$$

Let $v_{it}^{d}(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}) = U_{it}(\overline{S}_{it}, d_{it} = d) + \delta \mathbb{E}\left(V_{t+1}(\overline{S}_{i,t+1}, \boldsymbol{\epsilon}_{i,t+1}^{\theta} \mid \overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}, d_{it} = d) \right)$ denote the Bellman equation $V_t^d(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}, \epsilon_{it}^{d}(hs))$ net of the type one extreme value shock $\epsilon_{it}^{d}(hs)$. According to Rust (1987), the expected value function is

$$\mathbb{E}V_t(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}) = \sigma_{hs} \left\{ \mathfrak{e} + \log \left[\sum_{d} \exp\left(\frac{v_{it}^{d}(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta})}{\sigma_{hs}} \right) \right] \right\}, \quad t \in \{1,2\}, \tag{5}$$

and the choice probability is

$$\mathbb{P}(d_{it} = d \mid \overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta}) = \frac{\exp\left(\frac{v_{it}^{d}(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta})}{\sigma_{hs}} \right)}{\sum_{\hat{d}} exp\left(\frac{v_{it}^{\hat{d}}(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}^{\theta})}{\sigma_{hs}} \right)}, \quad t \in \{1,2\}, \tag{6}$$

where $\mathfrak{e}$ is the Euler-Mascheroni's constant.

Skill Evolution As for the transition of test scores $\boldsymbol{\Theta}_{it} = [\theta_{it}^{math}, \theta_{it}^{sci}, \theta_{it}^{mech}, \theta_{it}^{verb}]$ where $t \in \{1,2\}$, each element of next period's test scores $\boldsymbol{\Theta}_{i,t+1}$ is a linear combination of this period's

score $\mathbf{\Theta}_{it}$, high-school course-taking behavior $\boldsymbol{N}_{it}$, and an error term $\boldsymbol{\epsilon}_{it}^{\theta} = [\epsilon_{it}^{\theta}(math), \epsilon_{it}^{\theta}(sci), \epsilon_{it}^{\theta}(mech), \epsilon_{it}^{\theta}(verb)]'$ drawn from a multivariate normal distribution $MVN(0, \mathbf{\Sigma}_{hs})$.

Let $\boldsymbol{\gamma}_0 = [\gamma_0(math), \gamma_0(sci), \gamma_0(mech), \gamma_0(verb)]'$ denote the intercept. Let $\boldsymbol{\gamma}_\theta = [\gamma_\theta(math), \gamma_\theta(sci), \gamma_\theta(mech), \gamma_\theta(verb)]'$ denote the effects of this period's test score $\mathbf{\Theta}_{it}$ on next period's test score $\mathbf{\Theta}_{i,t+1}$. Let $\boldsymbol{\gamma}_h(s) = [\gamma_h^{math}(s), \gamma_h^{ps}(s), \gamma_h^{as}(s)]'$ denote the effect of each type of course taken on next period's s score, where $s \in \{math, sci, mech, verb\}$.

The transition function of test scores can be written as a combination of the intercept, the score effect, the course-taking effect, and an error term, i.e.,

$$
\begin{aligned}
\theta_{i,t+1}(math) &= \gamma_0(math) &&+ \gamma_\theta(math) \times \theta_{it}(math) &&+ \boldsymbol{N}_{it}\gamma_h(math) &&+ \epsilon_{it}^{\theta}(math), \\
\theta_{i,t+1}(sci) &= \gamma_0(sci) &&+ \gamma_\theta(sci) \times \theta_{it}(sci) &&+ \boldsymbol{N}_{it}\gamma_h(sci) &&+ \epsilon_{it}^{\theta}(sci), \\
\theta_{i,t+1}(mech) &= \gamma_0(mech) &&+ \gamma_\theta(mech) \times \theta_{it}(mech) &&+ \boldsymbol{N}_{it}\gamma_h(mech) &&+ \epsilon_{it}^{\theta}(mech), \\
\theta_{i,t+1}(verb) &= \underbrace{\gamma_0(verb)}_{\text{intercept}} &&+ \underbrace{\gamma_\theta(verb) \times \theta_{it}(verb)}_{\text{score effect}} &&+ \underbrace{\boldsymbol{N}_{it}\gamma_h(verb)}_{\text{course effect}} &&+ \underbrace{\epsilon_{it}^{\theta}(verb)}_{\text{error term}}, \quad t \in \{1, 2\}.
\end{aligned} \tag{7}
$$

From Period 3 onward, the ability level $\mathbf{\Theta}_{it}$ is fixed and denoted as $\mathbf{\Theta}_{i3} = [\theta_{i3}^{math}, \theta_{i3}^{sci}, \theta_{i3}^{mech}, \theta_{i3}^{verb}]$. Although it is reasonable to expect that human capital evolves with post-secondary education and job experiences, I assume that skills remain constant since there are relatively few observations who take the ASVAB tests after they have attained four-year colleges.

The later periods are discussed backward, i.e., first the occupation period, then the college major choice period, and then the post-secondary education choice period.

2.2 Career choice $(t = T)$: non-STEM job, pure-STEM job, applied-STEM job, or not working fulltime

The last period (T) is the career decision. At this stage, an individual i is characterized by the state variable vector $\overline{S}_{it} = (E_i, M_i, F_i, B_i)$, where $E_i \in \{0, 2, 4\}$ denotes the education choice, i.e., attending neither a two-year college nor a four-year college, attending a two-year college, and attending a four-year college respectively; $M_i \in \{1, 2, 3\}$ denotes the four-year college major type, i.e., non-STEM, pure-STEM, and applied-STEM;

Variables The individual is choosing among four occupation alternatives, i.e., $d_{it} = occ_i \in$ {non-STEM job, pure-STEM job, applied-STEM job, not working fulltime}, denoted by $d_{it} \in \{1, 2, 3, 4\}$ respectively. This period's utility function consists of a pecuniary component $\log Y(\cdot)$ if the individual works and a non-pecuniary component $\log P(\cdot)$.

The period utility function can be written as,

$$U_{it}(\overline{S}_{it}, d_{it}) = \mathbb{1}(d_{it} \in \{1,2,3\})\Big(\underbrace{\log Y(\overline{S}_{it}, d_{it})}_{\text{pecuniary return}} + \underbrace{\log P(\overline{S}_{it}, d_{it})}_{\text{gender composition preference}} + \underbrace{logQ(\overline{S}_{it}, d_{it})}_{\text{occ-major incompatibility effect}}\Big). \tag{8}$$

Wage offer equations More specifically, let β^k_{female} denote the female gender effect on wages given a type of occupation k. Let β^k_{black} denote the race effect on wages given a type of occupation k. Let β^k_{2yr} denote the two-year college degree effect on wages given a type of occupation k. Let $\boldsymbol{\beta}^k_{4yr} = [\beta^k_{4yr,ns}, \beta^k_{4yr,ps}, \beta^k_{4yr,as}]$ denote the four-year college degree effect on wage given a type of occupation k and a type of major degree non-STEM, pure-STEM, or applied-STEM. Let β^k denote the occupation fixed effect given a type of occupation k.

For a particular occupation k, the wage function $\log Y^k(\cdot)$ can be written as,

$$\begin{aligned}\log Y(\overline{S}_{it}, d_{it} = k) = &\underbrace{\mathbb{1}\{F_i = 1\} \times \beta^k_{female}}_{\text{female gender effect}} + \underbrace{\mathbb{1}\{B_i = 1\} \times \beta^k_{black}}_{\text{race effect}} \\ &+ \underbrace{\mathbb{1}\{E_i = 2\} \times \beta^k_{2yr}}_{\text{two-year college premium}} + \underbrace{\mathbb{1}\{E_i = 4\} \times \sum_{j=1}^{3}\left(\mathbb{1}\{M_i = j\} \times \beta^k_{4yr,j}\right)}_{\text{major-specific four-year college premium}} \\ &+ \underbrace{\beta^k}_{\text{occupation fixed effect}} + \underbrace{u^k_{it}}_{\text{error term}},\end{aligned} \tag{9}$$

where the error term u^k_{it} is drawn from a sector-specific normal distribution $N(0, \beta^k_\sigma)$, realized after the occupation has been chosen.

Non-pecuniary Benefits The non-pecuniary benefits consist of the part from gender composition and the part from occupation-major incompatibility.

Let $\alpha^c_{g,1}$ denote the linear effect, and $\alpha^c_{g,2}$ denote the quadratic effect of the opposite sex's share on the non-pecuniary benefit, where $c \in \{nb, b\}$[6] and $g \in \{male, female\}$. In other words, the non-pecuniary benefit an individual i gets from choosing an occupation k is associated with the proportion of the opposite sex choosing to work in k, denoted by $r_{female,k}$ for the female share and $r_{male,k}$ for the male share.

Let $\alpha^k_{maj \neq occ}$ denote the effect of occupation-major incompatibility on the non-pecuniary benefit of an individual having neither types of four-year STEM degrees entering a STEM occupation, where k $\in \{ps, as\}$ denote the occupation type pure-STEM and applied-STEM respectively. Let $\alpha^{ns}_{maj \neq occ}(s)$ denote the effect of occupation-major incompatibility on the non-pecuniary benefit of an individual having no four-year non-STEM degree entering a non-STEM occupation, where s $\in \{< 4yr, 4yr_ps, 4yr_as\}$ denote the degree type held by this individual. More specifically, $\alpha^{ns}_{maj \neq occ}(< 4yr)$ denotes the effect of having no four-year college degree on the non-pecuniary benefit of entering a non-STEM occupation, $\alpha^{ns}_{maj \neq occ}(4yr_ps)$ denote the effect of having no four-year non-STEM degree on the non-pecuniary benefit of an individual having a four-year pure-STEM degree entering a non-STEM occupation, and $\alpha^{ns}_{maj \neq occ}(4yr_as)$ denote the effect of having no four-year non-STEM degree on the non-pecuniary benefit of an individual having a four-year applied-STEM degree entering a non-STEM occupation.

Thus for a particular occupation k, the non-pecuniary benefit function can be written as,

$$
\begin{aligned}
\log P(\overline{S}_{it}, d_{it} = k) = {} & \mathbb{1}\{F_i = 0\} \times \mathbb{1}\{B_i = 0\} \times \Bigg(\underbrace{\alpha^{nb}_{male,1} \times r_{female,k} + \alpha^{nb}_{male,2} \times \left(r_{female,k}\right)^2}_{\text{gender composition effect for white or Hispanic males}} \Bigg) \\
& + \mathbb{1}\{F_i = 0\} \times \mathbb{1}\{B_i = 1\} \times \Bigg(\underbrace{\alpha^{b}_{male,1} \times r_{female,k} + \alpha^{b}_{male,2} \times \left(r_{female,k}\right)^2}_{\text{gender composition effect for black males}} \Bigg) \\
& + \mathbb{1}\{F_i = 1\} \times \mathbb{1}\{B_i = 0\} \times \Bigg(\underbrace{\alpha^{nb}_{female,1} \times r_{male,k} + \alpha^{nb}_{female,2} \times \left(r_{male,k}\right)^2}_{\text{gender composition effect for white or Hispanic females}} \Bigg) \\
& + \mathbb{1}\{F_i = 1\} \times \mathbb{1}\{B_i = 1\} \times \Bigg(\underbrace{\alpha^{b}_{female,1} \times r_{male,k} + \alpha^{b}_{female,2} \times \left(r_{male,k}\right)^2}_{\text{gender composition effect for black females}} \Bigg), \qquad (10)
\end{aligned}
$$

[6] I.e., $\{nonblack, black\}$.

$$\log Q(\overline{S}_{it}, d_{it} = k) = \underbrace{\mathbb{1}\{k = 2\} \times \mathbb{1}\{M_i \neq 2\} \times \alpha^{ps}_{maj \neq occ} + \mathbb{1}\{k = 3\} \times \mathbb{1}\{M_i \neq 3\} \times \alpha^{as}_{maj \neq occ}}_{\text{occupation-major incompatibility effect for STEM occupations}}$$
$$+ \mathbb{1}\{k = 1\} \times \Bigg(\mathbb{1}\{E_i \neq 4\} \times \alpha^{ns}_{maj \neq occ}(< 4yr)$$
$$\underbrace{+ \mathbb{1}\{M_i = 2\} \times \alpha^{ns}_{maj \neq occ}(4yr_ps) + \mathbb{1}\{M_i = 3\} \times \alpha^{ns}_{maj \neq occ}(4yr_as) \Bigg)}_{\text{occupation-major incompatibility effect for non-STEM occupations}}. \quad (11)$$

Note that for normalization we set the monetary and the non-pecuniary benefits of not working fulltime both to be 0.

Let $\epsilon^k_{it}(occ) \in \boldsymbol{\epsilon}_{it}(occ)$ denote the alternative-specific utility shocks that follow the standard extreme value type one distribution with a standard deviation of $\sigma_{occ} = 1$. The Bellman equation for choosing a particular occupation k is

$$V^k_t(\overline{S}_{it}, \epsilon^k_{it}(occ)) = U_{it}(\overline{S}_{it}, d_{it} = k) + \epsilon^k_{it}(occ), \quad t = T. \quad (12)$$

According to Rust (1987), the value function is,

$$V_t(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}(occ)) = \max_{d_{it}} \sum_k \left(V^k_t(\overline{S}_{it}, \epsilon^k_{it}(occ)) \times \mathbb{1}\{d_{it} = k\} \right), \quad t = T. \quad (13)$$

The expected value function is

$$\mathbb{E}V_t(\overline{S}_{it}) = \sigma_{occ} \left\{ \mathfrak{e} + \log \left[\sum_k \exp \left(\frac{U_{it}(\overline{S}_{it}, k)}{\sigma_{occ}} \right) \right] \right\}, \quad t = T, \quad (14)$$

and the choice probability is

$$\mathbb{P}(d_{it} = k \mid \overline{S}_{it}) = \frac{\exp\left(\frac{U_{it}(\overline{S}_{it}, k)}{\sigma_{occ}}\right)}{\sum_{\hat{k}} exp\left(\frac{U_{it}(\overline{S}_{it}, \hat{k})}{\sigma_{occ}}\right)}, \quad t = T, \quad (15)$$

where $\overline{S}_{it} = (E_i, M_i, \boldsymbol{\Theta}_{i3}, F_i, B_i)$ is the state variable, and $\mathfrak{e}$ is the Euler-Mascheroni's constant.

2.3 Major choice $(t = T - 1)$, given $E_i = 4$: non-STEM, pure-STEM, or applied-STEM major

For those who have chosen to attend a four-year college, two years later, they will decide on college majors.

At this stage, an individual i is characterized by the state variable vector $\overline{S}_{it} = (E_i = 4, \boldsymbol{\Theta}_{i3}, F_i, B_i)$, where $E_i = 4$ denotes that the individual has chosen to attend a four-year college; $\boldsymbol{\Theta}_{i3}$ is the test score vector that has built up through high-school coursework.

The individual is choosing among three four-year college major alternatives, i.e., $d_{it} = maj_i \in \{\text{non-STEM major, pure-STEM major, applied-STEM major}\}$, denoted by $d_{it} \in \{1, 2, 3\}$ respectively.

Let $\boldsymbol{\eta}^j_{\Theta} = [\eta^j(math), \eta^j(sci), \eta^j(mech), \eta^j(verb)]'$ denote the skill effect on utility given a type of four-year college major j. Let η^j_{female} denote the female gender effect on utility given a type of major j. Let η^j_{black} denote the race effect on utility given a type of major j. Let $\eta^j_{f\times b}$ denote the gender-race interaction effect on utility given a type of major j. Let η^j denote the major fixed effect on utility given a type of major j.

For a particular college major j, the utility function $U(S_{it}, j)$ can be written as,

$$\begin{aligned} U_{it}(\overline{S}_{it}, d_{it} = j) = \underbrace{\boldsymbol{\Theta}_{i3}\boldsymbol{\eta}^j_{\Theta}}_{\text{skill effect}} + \underbrace{\mathbb{1}\{F_i = 1\} \times \eta^j_{female}}_{\text{female gender effect}} + \underbrace{\mathbb{1}\{B_i = 1\} \times \eta^j_{black}}_{\text{race effect}} \\ + \underbrace{\mathbb{1}\{F_i = 1\} \times \mathbb{1}\{B_i = 1\} \times \eta^j_{f\times b}}_{\text{gender-race interaction effect}} + \underbrace{\eta^j}_{\text{constant}}, \quad t = T - 1. \end{aligned} \tag{16}$$

Let $\epsilon^j_{it}(maj) \in \boldsymbol{\epsilon}_{it}(maj)$ denote the alternative-specific utility shocks that follow the standard extreme value type one distribution with a standard deviation of $\sigma_{maj} = 1$. The Bellman equation for choosing a particular major j is

$$V^j_t(\overline{S}_{it}, \epsilon^j_{it}(maj)) = U_{it}(\overline{S}_{it}, d_{it} = j) + \delta^2 \mathbb{E}\left(V_{t+1}(\overline{S}_{i,t+1} \mid \overline{S}_{it}, d_{it} = j)\right) + \epsilon^j_{it}(maj), \quad t = T - 1, \quad E_i = 4. \tag{17}$$

The value function is,

$$V_t(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}(maj)) = \max_{d_{it}} \sum_j \left(V^j_t(\overline{S}_{it}, \epsilon^j_{it}(maj)) \times \mathbb{1}\{d_{it} = j\}\right), \quad t = T - 1, \quad E_i = 4. \tag{18}$$

Let $v^j_{it}(\overline{S}_{it}) = U_{it}(\overline{S}_{it}, d_{it} = j) + \delta^2 \mathbb{E}\left(V_{t+1}(\overline{S}_{i,t+1} \mid \overline{S}_{it}, d_{it} = j)\right)$ denote the Bellman equation $V^j_t(S_{it}, \epsilon^j_{it}(maj))$ net of the type one extreme value shock $\epsilon^j_{it}(maj)$. According to

Rust (1987), the expected value function is

$$\mathbb{E}V_t(\overline{S}_{it}) = \sigma_{maj}\left\{\mathfrak{e} + \log\left[\sum_j \exp\left(\frac{v_{it}^j(\overline{S}_{it})}{\sigma_{maj}}\right)\right]\right\}, \quad t = T-1, \quad E_i = 4, \tag{19}$$

and the choice probability is

$$\mathbb{P}(d_{it} = j \mid \overline{S}_{it}) = \frac{\exp\left(\frac{v_{it}^j(\overline{S}_{it})}{\sigma_{maj}}\right)}{\sum_{\hat{j}} exp\left(\frac{v_{it}^{\hat{j}}(\overline{S}_{it})}{\sigma_{maj}}\right)}, \quad t = T-1, \quad E_i = 4, \tag{20}$$

where $\mathfrak{e}$ is the Euler-Mascheroni's constant.

2.4 Post-secondary education choice $(t = 3)$: noncollege, two-year college, or four-year college

At the third period $t = 3$, the individual is choosing among not going to college, going to a two-year college, and going to a four-year college.

At the beginning of this stage, an individual i is characterized by the state variable vector $\overline{S}_{i3} = (\mathbf{\Theta}_{i3}, F_i, B_i, PG_i)$, where $\mathbf{\Theta}_{i3}$ is the test score vector that has built up through the high-school coursework.

The individual is choosing among three post-secondary education alternatives, i.e., $d_{i3} = e_i \in \{\text{noncollege, two-year college, four-year college}\}$, denoted by $d_{i3} \in \{0, 2, 4\}$ respectively.

Let $\epsilon_{i3}^e(coll) \in \boldsymbol{\epsilon}_{i3}(coll)$ denote the alternative-specific utility shocks that follow the standard extreme value type one distribution with a standard deviation of $\sigma_{coll} = 1$. The Bellman equation for choosing "noncollege", i.e., $d_{i3} = 0$ is

$$V_t^{d_{it}=0}(\overline{S}_{it}, \epsilon_{it}^{d_{it}=0}(coll)) = 0 + \mathbb{E}V_{t+1}(\overline{S}_{i,t+1} \mid \overline{S}_{it}, d_{it} = 0) + \epsilon_{it}^{d_{it}=0}(coll), \quad t = 3 \tag{21}$$

where $\overline{S}_{i,t+1} = (E_i = 0, M_i = n.a., \mathbf{\Theta}_{i3}, F_i, B_i)$. Note that there is no need to discount the continuation value term since choosing "noncollege" automatically updates the present period to the next period.

The Bellman equation for choosing to attend a two-year college $d_{i3} = 2$ is composed of a period utility $U_{i3}(\cdot)$ (to be defined later), a type one extreme value error term, and a continuation value, i.e.,

$$V_t^{d_{it}=2}(\overline{S}_{it}, \epsilon_{it}^{d_{it}=2}(coll)) = U_{it}(\overline{S}_{it}, d_{it}=2) + \delta^2 \mathbb{E}_{S_{i,t+1}|\overline{S}_{it}, d_{it}=2}\Big(V_{t+1}(S_{i,t+1} \mid \overline{S}_{it}, d_{it}=2)\Big) + \epsilon_{it}^{d_{it}=2}(coll), \quad t=3 \tag{22}$$

where $\overline{S}_{i,t+1} = (E_i = 2, M_i = n.a., \mathbf{\Theta}_{i3}, F_i, B_i)$ for occupation choice.

Similarly, the Bellman equation for choosing to attend a four-year college $d_{i3} = 4$ is

$$V_t^{d_{it}=4}(\overline{S}_{it}, \epsilon_{it}^{d_{it}=4}(coll)) = U_{it}(\overline{S}_{it}, d_{it}=4) + \delta^2 \mathbb{E}_{S_{i,t+1}|\overline{S}_{it}, d_{it}=4}\Big(V_{t+1}(S_{i,t+1} \mid \overline{S}_{it}, d_{it}=4)\Big) + \epsilon_{it}^{d_{it}=4}(coll), \quad t=3 \tag{23}$$

where $\overline{S}_{i,t+1} = (E_i = 4, \mathbf{\Theta}_{i3}, F_i, B_i)$ for college major choice.

Now let us define the period utility function of the individual choosing to attend a two-year or a four-year college. Let $\boldsymbol{\lambda}_{\Theta}^{e} = [\lambda^e(math), \lambda^e(sci), \lambda^e(mech), \lambda^e(verb)]'$ denote the skill effect on utility given a type of post-secondary choice e. Let λ_{female}^{e} and λ_{black}^{e} denote the female gender effect and the race effect, respectively, on utility given a type of post-secondary choice e. Let $\lambda_{parent2}^{e}$ and $\lambda_{parent3}^{e}$ denote the parents' education effect on utility given a type of post-secondary choice e. Let λ^e denote the post-secondary education choice fixed effect given e. The period utility function of the individual choosing a type of post-secondary education e can be written as

$$\begin{aligned} U_{i3}(\overline{S}_{i3}, d_{it}=e) = & \underbrace{\mathbf{\Theta}_{i3}\boldsymbol{\lambda}_{\Theta}^{e}}_{\text{skill effect}} + \underbrace{\mathbb{1}\{F_i=1\} \times \lambda_{female}^{e}}_{\text{female gender effect}} + \underbrace{\mathbb{1}\{B_i=1\} \times \lambda_{black}^{e}}_{\text{race effect}} \\ & + \underbrace{\mathbb{1}\{PG_i=2\} \times \lambda_{parent2}^{e} + \mathbb{1}\{PG_i=3\} \times \lambda_{parent3}^{e}}_{\text{parents' education effect}} \\ & + \underbrace{\lambda^e}_{\text{constant}}, \quad e \in \{2,4\}. \end{aligned} \tag{24}$$

The value function is,

$$V_t(\overline{S}_{it}, \boldsymbol{\epsilon}_{it}(coll)) = \max_{d_{it}} \sum_{e} \Big(V_t^e\big(\overline{S}_{it}, \epsilon_{it}^e(coll)\big) \times \mathbb{1}\{d_{it}=e\} \Big), \quad t=3. \tag{25}$$

Let $v_{it}^{e}(\overline{S}_{it})$ denote the Bellman equation $V_t^e(S_{it})$ net of the type one extreme value shock $\epsilon_{it}^{e}(coll)$. According to Rust (1987), the expected value function is

$$\mathbb{E}V_t(\overline{S}_{it}) = \sigma_{coll}\left\{\mathfrak{e} + \log\left[\sum_e \exp\left(\frac{v_{it}^{e}(\overline{S}_{it})}{\sigma_{coll}}\right)\right]\right\}, \quad t = 3, \tag{26}$$

and the choice probability is

$$\mathbb{P}(d_{it} = e \mid \overline{S}_{it}) = \frac{\exp\left(\frac{v_{it}^{e}(\overline{S}_{it})}{\sigma_{coll}}\right)}{\sum_{\hat{e}} exp\left(\frac{v_{it}^{\hat{e}}(\overline{S}_{it})}{\sigma_{coll}}\right)}, \quad t = 3, \tag{27}$$

where $\mathfrak{e}$ is the Euler-Mascheroni's constant, and

$$v_{it}^{e}(\overline{S}_{it}) = \begin{cases} \mathbb{E}V_{t+1}(\overline{S}_{i,t+1} \mid \overline{S}_{it}, d_{it} = e), & \text{if } e = 0 \\ U_{it}(\overline{S}_{it}, d_{it} = e) + \delta^2 \mathbb{E}\Big(V_{t+1}(\overline{S}_{i,t+1} \mid \overline{S}_{it}, d_{it} = e)\Big), & \text{if } e \in \{2, 4\}. \end{cases}$$

3 Estimation - Simulated Method of Moments

I use Simulated Method of Moments (SMM) for estimation. The detailed estimation strategy is specified in Appendix A. The summary statistics are specified in Table 1.

3.1 Weighting Matrix

To account for precision variation, I choose the weighting matrix to be a diagonal matrix, whose element is proportional to the inverse of the diagonal variance-covariance matrix of the empirical moments. The reason why I only make use of the diagonal elements of the variance-covariance matrix of the empirical moment instead of the whole matrix is due to the well-known problems associated with the use of the optimal weighting matrix (Altonji and Segal 1996).

Letting $\boldsymbol{\pi}$ denote the $N_{para} \times 1$ parameter vector, the estimation problem may be formally described as

$$\widehat{\boldsymbol{\pi}} = \arg\min_{\boldsymbol{\pi}} \left[\mathbf{m}_{sim}(\boldsymbol{\pi}) - \mathbf{m}_{data}\right]^{\top} \mathbf{W} \left[\mathbf{m}_{sim}(\boldsymbol{\pi}) - \mathbf{m}_{data}\right] \tag{28}$$

Table 1: Summary Statistics

	(1) Full Sample Male	(2) Full Sample Female	(3) D10 Male	(4) D10 Female	(5) D12 Male	(6) D12 Female
black	0.10	0.11	0.11	0.09	0.09	0.10
	(0.297)	(0.313)	(0.316)	(0.292)	(0.284)	(0.304)
neither parent completing $>$ 12 years	0.58	0.61	0.57	0.67	0.54	0.52
	(0.494)	(0.487)	(0.496)	(0.471)	(0.500)	(0.500)
$\geq$ 1 parent completing $>$ 12 years and neither $\geq$ 16	0.16	0.15	0.20	0.16	0.18	0.19
	(0.369)	(0.358)	(0.398)	(0.365)	(0.386)	(0.389)
$\geq$ 1 parent completing $\geq$ 16 years	0.26	0.24	0.23	0.17	0.28	0.30
	(0.437)	(0.425)	(0.424)	(0.377)	(0.451)	(0.457)
noncollege	0.59	0.52	0.59	0.54	0.55	0.43
	(0.492)	(0.500)	(0.493)	(0.499)	(0.499)	(0.496)
2-year college	0.08	0.13	0.07	0.16	0.06	0.15
	(0.274)	(0.342)	(0.257)	(0.368)	(0.240)	(0.355)
4-year college major: non-STEM	0.17	0.21	0.17	0.17	0.23	0.24
	(0.377)	(0.409)	(0.375)	(0.374)	(0.422)	(0.427)
4-year college major: pure-STEM	0.08	0.11	0.08	0.11	0.06	0.14
	(0.270)	(0.307)	(0.268)	(0.312)	(0.235)	(0.352)
4-year college major: applied-STEM	0.08	0.03	0.09	0.02	0.10	0.04
	(0.271)	(0.168)	(0.291)	(0.138)	(0.301)	(0.186)
Not working full-time	0.08	0.20	0.08	0.23	0.07	0.19
	(0.278)	(0.400)	(0.275)	(0.424)	(0.255)	(0.394)
full-time occupation: non-STEM	0.70	0.61	0.70	0.60	0.65	0.60
	(0.457)	(0.487)	(0.460)	(0.490)	(0.477)	(0.490)
full-time occupation: pure-STEM	0.07	0.13	0.09	0.10	0.07	0.14
	(0.259)	(0.337)	(0.293)	(0.296)	(0.258)	(0.352)
full-time occupation: applied-STEM	0.14	0.06	0.13	0.07	0.21	0.06
	(0.347)	(0.232)	(0.332)	(0.252)	(0.405)	(0.242)
# of G11 math courses taken	0.68	0.54	0.68	0.48	0.75	0.55
	(0.592)	(0.567)	(0.585)	(0.562)	(0.585)	(0.552)
# of G11 pure-STEM courses taken	0.92	0.87	0.92	0.78	1.02	0.92
	(0.870)	(0.876)	(0.862)	(0.884)	(0.944)	(0.888)
# of G11 applied-STEM courses taken	0.57	0.62	0.51	0.57	0.57	0.67
	(0.709)	(0.743)	(0.698)	(0.719)	(0.676)	(0.757)
# of G12 math courses taken	0.47	0.33	0.48	0.32	0.52	0.35
	(0.618)	(0.544)	(0.645)	(0.538)	(0.609)	(0.557)
# of G12 pure-STEM courses taken	1.37	1.25	1.34	1.19	1.44	1.26
	(1.074)	(1.055)	(1.032)	(1.091)	(1.050)	(1.010)
# of G12 applied-STEM courses taken	0.48	0.39	0.49	0.35	0.50	0.41
	(0.692)	(0.629)	(0.692)	(0.604)	(0.739)	(0.618)
math score	0.30	0.13	0.29	0.05	0.52	0.22
	(1.002)	(0.926)	(0.985)	(0.935)	(0.998)	(0.940)
science score	0.39	0.02	0.33	-0.04	0.51	0.11
	(0.936)	(0.851)	(0.941)	(0.901)	(0.855)	(0.790)
mechanical score	0.65	-0.29	0.63	-0.35	0.80	-0.23
	(0.951)	(0.710)	(0.949)	(0.712)	(0.931)	(0.718)
verb score	0.21	0.29	0.23	0.22	0.33	0.39
	(0.900)	(0.780)	(0.914)	(0.813)	(0.782)	(0.680)
Observations	1364	1450	389	386	242	326

Notes: Mean coefficients; sd in parentheses. "**D10** (**D12**)" denotes the sample containing the observa--tions of those taking the test with the 10th (12th) grade completed.

where $\mathbf{m}_{data}$ is the $N_{mom} \times 1$ vector of empirical moments, and $\mathbf{m}_{sim}(\boldsymbol{\pi})$ is the model moment vector given $\boldsymbol{\pi}$.

The variance-covariance matrix of the estimator is given by

$$\left[\mathbf{D}_{\mathbf{m}}^{\top}\mathbf{W}\mathbf{D}_{\mathbf{m}}\right]^{-1}\mathbf{D}_{\mathbf{m}}^{\top}\mathbf{W}\Sigma\mathbf{W}^{\top}\mathbf{D}_{\mathbf{m}}\left[\mathbf{D}_{\mathbf{m}}^{\top}\mathbf{W}\mathbf{D}_{\mathbf{m}}\right]^{-1}, \tag{29}$$

where Σ is the $N_{mom} \times N_{mom}$ variance-covariance matrix of the empirical moments, and $\mathbf{D_m} = \partial \mathbf{m}_{sim}(\boldsymbol{\pi}))/\partial \boldsymbol{\pi}$ is the $N_{mom} \times N_{para}$ derivative matrix of the moment conditions with respect to the model parameters at $\boldsymbol{\pi} = \widehat{\boldsymbol{\pi}}$.

3.2 Estimation Results

The estimation results are displayed in Tables 2 through 7. Below I will delve into the estimation results and emphasize the key findings in each model period.

High-school Course-taking Preferences The estimation results for high-school course-taking parameters exhibit three key features: (1) a strong preference effect, (2) a positive influence of math skills on course enrollments, and (3) a negative impact of mechanical skills.

Detailed results can be seen in Table 2, which presents the estimation results for utility parameters related to high-school course-taking behavior, with skill scores normalized between 0 and 10 for the sake of interpretation. One of the most significant findings is the substantial impact of gender compared to other factors such as skills and family background. The tables reveal that females are less likely than males to enroll in any of the three course types: math, pure-STEM, and applied-STEM. Additionally, the results show that math skills have a positive influence on enrollment in all three course types, while science skills positively impact the selection of math and pure-STEM courses exclusively. Conversely, mechanical skills exhibit a negative influence on enrollment in all three course types. In simpler terms, higher proficiency in math facilitates enrollment in any of the three course types, while increased mechanical skills result in reluctance to pursue any courses.

Given that the utility function in the last period where consumption occurs is in log form, the interpretation of the utility parameter estimates is as follows: a coefficient denoted by X implies that taking one additional course, in conjunction with a one-unit increase in the interaction term, affects utility as if the individual's consumption level is multiplied by e^X. For example, as shown in Table 2, in Grade 11, taking one more math course along with a

Table 2: Estimation Results of High School Course Choice Utility Parameters

		Grade 11 Est.	Std. Err.	Grade 12 Est.	Std. Err.
Math crs × math skill	$\kappa_{\theta H}^{math}(math)$	0.400	0.080	1.128	0.323
× science skill	$\kappa_{\theta H}^{math}(sci)$	0.161	0.085	1.443	0.352
× mechanical skill	$\kappa_{\theta H}^{math}(mech)$	-0.581	0.099	-1.483	0.309
× verb skill	$\kappa_{\theta H}^{math}(verb)$	-0.011	0.090	-1.312	0.381
× female	κ_{fh}^{math}	-1.635	0.223	-3.158	0.597
× black	κ_{bH}^{math}	-0.281	0.160	-1.444	0.259
× ≥ 1 parent completing $(12, 16)$ and neither ≥ 16 years of edu.	κ_{pH2}^{math}	-0.080	0.187	0.048	0.340
× ≥ 1 parent completing ≥ 16 years of edu.	κ_{pH3}^{math}	0.432	0.142	0.299	0.318
Pure-STEM crs × math skill	$\kappa_{\theta H}^{ps}(math)$	0.121	0.053	0.069	0.044
× science skill	$\kappa_{\theta H}^{ps}(sci)$	0.216	0.063	0.167	0.051
× mechanical skill	$\kappa_{\theta H}^{ps}(mech)$	-0.408	0.066	-0.161	0.052
× verb skill	$\kappa_{\theta H}^{ps}(verb)$	0.070	0.064	-0.090	0.057
× female	κ_{fh}^{ps}	-1.157	0.161	-0.432	0.122
× black	κ_{bH}^{ps}	-0.387	0.121	-0.266	0.080
× ≥ 1 parent completing $(12, 16)$ and neither ≥ 16 years of edu.	κ_{pH2}^{ps}	0.002	0.139	-0.032	0.111
× ≥ 1 parent completing ≥ 16 years of edu.	κ_{pH3}^{ps}	-0.063	0.127	0.119	0.109
Applied-STEM crs × math skill	$\kappa_{\theta H}^{as}(math)$	0.157	0.082	0.099	0.084
× science skill	$\kappa_{\theta H}^{as}(sci)$	-0.144	0.096	0.004	0.088
× mechanical skill	$\kappa_{\theta H}^{as}(mech)$	-0.243	0.081	-0.251	0.077
× verb skill	$\kappa_{\theta H}^{as}(verb)$	0.148	0.094	0.050	0.096
× female	κ_{fh}^{as}	-0.733	0.209	-1.190	0.223
× black	κ_{bH}^{as}	-0.522	0.161	-0.575	0.174
× ≥ 1 parent completing $(12, 16)$ and neither ≥ 16 years of edu.	κ_{pH2}^{as}	0.085	0.209	-0.038	0.202
× ≥ 1 parent completing ≥ 16 years of edu.	κ_{pH3}^{as}	-0.076	0.232	-0.249	0.225

one-unit increase in the math score decile has an impact on utility akin to multiplying the individual's consumption by $e^{0.400} \approx 1.492$.

Skill Formation The results obtained when estimating the parameters of skill evolution reveal one particularly significant feature: a substantial impact of taking pure-STEM courses on skill accumulation.

Detailed results can be seen in Table 3, which presents the estimates of the parameters governing high-school skill development based on value-added production functions. Skill scores are standardized relative to the population mean, with a value of $+1$ representing one standard deviation above the mean. It is evident that all four skills primarily evolve due to their inherent persistence. In terms of the course effects, all three course types exhibit positive impacts on all four skills. However, pure-STEM courses exert the most substantial influence on all skills, except for science and mechanical skill production, where taking

applied-STEM courses has the greatest course effect. It is noteworthy that choosing pure-STEM courses significantly enhances the production of all skills. To put it in perspective, taking an additional pure-STEM course results in approximately a $0.1 - 0.2$ times increase in skill, equivalent to one standard deviation, while enrolling in other STEM courses yields less than a 0.1 times increase.

Table 3: Select Estimation Results of High School Skill Evolution Parameters

		Estimate	Standard Error
Math skill intercept	$\gamma_0(math)$	-0.163	0.164
persistence	$\gamma_\theta(math)$	0.868	0.057
math course effect	$\gamma_h^{math}(math)$	0.067	0.133
pure-STEM course effect	$\gamma_h^{ps}(math)$	0.175	0.101
applied-STEM course effect	$\gamma_h^{as}(math)$	0.093	0.186
Science skill intercept	$\gamma_0(sci)$	-0.106	0.172
persistence	$\gamma_\theta(sci)$	0.852	0.053
math course effect	$\gamma_h^{math}(sci)$	0.081	0.124
pure-STEM course effect	$\gamma_h^{ps}(sci)$	0.098	0.108
applied-STEM course effect	$\gamma_h^{as}(sci)$	0.128	0.174
Mechanical skill intercept	$\gamma_0(mech)$	-0.160	0.159
persistence	$\gamma_\theta(mech)$	0.969	0.032
math course effect	$\gamma_h^{math}(mech)$	0.070	0.121
pure-STEM course effect	$\gamma_h^{ps}(mech)$	0.119	0.098
applied-STEM course effect	$\gamma_h^{as}(mech)$	0.121	0.174
Verbal skill intercept	$\gamma_0(verb)$	-0.144	0.150
persistence	$\gamma_\theta(verb)$	0.794	0.057
math course effect	$\gamma_h^{math}(verb)$	0.038	0.120
pure-STEM course effect	$\gamma_h^{ps}(verb)$	0.193	0.088
applied-STEM course effect	$\gamma_h^{as}(verb)$	0.082	0.173

Post-secondary Education Preferences The results obtained from estimating the parameters of post-secondary education choice reveal several noteworthy findings: (1) a significantly positive effect of parental education on both types of post-secondary education choice, (2) a notably negative impact of mechanical skills on both types of post-secondary education choice, (3) positive effects of math and science skills on selecting a two-year college education, and (4) positive effects of math and verbal skills on opting for a four-year college education.

Detailed results can be found in Table 4, which displays the estimates of the utility parameters concerning the choice of post-secondary education. The utility of not opting for

any form of post-secondary education is standardized to be 0, and skill scores are standardized relative to the population mean, with $+1$ representing one standard deviation above the mean.

Parents' education has significantly positive effects on pursuing either type of post-secondary education, especially four-year college education. More specifically, an individual having at least one parent completing more than 16 years of education is $e^{1.591} \approx 4.909$ ($e^{0.408} \approx 1.504$) times more likely to get a four-year (two-year) college degree as compared to getting no post-secondary education degree at all.[7] Regarding getting a four-year college degree, both math and verbal skills have significantly positive effects, i.e., a one-standard-deviation unit increase in math scores renders the individual $e^{1.656} \approx 5.238$ times more likely, and verbal scores render $e^{0.499} \approx 1.647$ times more likely than not obtaining a college degree. Regarding obtaining a two-year college degree, science skills have a significantly positive effect (i.e., a one-standard-deviation unit increase renders $e^{0.472} \approx 1.603$ times more likely than not obtaining a college degree), while mechanical skills have a negative effect (i.e., a one-standard-deviation unit increase renders $e^{-0.282} \approx 0.754$ times more likely than not obtaining a college degree). The fact that mechanical skills undermine an individual's odds of continuing with post-secondary education relates to the fact that mechanical skills are valued in the labor market regardless of whether one has a college degree.[8]

Four-year College Major Preferences The results acquired from estimating the parameters of four-year college major choice unveil three notable findings: (1) a significantly positive effect of science skills on the likelihood of getting a pure-STEM degree, (2) the positive effects of math, science, and mechanical skills on getting an applied-STEM degree, and (3) the negative impact of verbal skills on getting either type of the STEM degree.

The detailed results can be seen in Table 5, which contains the estimates of the utility parameters for the choice of four-year college degrees, conditional on one having a four-year college degree. The utility of choosing a non-STEM college degree is regularized to be

[7] The coefficients are interpreted as (approximately) the percent change in the risk or odds, i.e., the ratio of the probability of staying in the category of interest to that of staying in the base category if the variable increases by one unit while all other ones are held constant. In other words, the exponentials of the coefficients represent how many times as likely to stay in the category of interest as compared to the base category if the variable increases by one unit while all other ones are held constant.

[8] See Chuan and Zhang (2021) for more details.

Table 4: Estimation Results of Post-secondary Education Parameters

		Estimate	Standard Error
($<$ 4-year college) math skill	$\lambda^{2yr}(math)$	0.377	0.062
science skill	$\lambda^{2yr}(sci)$	0.472	0.081
mechanical skill	$\lambda^{2yr}(mech)$	-0.282	0.115
verbal skill	$\lambda^{2yr}(verb)$	-0.104	0.091
female	λ^{2yr}_{female}	0.700	0.201
black	λ^{2yr}_{black}	0.172	0.164
$\geq$ 1 parent completing $(12, 16)$ and neither $\geq$ 16 years of edu.	$\lambda^{2yr}_{parent2}$	0.369	0.180
$\geq$ 1 parent completing $\geq$ 16 years of edu.	$\lambda^{2yr}_{parent3}$	0.408	0.201
constant	λ^{2yr}	-2.118	0.149
(4-year college) math skill	$\lambda^{4yr}(math)$	1.656	0.101
science skill	$\lambda^{4yr}(sci)$	-0.258	0.176
mechanical skill	$\lambda^{4yr}(mech)$	-0.706	0.205
verbal skill	$\lambda^{4yr}(verb)$	0.499	0.166
female	λ^{4yr}_{female}	0.103	0.235
black	λ^{4yr}_{black}	0.004	0.208
$\geq$ 1 parent completing $(12, 16)$ and neither $\geq$ 16 years of edu.	$\lambda^{4yr}_{parent2}$	0.605	0.171
$\geq$ 1 parent completing $\geq$ 16 years of edu.	$\lambda^{4yr}_{parent3}$	1.591	0.165
constant	λ^{4yr}	-2.802	0.193

0. Math skills negatively impact one's odds of getting a pure-STEM degree and positively affect one's odds of getting an applied-STEM degree. More specifically, a one-standard-deviation increase in math skills undermines one's odds of getting a pure-STEM four-year college major by $e^{-0.618} \approx 0.539$ times and boosters one's odds of getting an applied-STEM four-year college major by $e^{0.256} \approx 1.292$ times, both in reference to getting a non-STEM four-year college major.

Science skills and mechanical skills positively impact one's odds of getting a pure-STEM degree and an applied-STEM degree, respectively. More specifically, a one-standard-deviation increase in science skills is particularly important in bolstering one's odds of obtaining a pure-STEM four-year college major (i.e., by $e^{1.105} \approx 3.019$ times), and so is a one-standard-deviation increase in mechanical skills in increasing one's odds of achieving an applied-STEM 4-year college major (i.e., by $e^{0.944} \approx 2.570$ times).

It is noticeable that verbal skills have a negative effect on the odds of getting either type of the STEM degrees with respect to the non-STEM one. In particular, a one-standard-deviation increase in verbal skills undermines one's odds of getting an applied-STEM degree

by $e^{-1.348} \approx 0.260$ times. It is also found in the estimation that black females have the highest likelihood of entering an applied-STEM major in comparison to the other three gender-race demographic groups, ceteris paribus.

Table 5: Select Estimation Results of Four-year College Major Parameters

		Estimate	Standard Error
(pure-STEM 4-year college major) math skill	$\eta^{ps}(math)$	-0.618	0.124
science skill	$\eta^{ps}(sci)$	1.105	0.133
mechanical skill	$\eta^{ps}(mech)$	-0.049	0.161
verbal skill	$\eta^{ps}(verb)$	-0.508	0.217
female	η^{ps}_{female}	1.062	0.156
black	η^{ps}_{black}	2.000	0.186
constant	η^{ps}	-2.040	0.214
(applied-STEM 4-year college major) math skill	$\eta^{as}(math)$	0.256	0.163
science skill	$\eta^{as}(sci)$	0.487	0.169
mechanical skill	$\eta^{as}(mech)$	0.944	0.154
verbal skill	$\eta^{as}(verb)$	-1.348	0.214
female	η^{as}_{female}	-0.101	0.177
black	η^{as}_{black}	1.682	0.226
constant	η^{as}	-1.779	0.232

Wage Offer Functions The results derived from estimating the parameters of wage offer functions unveil a couple of noteworthy findings. Firstly, the hierarchy of occupations in terms of base salary, from lowest to highest, is as follows: non-STEM, pure-STEM, and applied-STEM. Secondly, the hierarchy of occupations in terms of the female wage penalty, from lowest to highest, is in the opposite order compared to the base salary: applied-STEM, pure-STEM, and non-STEM. Detailed results are presented in Table 6, which contains the estimates of wage offer parameters. The wage represents an individual's annual labor income in $10,000$ US dollars in 2020.

There are three sets of wage offer parameters for each type of occupation: non-STEM, pure-STEM, and applied-STEM. The base wage offer is relative to a white or Hispanic male with only a high-school degree. Notably, applied-STEM occupations offer the highest base salary, pure-STEM occupations offer the second-highest, and non-STEM offers the lowest. In any occupation category, being either female or black has a negative impact on wage offers, while holding various degrees - two-year college degrees, four-year non-STEM degrees, four-year pure-STEM degrees, and four-year applied-STEM degrees - has positive effects on

wage offers. Although two-year degrees have a relatively small positive effect on wage offers in non-STEM occupations, resulting in a 3.3% wage increase, they have a more substantial impact in pure-STEM occupations, with a 55.6% wage increase, and in applied-STEM occupations, with a 25.5% wage increase. It's interesting to note that the effect of female gender on wages is smallest in applied-STEM occupations, where female representation is the lowest and the base salary is the highest.

Table 6: Select Estimation Results of Occupation Pecuniary Parameters

		Estimate	Standard Error
(non-STEM occupation) female	β^{ns}_{female}	-0.506	0.052
black	β^{ns}_{black}	-0.161	0.057
2-year degree	β^{ns}_{2yr}	0.033	0.082
4-year non-STEM degree	$\beta^{ns}_{4yr,ns}$	0.516	0.061
4-year pure-STEM degree	$\beta^{ns}_{4yr,ps}$	0.360	0.105
4-year applied-STEM degree	$\beta^{ns}_{4yr,as}$	0.293	0.176
constant	β^{ns}	0.713	0.036
(pure-STEM occupation) female	β^{ps}_{female}	-0.399	0.137
black	β^{ps}_{black}	-0.444	0.124
2-year degree	β^{ps}_{2yr}	0.556	0.167
4-year non-STEM degree	$\beta^{ps}_{4yr,ns}$	0.738	0.127
4-year pure-STEM degree	$\beta^{ps}_{4yr,ps}$	0.389	0.157
4-year applied-STEM degree	$\beta^{ps}_{4yr,as}$	0.326	0.252
constant	β^{ps}	0.766	0.095
(applied-STEM occupation) female	β^{as}_{female}	-0.194	0.077
black	β^{as}_{black}	-0.583	0.120
2-year degree	β^{as}_{2yr}	0.255	0.132
4-year non-STEM degree	$\beta^{as}_{4yr,ns}$	0.329	0.092
4-year pure-STEM degree	$\beta^{as}_{4yr,ps}$	0.427	0.128
4-year applied-STEM degree	$\beta^{as}_{4yr,as}$	0.423	0.096
constant	β^{as}	1.211	0.075

Non-pecuniary Benefit Preferences The estimates of utility parameters for non-pecuniary benefits offer three significant insights. Firstly, females tend to have a preference for working in gender-balanced or female-dominated occupations, whereas males benefit from workplace gender diversity with no strong distaste for working in a gender-imbalanced occupation. Secondly, both pure-STEM and applied-STEM occupations incur substantial and comparable occupation-major incompatibility costs. Thirdly, the incompatibility costs associated with non-STEM occupations vary based on educational status. Notably, only four-

year college applied-STEM majors transitioning into non-STEM occupations experience a significant incompatibility cost, which is less severe compared to the incompatibility cost of the two STEM occupation types.[9]

Table 7 provides the non-pecuniary benefit estimates related to workplace gender composition and occupation-major incompatibility. Additionally, Figure 1 illustrates the corresponding hyperbolas. The plot reveals that both male hyperbolas are positioned to the right of the female hyperbolas, in line with the findings from Schuh (2022), indicating that males tend to prefer workplace gender diversity while females favor being in a workplace where their gender is the majority.

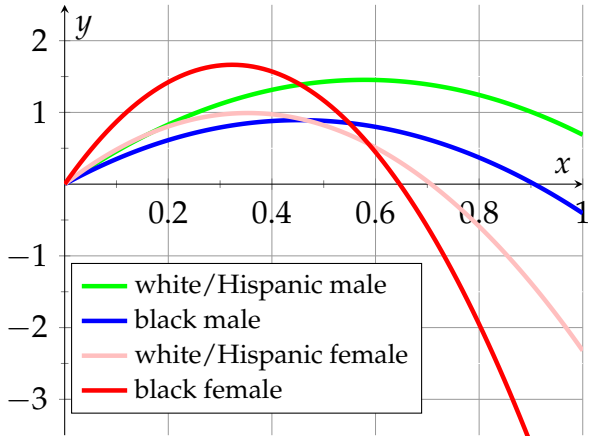
Table 7: Estimation Results of Occupation Non-Pecuniary Parameters

		Estimate	Standard Error
Nonblack Male preference for female share, linear	$\alpha^{nb}_{male,1}$	5.019	1.008
quadratic	$\alpha^{nb}_{male,2}$	-4.330	1.744
Black male preference for female share, linear	$\alpha^{b}_{male,1}$	3.937	1.384
quadratic	$\alpha^{b}_{male,2}$	-4.343	2.449
Nonblack female preference for male share, linear	$\alpha^{nb}_{female,1}$	5.608	0.990
quadratic	$\alpha^{nb}_{female,2}$	-7.930	1.652
Black female preference for male share, linear	$\alpha^{b}_{female,1}$	10.293	1.283
quadratic	$\alpha^{b}_{female,2}$	-15.916	2.047
Occ-major incompatibility, non-STEM occ., $<$ 4-year college	$\alpha^{ns}_{maj\neq occ}(<4yr)$	0.061	0.138
4-year college pure-STEM major	$\alpha^{ns}_{maj\neq occ}(4yr_ps)$	0.516	0.215
4-year college applied-STEM major	$\alpha^{ns}_{maj\neq occ}(4yr_as)$	-1.036	0.339
Occ-major incompatibility, pure-STEM occ.	$\alpha^{ps}_{maj\neq occ}$	-2.656	0.211
Occ-major incompatibility, applied-STEM occ.	$\alpha^{as}_{maj\neq occ}$	-2.404	0.208

From the estimations of the occupation-major incompatibility parameters, it is evident that substantial costs associated with occupation-major incompatibility exist in both pure-STEM and applied-STEM occupations. To elaborate further, let us consider specific cases. When comparing an individual holding a four-year college pure-STEM (applied-STEM) major to someone lacking such a degree but attempting to enter the same field, a significant penalty becomes apparent for the latter. This penalty can be analogously understood as a reduction in their annual consumption by a factor of $e^{-2.656} \approx 0.070 (e^{-2.404} \approx 0.090)$. Additionally, when comparing an individual with a four-year college non-STEM major to

[9]It is worth mentioning that occupation-major incompatibility parameters differ with educational status in non-STEM occupations but remain consistent in pure- and applied-STEM occupations. This distinction arises from the greater diversity of non-STEM occupations, which necessitates a more nuanced classification of job entrants.

Figure 1: Plots for Non-Pecuniary Utility Parameters



an applied-STEM major endeavoring to enter the non-STEM field, a similar penalty cost emerges for the latter, resulting in a reduction of their annual consumption by a factor of $e^{-1.036} \approx 0.355$. What is particularly intriguing is that there appears to be no cost but rather a reward for individuals with educational levels below a four-year college education or those holding a four-year college pure-STEM major when they seek entry into the non-STEM field. This counterintuitive result may be attributed to various factors, such as the evolving demands of the labor market, the transferability of skills across disciplines, or the changing landscape of non-STEM occupations. Further research and analysis are required to provide a comprehensive understanding of this phenomenon.

4 Counterfactual Results

This section provides a description of the results obtained from four exercises: (1) the assessment of the importance of gender gap sources, which analyzes gender gaps into those stemming from different sources such as skill gaps or preference gaps, (2) the analysis of the gender composition effect, which examines both genders' preferences for occupational gender composition, and two policy experiments, i.e., (3) wage policy experiment, where simulations are implemented under different wage policies, and (4) course requirement policy experiment, where simulations are done under different coursework policies. The first two exercises aim to analyze gender gaps into those stemming from different sources, and the last two policy experiments explore ways to close gender gaps and boost STEM partici-

pation. Below I will provide a detailed explanation of the results for each exercise.

4.1 Effects of Gender Disparities in Skills and Preferences

In Tables 8-13, Rows 1-2 show the data moments by gender. Rows 3-4 show the simulation moments by gender, which are also the baseline for comparative analysis. Rows 5 - 10 present the results of six counterfactuals: the simulation results of female individuals (1) if their initial math skill levels are shifted to where their means are equal to those of males (Row 5), (2) if their initial science skill levels are shifted to where their means are equal to those of males (Row 6), (3) if their initial mechanical skill levels are shifted to where their means are equal to those of males (Row 7), (4) if their initial verbal skill levels are shifted to where their means are equal to those of males (Row 8), (5) if their initial skill levels, math, science, mechanical, and verbal, are all shifted to where their means are equal to those of males (Row 9), and (6) if the female individuals have the same preference as that of those male individuals (Row 10). Next to the column of each choice, Tables 8-13 also display the $male - female$ gender gaps Δ_{M-F} in Columns 2,5, and 8 and their corresponding percentage change Δ in Columns 3, 6, and 9.

Regarding counterfactual choice distributions, Table 8 shows the counterfactual results of the numbers of STEM courses taken in the $11th$ and $12th$ grades combined, for which the simulations are implemented on the individuals observed to complete the $10th$ grade upon the ASVAB test (**D10**). Tables 9, 10, and 13 show the counterfactual results of college major choice distribution, occupation choice distribution, and post-secondary education choice distribution, respectively, for which the simulations are implemented on the individuals observed to complete the $10th$ grade upon the ASVAB test (**D10**). Tables 11 and 12 show the counterfactual results of college major choice distribution and occupation choice distribution, respectively, for which the simulations are implemented on the individuals observed to complete the $12th$ grade upon the ASVAB test (**D12**).

The findings from Exercises (1) shed light on the evolving roles of skill disparities and choice-specific preference gaps in driving gender disparities across the decision-making journey, spanning from high-school course selection to occupational choices. Notably, as individuals progress through various stages of education and their careers, the influence of choice-specific preference gaps diminishes, while the impact of skill disparities becomes

Table 8: Counterfactual of # of Courses Taken in Grades 11 & 12 (Starting from Grade 10)

	math	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	1.159	-	-	2.259	-	-	0.997	-	-
female (data)	0.800	0.359	-	1.966	0.293	-	0.920	0.077	-
male (baseline)	1.240	-	-	2.493	-	-	1.123	-	-
female (baseline)	0.604	0.636	-	1.794	0.700	-	0.753	0.370	-
female, with male MATH skill dist.	0.747	0.493	22.410	1.874	0.619	11.493	0.801	0.322	13.077
	[0.587 0.906]	-	-	[1.645 2.104]	-	-	[0.609 0.994]	-	-
female, with male SCI skill dist.	0.794	0.446	29.897	2.013	0.480	31.339	0.697	0.426	-14.984
	[0.794]	-	-	[2.013]	-	-	[0.697]	-	-
female, with male MECH skill dist.	0.144	1.096	-72.418	1.035	1.459	-108.513	0.451	0.672	-81.463
	[0.059 0.228]	-	-	[0.759 1.31]	-	-	[0.238 0.664]	-	-
female, with male VERB skill dist.	0.602	0.638	-0.407	1.794	0.700	0.017	0.754	0.369	0.363
	[0.476 0.727]	-	-	[1.584 2.004]	-	-	[0.582 0.926]	-	-
female, with male FULL skill dist.	0.248	0.992	-56.069	1.269	1.224	-75.003	0.460	0.663	-79.097
	[0.125 0.37]	-	-	[0.977 1.561]	-	-	[0.244 0.676]	-	-
female, with male preference	2.046	-0.806	226.846	3.277	-0.784	212.036	1.661	-0.537	245.102
	[1.514 2.579]	-	-	[2.854 3.7]	-	-	[1.253 2.068]	-	-

Notes: The columns denote the types of high-school STEM coursework - math, pure-STEM, and applied-STEM.

Table 9: Counterfactual of Four-year College Major Choice (Starting from Grade 10, Absolute Measures)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.187	-	-	0.044	-	-	0.083	-	-
female (data)	0.196	-0.009	-	0.112	-0.068	-	0.028	0.054	-
male (baseline)	0.194	-	-	0.051	-	-	0.084	-	-
female (baseline)	0.193	0.000	-	0.113	-0.062	-	0.028	0.056	-
female, with male MATH skill dist.	0.225	-0.031	8051.433	0.120	-0.069	-11.722	0.034	0.050	10.897
	[0.194 0.256]	-	-	[0.102 0.139]	-	-	[0.025 0.043]	-	-
female, with male SCI skill dist.	0.178	0.015	-3837.751	0.136	-0.085	-37.290	0.029	0.055	1.597
	[0.154 0.202]	-	-	[0.109 0.164]	-	-	[0.022 0.036]	-	-
female, with male MECH skill dist.	0.121	0.072	-18309.112	0.069	-0.017	71.844	0.038	0.046	17.476
	[0.077 0.166]	-	-	[0.051 0.087]	-	-	[0.012 0.064]	-	-
female, with male VERB skill dist.	0.193	0.000	52.765	0.113	-0.062	-0.184	0.028	0.056	-0.295
	[0.165 0.221]	-	-	[0.096 0.131]	-	-	[0.021 0.035]	-	-
female, with male FULL skill dist.	0.134	0.060	-15139.696	0.092	-0.040	34.589	0.049	0.035	37.603
	[0.092 0.176]	-	-	[0.07 0.114]	-	-	[0.02 0.078]	-	-
female, with male preference	0.270	-0.076	19413.563	0.071	-0.019	68.892	0.043	0.041	26.801
	[0.167 0.372]	-	-	[0.045 0.096]	-	-	[0.019 0.067]	-	-

Notes: The columns denote the types of four-year college majors - non-STEM, pure-STEM, and applied-STEM.

Table 10: Counterfactual of Occupation Choice (Starting from Grade 10)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.696	-	-	0.095	-	-	0.126	-	-
female (data)	0.602	0.094	-	0.095	-0.000	-	0.067	0.059	-
male (baseline)	0.716	-	-	0.075	-	-	0.126	-	-
female (baseline)	0.622	0.094	-	0.119	-0.045	-	0.057	0.069	-
female, with male MATH skill dist.	0.621	0.095	-1.051	0.122	-0.048	-6.200	0.059	0.066	3.475
	[0.588 0.654]	-	-	[0.101 0.143]	-	-	[0.042 0.077]	-	-
female, with male SCI skill dist.	0.617	0.099	-5.568	0.128	-0.053	-18.325	0.057	0.069	-0.353
	[0.583 0.65]	-	-	[0.106 0.149]	-	-	[0.04 0.074]	-	-
female, with male MECH skill dist.	0.624	0.092	2.456	0.101	-0.026	41.633	0.062	0.064	7.203
	[0.588 0.66]	-	-	[0.08 0.122]	-	-	[0.041 0.083]	-	-
female, with male VERB skill dist.	0.622	0.094	0.102	0.119	-0.045	0.125	0.057	0.069	-0.124
	[0.589 0.655]	-	-	[0.099 0.139]	-	-	[0.04 0.073]	-	-
female, with male FULL skill dist.	0.614	0.102	-7.936	0.110	-0.036	20.573	0.067	0.059	14.868
	[0.578 0.651]	-	-	[0.089 0.131]	-	-	[0.045 0.09]	-	-
female, with male preference	0.629	0.087	7.565	0.104	-0.029	34.877	0.066	0.060	12.721
	[0.592 0.666]	-	-	[0.084 0.124]	-	-	[0.044 0.088]	-	-

Notes: The columns denote occupation types - non-STEM, pure-STEM, and applied-STEM.

Table 11: Counterfactual of Four-year College Major Choice (Starting from Grade 12, Absolute Measures)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.206	-	-	0.050	-	-	0.110	-	-
female (data)	0.230	-0.024	-	0.130	-0.080	-	0.031	0.079	-
male (baseline)	0.199	-	-	0.052	-	-	0.106	-	-
female (baseline)	0.234	-0.035	-	0.127	-0.075	-	0.030	0.077	-
female, with male MATH skill dist.	0.290	-0.090	-158.250	0.133	-0.082	-8.341	0.040	0.067	12.879
	[0.271 0.308]	-	-	[0.121 0.146]	-	-	[0.034 0.045]	-	-
female, with male SCI skill dist.	0.199	0.000	100.405	0.164	-0.113	-49.629	0.031	0.076	1.138
	[0.185 0.214]	-	-	[0.142 0.187]	-	-	[0.026 0.036]	-	-
female, with male MECH skill dist.	0.165	0.034	198.390	0.082	-0.030	59.478	0.055	0.051	33.215
	[0.131 0.199]	-	-	[0.059 0.105]	-	-	[0.035 0.076]	-	-
female, with male VERB skill dist.	0.229	-0.029	16.208	0.128	-0.076	-0.841	0.032	0.075	2.628
	[0.215 0.242]	-	-	[0.116 0.139]	-	-	[0.029 0.035]	-	-
female, with male FULL skill dist.	0.172	0.027	177.105	0.116	-0.064	15.086	0.081	0.026	66.663
	[0.146 0.198]	-	-	[0.087 0.145]	-	-	[0.054 0.108]	-	-
female, with male preference	0.268	-0.068	-95.765	0.067	-0.015	80.520	0.035	0.071	7.025
	[0.252 0.284]	-	-	[0.06 0.073]	-	-	[0.032 0.038]	-	-

Notes: The columns denote the types of four-year college majors - non-STEM, pure-STEM, and applied-STEM.

Table 12: Counterfactual of Occupation Choice (Starting from Grade 12)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.654	-	-	0.071	-	-	0.205	-	-
female (data)	0.602	0.052	-	0.144	-0.073	-	0.062	0.143	
male (baseline)	0.708	-	-	0.074	-	-	0.138	-	-
female (baseline)	0.618	0.090	-	0.124	-0.051	-	0.058	0.080	-
female, with male MATH skill dist.	0.616	0.092	-2.162	0.127	-0.054	-5.849	0.063	0.075	6.120
	[0.585 0.648]	-	-	[0.106 0.148]	-	-	[0.045 0.08]	-	-
female, with male SCI skill dist.	0.608	0.100	-11.662	0.139	-0.065	-28.507	0.058	0.080	-0.629
	[0.576 0.639]	-	-	[0.116 0.161]	-	-	[0.041 0.075]	-	-
female, with male MECH skill dist.	0.612	0.096	-7.553	0.106	-0.033	35.947	0.072	0.066	17.132
	[0.577 0.646]	-	-	[0.085 0.127]	-	-	[0.051 0.093]	-	-
female, with male VERB skill dist.	0.617	0.091	-1.595	0.124	-0.051	0.123	0.059	0.078	1.458
	[0.586 0.648]	-	-	[0.104 0.145]	-	-	[0.042 0.076]	-	-
female, with male FULL skill dist.	0.594	0.114	-27.058	0.119	-0.046	9.999	0.083	0.055	31.398
	[0.559 0.63]	-	-	[0.098 0.141]	-	-	[0.058 0.108]	-	-
female, with male preference	0.628	0.080	10.878	0.102	-0.029	43.719	0.064	0.074	6.780
	[0.597 0.659]	-	-	[0.083 0.121]	-	-	[0.046 0.081]	-	-

Notes: The columns denote occupation types - non-STEM, pure-STEM, and applied-STEM.

Table 13: Counterfactual of Post-secondary Education Choice (Starting from Grade 10)

	high sch. grad.	Δ_{M-F}	%	**two-year coll.**	Δ_{M-F}	%	**four-year coll.**	Δ_{M-F}	%
male (data)	0.611	-	-	0.076	-	-	0.314	-	-
female (data)	0.531	0.079	-	0.132	-0.057	-	0.337	-0.023	-
male (baseline)	0.594	-	-	0.077	-	-	0.329	-	-
female (baseline)	0.529	0.066	-	0.137	-0.060	-	0.334	-0.005	-
female, with male MATH skill dist.	0.487	0.107	-63.538	0.134	-0.057	5.551	0.379	-0.050	-850.570
	[0.444 0.529]	-	-	[0.12 0.147]	-	-	[0.335 0.423]	-	-
female, with male SCI skill dist.	0.506	0.088	-33.897	0.151	-0.074	-22.148	0.343	-0.014	-167.730
	[0.466 0.547]	-	-	[0.136 0.165]	-	-	[0.302 0.384]	-	-
female, with male MECH skill dist.	0.645	-0.051	177.463	0.127	-0.051	16.254	0.228	0.101	2013.884
	[0.585 0.704]	-	-	[0.105 0.15]	-	-	[0.166 0.289]	-	-
female, with male VERB skill dist.	0.528	0.066	-0.106	0.137	-0.060	0.142	0.334	-0.005	-2.929
	[0.487 0.569]	-	-	[0.124 0.15]	-	-	[0.293 0.375]	-	-
female, with male FULL skill dist.	0.586	0.008	87.665	0.140	-0.063	-3.793	0.274	0.055	1129.516
	[0.528 0.644]	-	-	[0.116 0.163]	-	-	[0.214 0.335]	-	-
female, with male preference	0.534	0.060	9.023	0.083	-0.006	90.671	0.383	-0.054	-921.065
	[0.44 0.628]	-	-	[0.064 0.101]	-	-	[0.277 0.489]	-	-

Notes: The columns denote post-secondary types - no post-secondary education, two-year college, and four-year college.

increasingly critical. More specifically, according to the estimation results, male students exhibit a greater inclination toward enrolling in STEM-related courses compared to their female counterparts. This preference gap predominantly contributes to disparities in course selection when contrasted with gender skill gaps. When imposing male preferences on females, not only are gender course-taking disparities eliminated, but they are even reversed,

with females surpassing males by a margin greater than their initial lag (Table 8). However, despite the boost in female STEM-related course enrollment and the removal of various choice-specific preference gaps, it fails to deliver a significantly more substantial improvement in mechanical skills compared to other skills. Consequently, this surge in female STEM-related course enrollment, coupled with the elimination of choice-specific preference gaps, does alleviate female under-representation in the applied-STEM field, i.e., it leads to a slight decrease in the under-representation, with a 27% decrease in the gender participation gap in choosing applied-STEM majors (Table 9) and a 13% decrease in the gender gap in selecting applied-STEM occupations (Table 10).

Under the influence of choice-specific preference gaps, the impact of gender skill gaps, especially in mechanical skills, on generating gender disparities in the applied-STEM field is relatively minor during the course selection phase but becomes increasingly significant in later stages. To elaborate, closing gaps in math or science skills at the beginning of the $11th$ grade reduces gender disparities in course enrollment across most STEM-related subjects by approximately 20%. However, closing all skill gaps at the beginning of the $11th$ grade exacerbates female under-representation in all STEM courses by more than 50%. This counterintuitive outcome is attributable to the adverse effect that mechanical skills have on course enrollment. Nevertheless, the effectiveness of closing gender skill gaps at the beginning of the $11th$ grade becomes more apparent in later stages. Closing all skill gaps at the beginning of the $11th$ grade substantially diminishes female under-representation in applied-STEM majors by 38% (Table 9) and in applied-STEM occupations by 15% (Table 10). Notably, closing the gender gap in mechanical skills at the beginning of the $11th$ grade alone reduces female under-representation in applied-STEM majors by 17% (Table 9) and in applied-STEM occupations by 7% (Table 10). In addition, closing the gender gap in math skills alone at the beginning of the $11th$ grade reduces female under-representation in applied-STEM majors by 11% (Table 9) and in applied-STEM occupations by 3% (Table 10).

In addition to simulations done from when the individuals start the $11th$ grade, I also implement simulations from when the individuals exit high school. The results reveal that closing the gender skill gaps upon exiting high school reduces female under-representation by 67% in applied-STEM majors (Table 11) and by 31% in applied-STEM occupations (Table 12). Closing the gender gap in mechanical skills upon exiting high school alone reduces female under-representation in applied-STEM majors by 33% (Table 11) and in applied-STEM

occupations by 17% (Table 12).

In sum, the results from Exercises (1) underscore the importance of skill-focused policies aimed at narrowing skill gaps, particularly in mechanical skills, rather than in other STEM-related skills.

4.2 The Trade-off between College Enrollment and Participation in Applied-STEM Fields

A noteworthy observation arising from counterfactual simulations is the significant trade-off between the college completion rate and the rate of holding four-year college applied-STEM majors when using the closure of the mechanical skill gap as a measure to enhance female participation in applied-STEM fields. These countervailing forces align with the contrasting effects of mechanical skills on college enrollment and the selection of applied-STEM majors, as previously revealed in the estimation results presented in Section 3.

The impact of closing gender gaps across the entire spectrum of skills is twofold. It leads to an 18% decrease in the female four-year college completion rate (from 0.334 to 0.274, as shown in Table 13). Simultaneously, it results in a 75% increase in the rate of female holding of applied-STEM majors (from 0.028 to 0.049, as displayed in Table 9) and an 18% rise in the rate of female holding of applied-STEM occupations (from 0.057 to 0.067, as indicated in Table 10).

To gain a clearer understanding of the role of mechanical skills, let's delve into the counterfactual results when closing the gender gap in mechanical skills. This shift in the skill distribution leads to a 32% reduction in the female four-year college completion rate (reducing it from 0.334 to 0.228, as depicted in Table 13). However, it also results in a 36% increase in the rate of females holding applied-STEM majors (raising it from 0.028 to 0.038, as presented in Table 9), and a 9% increase in the rate of females holding applied-STEM occupations (increasing it from 0.057 to 0.062, as outlined in Table 10).

4.3 Effects of Non-pecuniary and Pecuniary Benefits

In Tables 14-16, Rows 1-2 show the data moments by gender. Rows 3-4 show the simulation moments by gender, which are also the baseline for comparative analysis. Row 5 presents the simulation results of female individuals if their preference for workplace gender composition is removed. In Tables 17-19, Rows 1-2 show the data moments by gender. Rows 3-4 show the simulation moments by gender, which are also the baseline for comparative analysis. Rows 5 - 8 present the results of four counterfactuals, respectively: the simulation results of female individuals (1) if females have the same wage offer as men do in the applied-STEM sector (Row 5), (2) if females have the same wage offer as men do in the two STEM sectors, pure and applied (Row 6), (3) if the wage offer of the applied-STEM sector for women increases to the point where the gender composition in this sector is balanced (Row 7), and (4) if the wage offer of the applied-STEM sector for women increases to the point where the female representation is the same as if there were no female preference for workplace gender composition (Row 8). Next to the column of each choice, Tables 14-19 also display the $male - female$ gender gaps Δ_{M-F} in Columns 2,5, and 8 and their corresponding percentage change Δ in Columns 3,6, and 9.

Table 14: Counterfactual of # of Courses Taken in Grades 11 & 12 (Starting from Grade 10)

	math	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	1.159	-	-	2.259	-	-	0.997	-	-
female (data)	0.800	0.359	-	1.966	0.293	-	0.920	0.077	-
male (baseline)	1.240	-	-	2.493	-	-	1.123	-	-
female (baseline)	0.604	0.636	-	1.794	0.700	-	0.753	0.370	-
female, with no GC preference	0.604	0.635	0.040	1.795	0.698	0.169	0.753	0.370	0.149
	[0.479 0.73]	-	-	[1.585 2.005]	-	-	[0.58 0.927]	-	-

Notes: The columns denote the types of high-school STEM coursework - math, pure-STEM, and applied-STEM.

The results obtained from Exercises (2) and (3) unveil the impact of gender composition preferences and varying wage offers on the emergence of gender disparities in educational and career choices. Building upon these two exercises, I also quantify the monetary compensation required to counteract the degree of female under-representation in applied-STEM fields resulting from women's reluctance to enter male-dominated occupations. To delve into specifics, even without a significant improvement in females' STEM-related course enrollment (Table 14), eliminating the estimated preference for female-dominated workplaces reduces female under-representation by 29% in applied-STEM majors (Table 15) and a sub-

Table 15: Counterfactual of Four-year College Major Choice (Starting from Grade 10, Absolute Measures)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.187	-	-	0.044	-	-	0.083	-	-
female (data)	0.196	-0.009	-	0.112	-0.068	-	0.028	0.054	-
male (baseline)	0.194	-	-	0.051	-	-	0.084	-	-
female (baseline)	0.193	0.000	-	0.113	-0.062	-	0.028	0.056	-
female, with no GC preference	0.193	0.000	41.374	0.096	-0.045	27.031	0.044	0.040	29.219
	[0.165 0.222]	-	-	[0.081 0.112]	-	-	[0.034 0.054]	-	-

Notes: The columns denote the types of four-year college majors - non-STEM, pure-STEM, and applied-STEM.

Table 16: Counterfactual of Occupation Choice (Starting from Grade 10)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.696	-	-	0.095	-	-	0.126	-	-
female (data)	0.602	0.094	-	0.095	-0.000	-	0.067	0.059	
male (baseline)	0.716	-	-	0.075	-	-	0.126	-	-
female (baseline)	0.622	0.094	-	0.119	-0.045	-	0.057	0.069	-
female, with no GC preference	0.481	0.235	-149.742	0.073	0.001	102.946	0.116	0.010	85.274
	[0.445 0.517]	-	-	[0.059 0.087]	-	-	[0.084 0.147]	-	-

Notes: The columns denote occupation types - non-STEM, pure-STEM, and applied-STEM.

Table 17: Counterfactual of # of Courses Taken in Grades 11 & 12 (Starting from Grade 10)

	math	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	1.159	-	-	2.259	-	-	0.997	-	-
female (data)	0.800	0.359	-	1.966	0.293	-	0.920	0.077	-
male (baseline)	1.240	-	-	2.493	-	-	1.123	-	-
female (baseline)	0.604	0.636	-	1.794	0.700	-	0.753	0.370	-
female, with male wage offer - applied-STEM	0.604	0.636	0.011	1.794	0.700	0.003	0.753	0.370	0.057
	[0.479 0.73]	-	-	[1.584 2.004]	-	-	[0.581 0.925]	-	-
female, with male wage offer - pure- & applied-STEM	0.605	0.635	0.060	1.796	0.698	0.274	0.754	0.370	0.198
	[0.479 0.73]	-	-	[1.586 2.006]	-	-	[0.581 0.926]	-	-
female, with applied-STEM wage offer increased by 89.57%	0.605	0.635	0.061	1.795	0.699	0.105	0.754	0.370	0.179
	[0.479 0.73]	-	-	[1.584 2.005]	-	-	[0.58 0.927]	-	-
female, with applied-STEM wage offer increased by 80.03%	0.605	0.635	0.061	1.794	0.699	0.082	0.753	0.370	0.169
	[0.479 0.73]	-	-	[1.584 2.005]	-	-	[0.58 0.926]	-	-

Notes: The columns denote the types of high-school STEM coursework - math, pure-STEM, and applied-STEM.

Table 18: Counterfactual of Four-year College Major Choice (Starting from Grade 10, Absolute Measures)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.187	-	-	0.044	-	-	0.083	-	-
female (data)	0.196	-0.009	-	0.112	-0.068	-	0.028	0.054	-
male (baseline)	0.194	-	-	0.051	-	-	0.084	-	-
female (baseline)	0.193	0.000	-	0.113	-0.062	-	0.028	0.056	-
female, with male wage offer - applied-STEM	0.192	0.002	-346.415	0.113	-0.061	1.008	0.030	0.054	4.473
	[0.164 0.22]	-	-	[0.095 0.13]	-	-	[0.023 0.038]	-	-
female, with male wage offer - pure- & applied-STEM	0.187	0.006	-1510.568	0.125	-0.074	-19.173	0.029	0.055	2.550
	[0.16 0.215]	-	-	[0.106 0.144]	-	-	[0.022 0.037]	-	-
female, with applied-STEM wage offer increased by 89.57%	0.188	0.006	-1305.053	0.107	-0.056	9.377	0.041	0.043	22.784
	[0.161 0.215]	-	-	[0.091 0.124]	-	-	[0.031 0.05]	-	-
female, with applied-STEM wage offer increased by 80.03%	0.189	0.005	-1112.814	0.108	-0.057	7.673	0.039	0.045	19.596
	[0.161 0.216]	-	-	[0.092 0.125]	-	-	[0.029 0.048]	-	-

Notes: The columns denote the types of four-year college majors - non-STEM, pure-STEM, and applied-STEM.

Table 19: Counterfactual of Occupation Choice (Starting from Grade 10)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.696	-	-	0.095	-	-	0.126	-	-
female (data)	0.602	0.094	-	0.095	-0.000	-	0.067	0.059	-
male (baseline)	0.716	-	-	0.075	-	-	0.126	-	- %
female (baseline)	0.622	0.094	-	0.119	-0.045	-	0.057	0.069	- %
female, with male wage offer - applied-STEM	0.615	0.101	-7.180	0.118	-0.043	3.307	0.068	0.058	15.568
	[0.582 0.648]	-	-	[0.098 0.138]	-	-	[0.048 0.087]	-	-
female, with male wage offer - pure- & applied-STEM	0.584	0.132	-40.475	0.163	-0.088	-96.898	0.064	0.062	10.406
	[0.549 0.618]	-	-	[0.137 0.189]	-	-	[0.046 0.083]	-	-
female, with applied-STEM wage offer increased by 89.57%	0.577	0.139	-47.343	0.110	-0.035	21.585	0.126	-0.000	100.032
	[0.541 0.613]	-	-	[0.091 0.129]	-	-	[0.092 0.16]	-	-
female, with applied-STEM wage offer increased by 80.03%	0.583	0.133	-40.943	0.111	-0.037	18.018	0.116	0.010	85.288
	[0.548 0.619]	-	-	[0.092 0.13]	-	-	[0.084 0.147]	-	-

Notes: The columns denote occupation types - non-STEM, pure-STEM, and applied-STEM.

stantial 85% in applied-STEM occupations (Table 16).

In terms of monetary incentives, the tendency of women to avoid male-dominated occupations can be offset by an 80% increase in the wage offer for females in the applied-STEM sector (Table 19). Similar to the effect of eliminating the female preference for workplace gender composition, solely equalizing wage offers by gender in STEM sectors, which involves providing females working in the applied-STEM sector with a 19% wage bonus and those working in the pure-STEM sector with a 40% wage bonus, barely changes the landscape of STEM-related course enrollment (Table 17). In contrast, such a wage change reduces female under-representation by 3% in applied-STEM majors (Table 18) and by 10% in applied-STEM occupations (Table 19).

The findings derived from Exercises (2) and (3) underscore the effectiveness of wage compensation policies in diminishing female under-representation in the applied-STEM field, especially in the applied-STEM occupations. They also speak to the necessity of addressing women's aversion to male-dominated occupations by addressing the underlying institutional factors.

4.4 Effects of Course Requirement Policies

In Tables 20-22, Rows 1-2 show the data moments by gender. Rows 3-4 show the simulation moments by gender, which are also the baseline for comparative analysis. Rows 5 - 10 present the results of three counterfactuals respectively: the simulation results of female individuals (1) if to have at least one math course is required (Row 5-6), (2) if to have at least one STEM course is required, pure or applied (Row 7-8), and (3) if to have at least one math course AND one STEM course is required, pure or applied (Row 9-10). Next to the column of each choice, Tables 20-22 also display the $male - female$ gender gaps Δ_{M-F} in Columns 2,5, and 8 and their corresponding percentage change Δ in Columns 3, 6, and 9.

Exercise (4) highlights the limited effectiveness of certain gender-neutral policies, specifically those mandating an increase in the number of STEM-related courses to be taken. While this policy succeeds in reducing the gender gap in STEM-related course enrollment during high school by around 30-50% (Table 20), it proves ineffective in diminishing, and in fact, slightly exacerbates female under-representation in applied-STEM fields at later stages of

Table 20: # of Courses Taken (Grades 11 & 12) - Closing Gender Course-taking Gaps through Course-taking Requirements

	math	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	1.159	-	-	2.259	-	-	0.997	-	-
female (data)	0.800	0.359	-	1.966	0.293	-	0.920	0.077	-
male (baseline)	1.240	-	-	2.493	-	-	1.123	-	-
female (baseline)	0.604	0.636	-	1.794	0.700	-	0.753	0.370	-
male (require $\geq$ 1 math crs.)	2.677	-	-	2.377	-	-	1.128	-	-
	[2.535 2.819]	-	-	[1.968 2.785]	-	-	[0.837 1.42]	-	-
female (require $\geq$ 1 math crs.)	2.381	0.296	53.445	1.659	0.718	-2.583	0.759	0.370	0.211
	[2.274 2.489]	-	-	[1.35 1.968]	-	-	[0.477 1.041]	-	-
male (require $\geq$ 1 p-S or a-S crs.)	1.235	-	-	3.077	-	-	1.409	-	-
	[1.003 1.467]	-	-	[2.804 3.349]	-	-	[1.118 1.699]	-	-
female (require $\geq$ 1 p-S or a-S crs.)	0.590	0.646	-1.582	2.585	0.492	29.678	1.160	0.248	32.923
	[0.46 0.719]	-	-	[2.361 2.808]	-	-	[0.947 1.373]	-	-
male (require $\geq$ 1 math and $\geq$ 1 p-S or a-S crs.)	2.673	-	-	2.962	-	-	1.442	-	-
	[2.527 2.818]	-	-	[2.577 3.347]	-	-	[1.129 1.755]	-	-
female (require $\geq$ 1 math and $\geq$ 1 p-S or a-S crs.)	2.373	0.300	52.862	2.459	0.503	28.105	1.194	0.248	33.087
	[2.266 2.479]	-	-	[2.12 2.798]	-	-	[0.878 1.51]	-	-

Notes: The columns denote the types of high-school STEM coursework - math, pure-STEM, and applied-STEM.

Table 21: Four-year College Major Choice - Closing Gender College Major Gaps through Coursetaking Requirements (Absolute Measures)

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.187	-	-	0.044	-	-	0.083	-	-
female (data)	0.196	-0.009	-	0.112	-0.068	-	0.028	0.054	-
male (baseline)	0.194	-	-	0.051	-	-	0.084	-	-
female (baseline)	0.193	0.000	-	0.113	-0.062	-	0.028	0.056	-
male (require $\geq$ 1 math crs.)	0.192	-	-	0.052	-	-	0.094	-	-
	[0.119 0.265]	-	-	[0.031 0.073]	-	-	[0.046 0.141]	-	-
female (require $\geq$ 1 math crs.)	0.198	-0.006	1594.070	0.117	-0.064	-3.712	0.030	0.064	-13.023
	[0.142 0.253]	-	-	[0.084 0.149]	-	-	[0.016 0.044]	-	-
male (require $\geq$ 1 p-S or a-S crs.)	0.204	-	-	0.053	-	-	0.093	-	-
	[0.137 0.272]	-	-	[0.038 0.068]	-	-	[0.058 0.129]	-	-
female (require $\geq$ 1 p-S or a-S crs.)	0.217	-0.013	3284.804	0.118	-0.065	-5.274	0.030	0.063	-12.929
	[0.178 0.256]	-	-	[0.099 0.137]	-	-	[0.022 0.037]	-	-
male (require $\geq$ 1 math and $\geq$ 1 p-S or a-S crs.)	0.206	-	-	0.056	-	-	0.100	-	-
	[0.12 0.293]	-	-	[0.034 0.079]	-	-	[0.047 0.154]	-	-
female (require $\geq$ 1 math and $\geq$ 1 p-S or a-S crs.)	0.216	-0.010	2623.936	0.125	-0.069	-10.851	0.035	0.066	-16.791
	[0.158 0.274]	-	-	[0.091 0.158]	-	-	[0.02 0.049]	-	-

Notes: The columns denote the types of four-year college majors - non-STEM, pure-STEM, and applied-STEM.

Table 22: Occupation Choice - Closing Gender Occupation Gaps through Coursetaking Requirements

	non-STEM	Δ_{M-F}	%	**pure-STEM**	Δ_{M-F}	%	**applied-STEM**	Δ_{M-F}	%
male (data)	0.696	-	-	0.095	-	-	0.126	-	-
female (data)	0.602	0.094	-	0.095	-0.000	-	0.067	0.059	-
male (baseline)	0.716	-	-	0.075	-	-	0.126	-	-
female (baseline)	0.622	0.094	-	0.119	-0.045	-	0.057	0.069	-
male (require $\geq$ 1 math crs.)	0.711	-	-	0.076	-	-	0.130	-	-
	[0.674 0.747]	-	-	[0.058 0.094]	-	-	[0.097 0.163]	-	-
female (require $\geq$ 1 math crs.)	0.623	0.087	7.173	0.120	-0.044	0.831	0.057	0.073	-6.256
	[0.589 0.658]	-	-	[0.098 0.143]	-	-	[0.039 0.076]	-	-
male (require $\geq$ 1 p-S or a-S crs.)	0.711	-	-	0.076	-	-	0.131	-	-
	[0.678 0.743]	-	-	[0.058 0.093]	-	-	[0.103 0.159]	-	-
female (require $\geq$ 1 p-S or a-S crs.)	0.624	0.086	8.385	0.121	-0.045	-0.332	0.056	0.075	-8.471
	[0.591 0.658]	-	-	[0.1 0.141]	-	-	[0.04 0.073]	-	-
male (require $\geq$ 1 math and $\geq$ 1 p-S or a-S crs.)	0.707	-	-	0.076	-	-	0.134	-	-
	[0.668 0.746]	-	-	[0.058 0.095]	-	-	[0.098 0.171]	-	-
female (require $\geq$ 1 math and $\geq$ 1 p-S or a-S crs.)	0.620	0.087	7.209	0.124	-0.047	-5.870	0.059	0.075	-9.430
	[0.585 0.654]	-	-	[0.101 0.146]	-	-	[0.041 0.078]	-	-

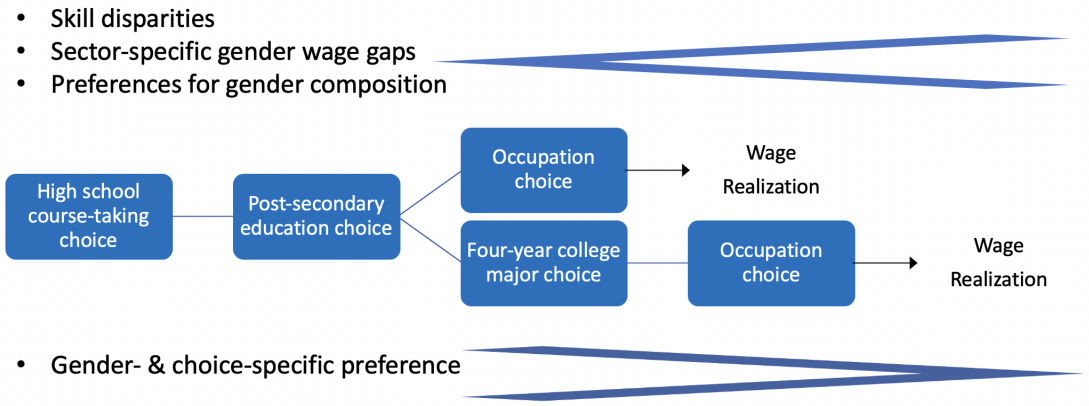
Notes: The columns denote occupation types - non-STEM, pure-STEM, and applied-STEM.

education (Table 21) and career choices (Table 22). The inefficacy of this approach can be attributed to its gender-neutral nature, which results in a failure to substantially narrow the gender skill gaps, particularly in mechanical skills. These findings underscore the necessity of implementing policies that are not only female-targeted but also specifically focused on improving mechanical skills. Such policies are essential for addressing the issue of inadequate female representation in the applied-STEM field.

4.5 Summary: Lifecycle-Varying Effects of Different Factors on Gender Choice Disparities

In summarizing the results from the counterfactual analysis above, it becomes evident that concerning gender disparities throughout the sequence of choices, as depicted in Figure 2 — including taking high-school courses, selecting post-secondary education, choosing a college major in the case of pursuing a four-year degree, and deciding on an occupation, the influence of gender preference gaps gradually diminishes. In contrast, the significance of gender skill gaps, wage disparities, and females' aversion to male-dominated occupations becomes increasingly pivotal in shaping these gender-based disparities in choices.

Figure 2: Importance Assessment Result



5 Conclusion

In conclusion, this paper has undertaken a thorough examination of gender disparities in STEM participation, delving into the intricate interplay among skills, preferences, and policy interventions. To recapitulate the key findings, three significant takeaways have emerged from my analysis.

First and foremost, subject-specific preferences have emerged as the dominant influence in shaping high-school coursework choices. Recent research, exemplified by Todd (2022) and Ceci (2014), highlights the narrowing of gender-based disparities in coursework choices, coinciding with the convergence of math and science skills, and the underlying subject-specific preference gaps.

Secondly, the enduring gender gap in mechanical skills, relative to the closing gender gap in math skills, elucidates the persistently low proportion of women opting for applied-STEM majors within the broader population. This trend is even more pronounced in the college demographic, despite the increasing number of women pursuing higher education since the 1980s.

Lastly, women's aversion to male-dominated occupations plays a pivotal role in the sustained underrepresentation of females in applied-STEM fields. While reducing the gender skill gap in mechanical skills can alleviate a portion of the underrepresentation in applied-STEM majors, it may not have as significant an impact on closing the gender gap in applied-STEM occupations.

By shedding light on the evolving dynamics of these factors, this research provides valuable insights for policymakers, educators, and stakeholders striving to foster a more inclusive and diverse STEM landscape. The fact that the initial gender skill gaps at the end of the $10th$ grade coupled with women's distaste for working in male-dominated occupations lead to women's reluctance to choose the applied-STEM fields, calls for pre-Grade-10 intervention policies that are not only female-targeted but also specifically focused on improving mechanical skills such as female-targeted summer camps and extracurricular activities. It is also a policy imperative to address females' concerns about working in a male-dominated environment, e.g., the likelihood of encountering discrimination and harassment (Folke and Rickne 2022), access to gender-specific accommodations and facilities such as flexible work arrangements and nursing facilities, the propensity to be assigned non-promotable tasks (Babcock et al. 2017), and the patterns of information exchange within and across gender lines (Hampole et al. 2021; Gallen and Wasserman 2021; Gallen and Wasserman 2023).

This analysis also reveals contrasting effects of mechanical skills on college enrollment and the selection of applied-STEM majors. According to the estimation results, having a higher level of mechanical skills alone makes the noncollege option more attractive upon high-school graduation. However, when combined with a higher level of mathematical skills, it significantly increases the likelihood of choosing a four-year college major in applied-STEM fields. This finding aligns with the simulation result, suggesting that raising girls' mechanical skill levels to match those of boys would lead to a decrease in the overall female four-year college completion rate. Simultaneously, it would increase the percentage of women pursuing four-year college applied-STEM majors and opting for careers in applied-STEM occupations. It's crucial to note that this decrease in the female four-year college completion rate should not be seen as a setback in the progress of women's higher education attainment. Instead, it reflects positive strides toward achieving gender equity by improving female mechanical skills and reducing occupational segregation.

References

Altonji, Joseph G and Lewis M Segal, "Small-sample bias in GMM estimation of covariance structures," *Journal of Business & Economic Statistics*, 1996, *14* (3), 353–366.

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

_ , V Joseph Hotz, and Songman Kang, "Modeling college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 2012, *166* (1), 3–16.

Babcock, Linda, Maria P Recalde, Lise Vesterlund, and Laurie Weingart, "Gender differences in accepting and receiving requests for tasks with low promotability," *American Economic Review*, 2017, *107* (3), 714–747.

Bartolj, Tjaša and Sašo Polanec, "College major choice and ability: Why is general ability not enough?," *Economics of Education Review*, 2012, *31* (6), 996–1016.

Beede, David N, Tiffany A Julian, David Langdon, George McKittrick, Beethika Khan, and Mark E Doms, "Women in STEM: A gender gap to innovation," *Economics and Statistics Administration Issue Brief*, 2011, (04-11).

Bostwick, Valerie K. and Bruce A. Weinberg, "Nevertheless She Persisted? Gender Peer Effects in Doctoral STEM Programs," *Journal of Labor Economics*, apr 1 2022, *40* (2), 397–436.

Card, David and A Abigail Payne, "High school choices and the gender gap in STEM," *Economic Inquiry*, 2021, *59* (1), 9–28.

Carrell, Scott E., Marianne E. Page, and James E. West, "Sex and Science: How Professor Gender Perpetuates the Gender Gap*," *Quarterly Journal of Economics*, 8 2010, *125* (3), 1101–1144.

Ceci, Stephen J, Donna K Ginther, Shulamit Kahn, and Wendy M Williams, "Women in academic science: A changing landscape," *Psychological science in the public interest*, 2014, *15* (3), 75–141.

Cheryan, Sapna, Sianna A Ziegler, Amanda K Montoya, and Lily Jiang, "Why are some STEM fields more gender balanced than others?," *Psychological bulletin*, 2017, *143* (1), 1.

Chuan, Amanda and Weilong Zhang, "Non-College Occupations, Workplace Routinization, and the Gender Gap in College Enrollment," 2021.

Cortes, Patricia and Jessica Pan, "Occupation and gender," *The Oxford handbook of women and the economy*, 2018, pp. 425–452.

Delfino, Alexia, "Breaking Gender Barriers: Experimental Evidence on Men in Pink-Collar Jobs," 2021.

Folke, Olle and Johanna Rickne, "Sexual harassment and gender inequality in the labor market," *The Quarterly Journal of Economics*, 2022, *137* (4), 2163–2212.

Gallen, Yana and Melanie Wasserman, "Informed Choices," *Employment Research*, 2021, *28* (3), 5–7.

_ **and** _ , "Does information affect homophily?," *Journal of Public Economics*, 2023, *222*, 104876.

Hampole, Menaka, Francesca Truffa, and Ashley Wong, "Peer effects and the gender gap in corporate leadership: Evidence from MBA students," Technical Report, Working Paper 2021.

Harvill, Eleanor L, "High school math curriculum, students' course selection and education outcomes." PhD dissertation, University of Pennsylvania 2011.

Hill, Catherine, Christianne Corbett, and Andresse St Rose, *Why so few? Women in science, technology, engineering, and mathematics.*, ERIC, 2010.

Kahn, Shulamit and Donna Ginther, "Women and STEM," Technical Report, National Bureau of Economic Research 2017.

Keane, Michael P and Kenneth I Wolpin, "The career decisions of young men," *Journal of political Economy*, 1997, *105* (3), 473–522.

Kinsler, Josh and Ronni Pavan, "The specificity of general human capital: Evidence from college major choice," *Journal of Labor Economics*, 2015, *33* (4), 933–972.

Kuhn, Peter and Kailing Shen, "Gender discrimination in job ads: Evidence from china," *The Quarterly Journal of Economics*, 2013, *128* (1), 287–336.

Larson-Koester, Miriam, "Occupation Gender Segregation: Empirical Evidence from a Matching Model with Transfers," 2017.

Legewie, Joscha and Thomas A DiPrete, "School context and the gender gap in educational achievement," *American Sociological Review*, 2012, *77* (3), 463–485.

Morando, Greta, "Peer gender and STEM specialization," *Applied Economics Letters*, jul 23 2020, *28* (12), 1041–1045.

Mouganie, Pierre and Yaojing Wang, "High-Performing Peers and Female STEM Choices in School," *Journal of Labor Economics*, 7 2020, *38* (3), 805–841.

Rosser, Sue V and Eliesh O'Neil Lane, "Key barriers for academic institutions seeking to retain female scientists and engineers: Family-unfriendly policies. Low Numbers, stereotypes, and harassment," *Journal of Women and Minorities in Science and Engineering*, 2002, *8* (2).

Roy, Andrew Donald, "Some thoughts on the distribution of earnings," *Oxford economic papers*, 1951, *3* (2), 135–146.

Rust, John, "Optimal replacement of GMC bus engines: An empirical model of Harold Zurcher," *Econometrica: Journal of the Econometric Society*, 1987, pp. 999–1033.

Schuh, Rachel, "Miss-Allocation: The Value of Workplace Gender Composition and Occupational Segregation," 2022.

Shan, Xiaoyue, "Does minority status drive women out of male-dominated fields?," *Unpublished manuscript*, 2021.

Sloane, Carolyn M, Erik G Hurst, and Dan A Black, "College Majors, Occupations, and the Gender Wage Gap," *Journal of Economic Perspectives*, 2021, *35* (4), 223–48.

Spertus, Ellen, "Why are there so few female computer scientists?," 1991.

Stinebrickner, Ralph and Todd R Stinebrickner, "Academic performance and college dropout: Using longitudinal expectations data to estimate a learning model," *Journal of Labor Economics*, 2014, *32* (3), 601–644.

_ **and** _ , "A major in science? Initial beliefs and final outcomes for college major and dropout," *Review of Economic Studies*, 2014, *81* (1), 426–472.

Stinebrickner, Todd R and Ralph Stinebrickner, "Learning about academic ability and the college dropout decision," *Journal of Labor Economics*, 2012, *30* (4), 707–748.

Sullivan, Paul, "Empirical evidence on occupation and industry specific human capital," *Labour economics*, 2010, *17* (3), 567–580.

Todd, Petra and Shasha Wang, "How Early Skill Gaps Contribute to Gender Differences in STEM Major and Occupation Choices: A Random Forest Approach," 2022.

A Detailed Estimation Strategy

I use Simulated Method of Moments (SMM) for estimation. More specifically, I divide the dataset into four subsets according to the highest grade completed when they took the ASVAB test - those taking the test with the 9th grade completed (**D9**), those taking the test with the 10th grade completed (**D10**), those taking the test with the 11th grade completed (**D11**), and those taking the test with the 12th grade completed (**D12**). Thanks to the randomness of the data, subsamples according to different grade's completion can be comparable to each other, as shown in Table 1, where the summary statistics of the full sample, the **D10** sample, and the **D12** sample are presented.

I listed in Table 23 all the moments and their corresponding datasets used to estimate the parameters.

Table 23: Matching Moments in GMM

id	Datasets		Moments	#	Param.
	For data	For model			
1	**D9, D10, D11, D12**	**D12**	Occupation distributions (i.e., non-STEM, pure-STEM, applied-STEM, and staying at home) conditional on the four gender-race categories[10], on the five education status,[11] and on different skill levels.[12]	$4 \times 4 + 5 \times 4 + 3 \times 4 \times 4 = 84$	β, α
2	**D9, D10, D11, D12**	**D12**	The four-year college major distribution (i.e., non-STEM, pure-STEM, applied-STEM) conditional on the four gender-race categories and different skill levels	$4 \times 3 + 3 \times 3 \times 4 + 4 \times 3 \times 6 = 120$	η

[10] I.e., male nonblack, male black, female nonblack, and female black.

[11] I.e., noncollege, two-year college, four-year college with non-STEM degrees, four-year college with pure-STEM degrees, and four-year college with applied-STEM degrees.

[12] For moments conditional on a single skill, I categorize each skill into three levels - low, middle, and high according to percentile ranks. For moments conditional on two skill sets, I categorize each skill into two levels $\{low, high\}$ and take the product of any two skill sets $\{(low, low), (low, high), (high, low), (high, high)\}$.

3	**D9, D10, D11, D12**	**D12**	Post-secondary education distribution (i.e., going to a two-year college, going to a four-year college, and not going to either type of college) conditional on the four gender-race categories, on the product space of parents' education level and gender,[13] and on different skill levels.	$4 \times 3 + 6 \times 3 + 3 \times 3 \times 4 + 4 \times 3 \times 6 = 138$	λ
4	**D9, D10, D11, D12**	**D12**	Average log wage conditional on occupation type and gender, on occupation type and race, on occupation type and education status, and on different skill levels	$6 + 6 + 15 + 3 \times 4 = 39$	β
5	**D9, D10, D11, D12**	**D12**	The gender-occupation-specific variance of log wage	6	β_{σ}^{ns}, β_{σ}^{ps}, β_{σ}^{as}
6	**D11**	**D10**	The (expected) skill profile's mean vectors for four gender-race-specific groups (i.e., male/female $\times$ nonblack/black) and for six gender-parental-education groups, and the gender-specific variance-covariance matrices	$4 \times 10 + 10 \times 2 = 60$	γ
7	**D12**	**D11**	A/A	A/A	A/A
8	**D10**	**D10**	The (expected) number of the three types of courses taken (i.e., math, pure-STEM, applied-STEM) conditional on different skill levels, on genders, on races, and on parents' education levels	$3 \times 3 \times 4 + 3 \times 2 \times 2 + 3 \times 3 = 57$	κ
9	**D11**	**D11**	A/A	A/A	A/A
Total #				$N_{mom} = 621$, $N_{para} = 149$	

Note: A/A means "as above".

B Model Fit

This section presents the model fit plots in Figures 3-8.

[13] I categorize parents' education into 3 levels - low, i.e., both parents completing no more than 12 years of education, middle, i.e., at least one parent completing $(12, 16)$ years of education and neither completing more, and high, i.e., at least one parent completing above 12 years of education.

Figure 3: Moments Matching - Post-secondary Education Distribution

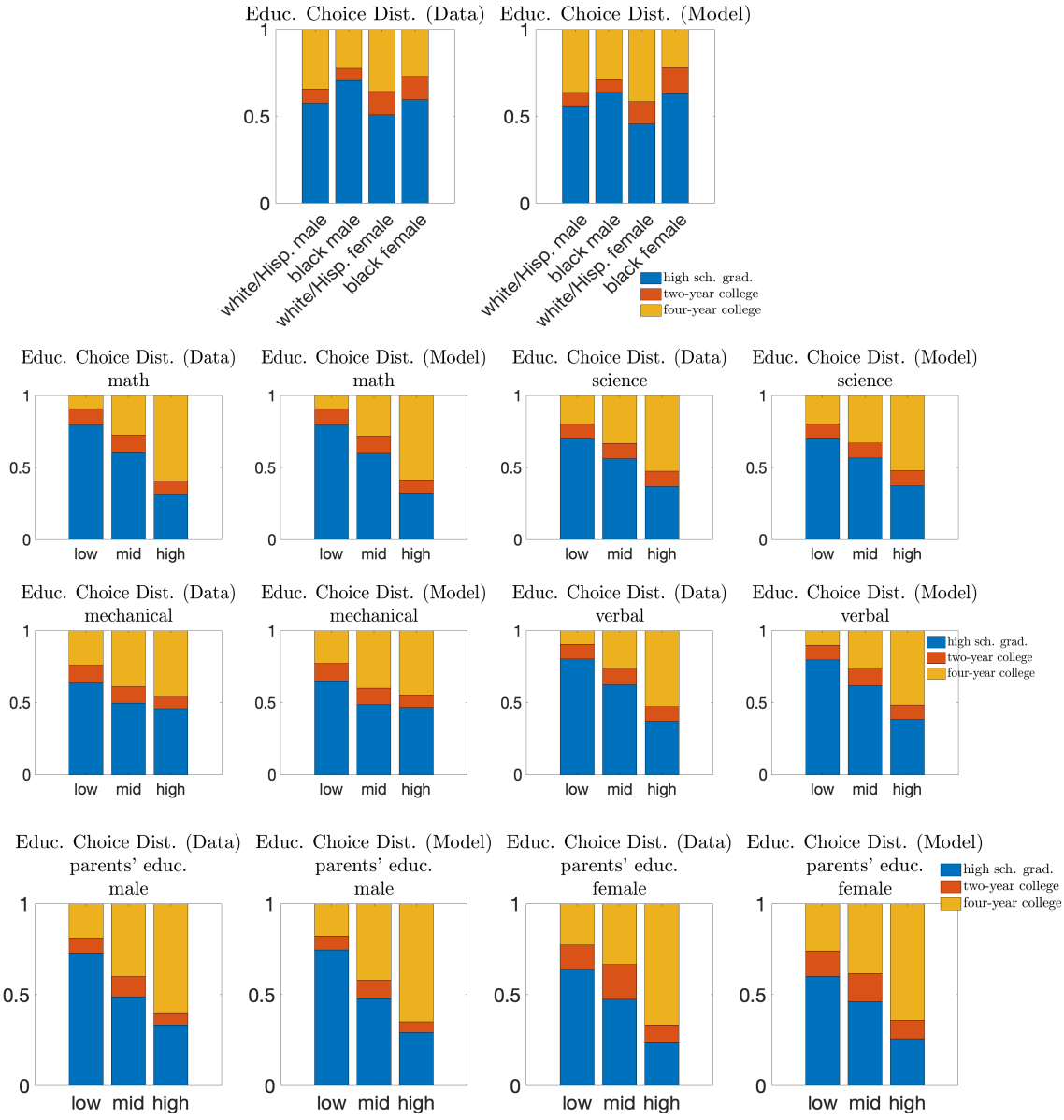


Figure 4: Moments Matching - Four-year College Major Distribution

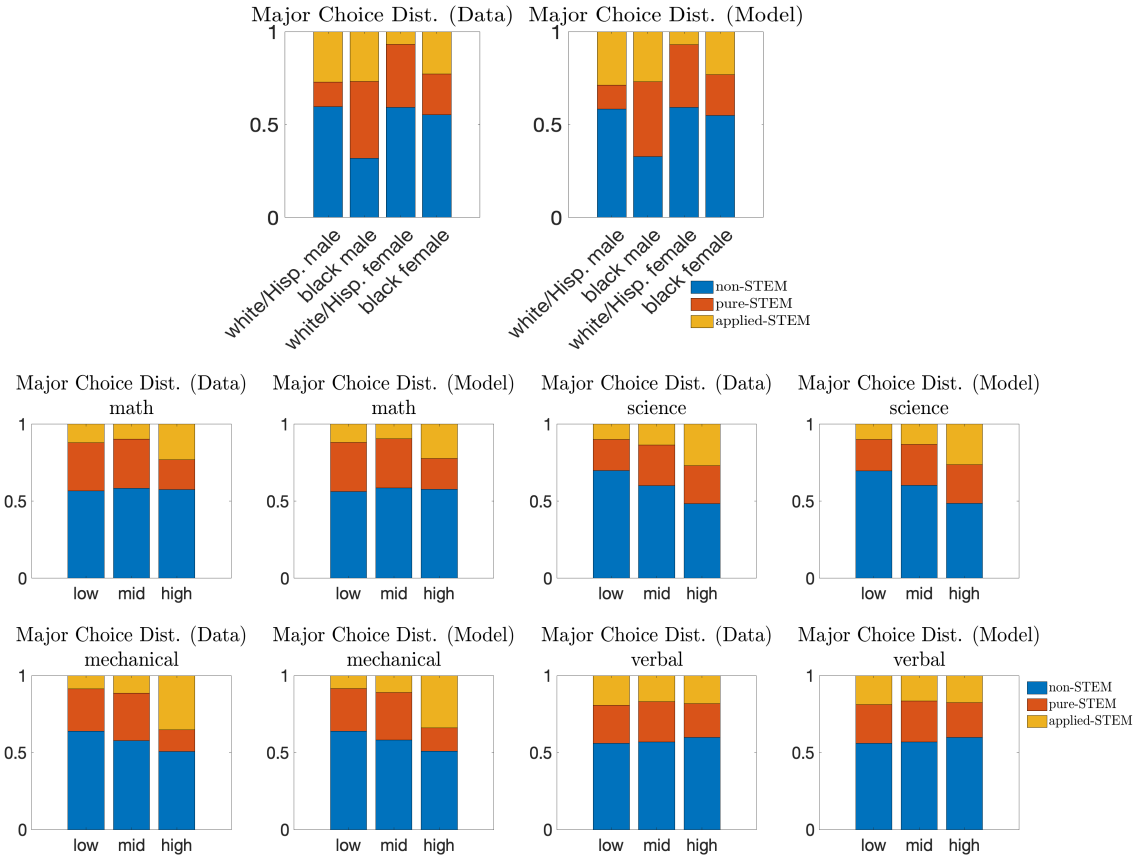


Figure 5: Moments Matching - Occupation Choice Distribution

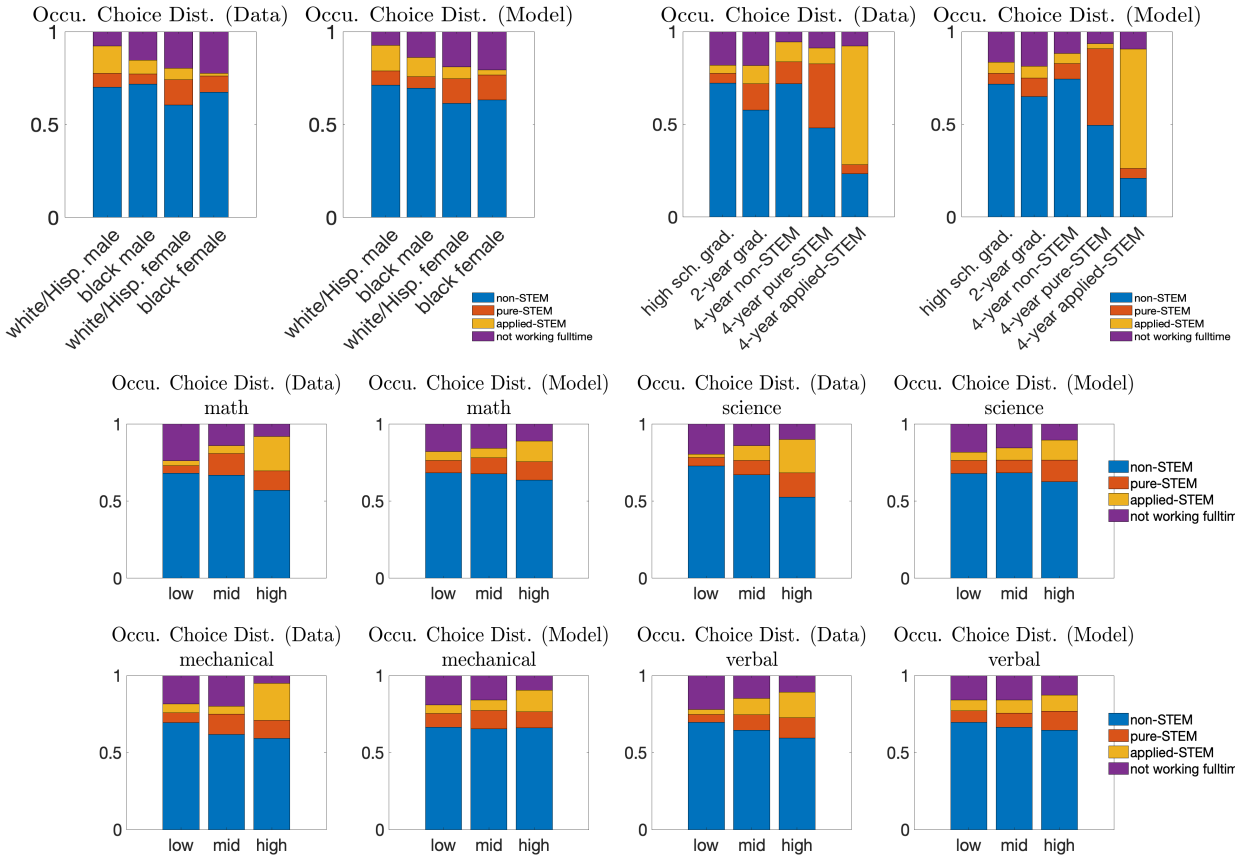


Figure 6: Moments Matching - High-school Coursework Moments

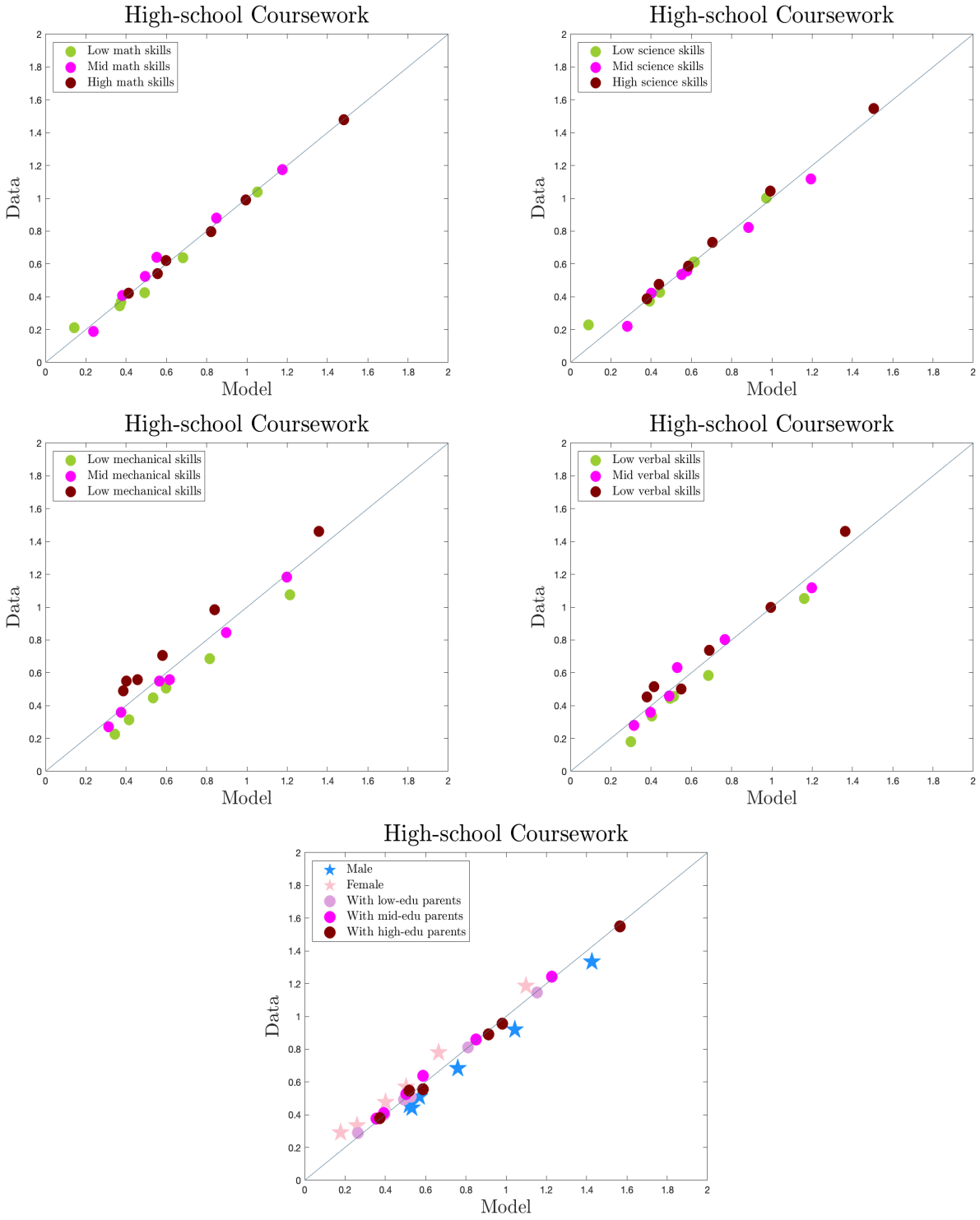


Figure 7: Moments Matching - Skill Score Moments

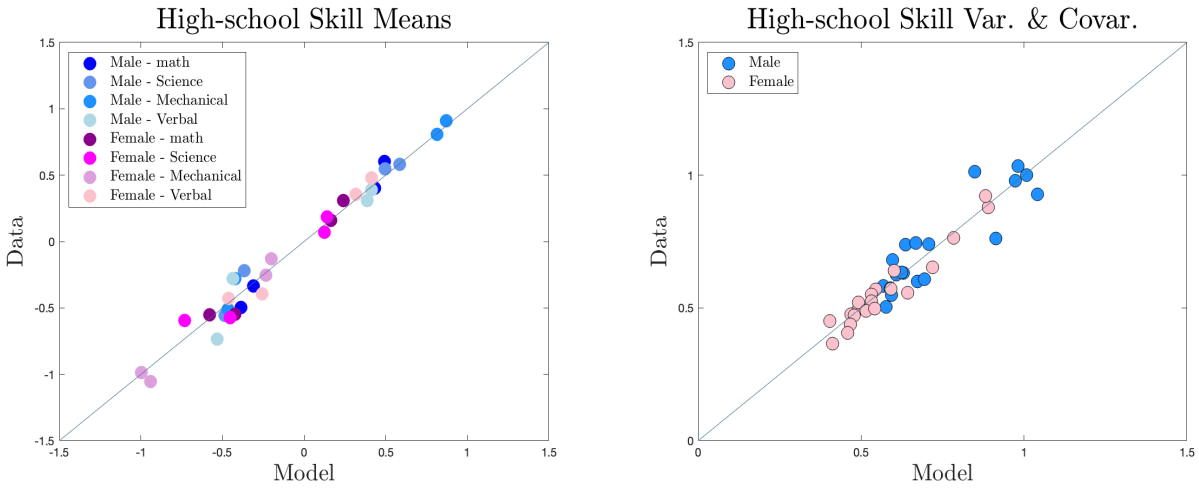
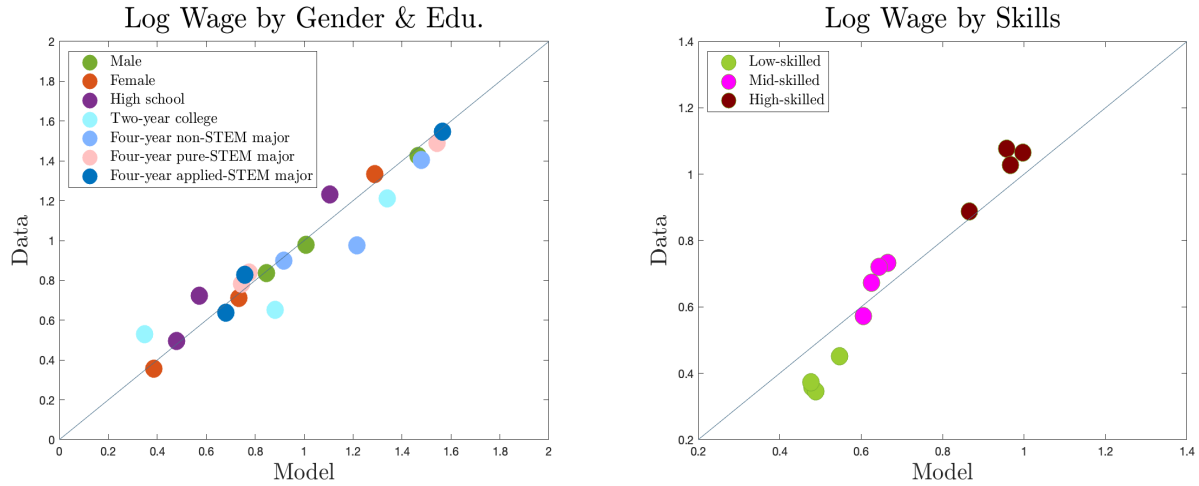


Figure 8: Moments Matching - Wage Moments



C Model Timeline

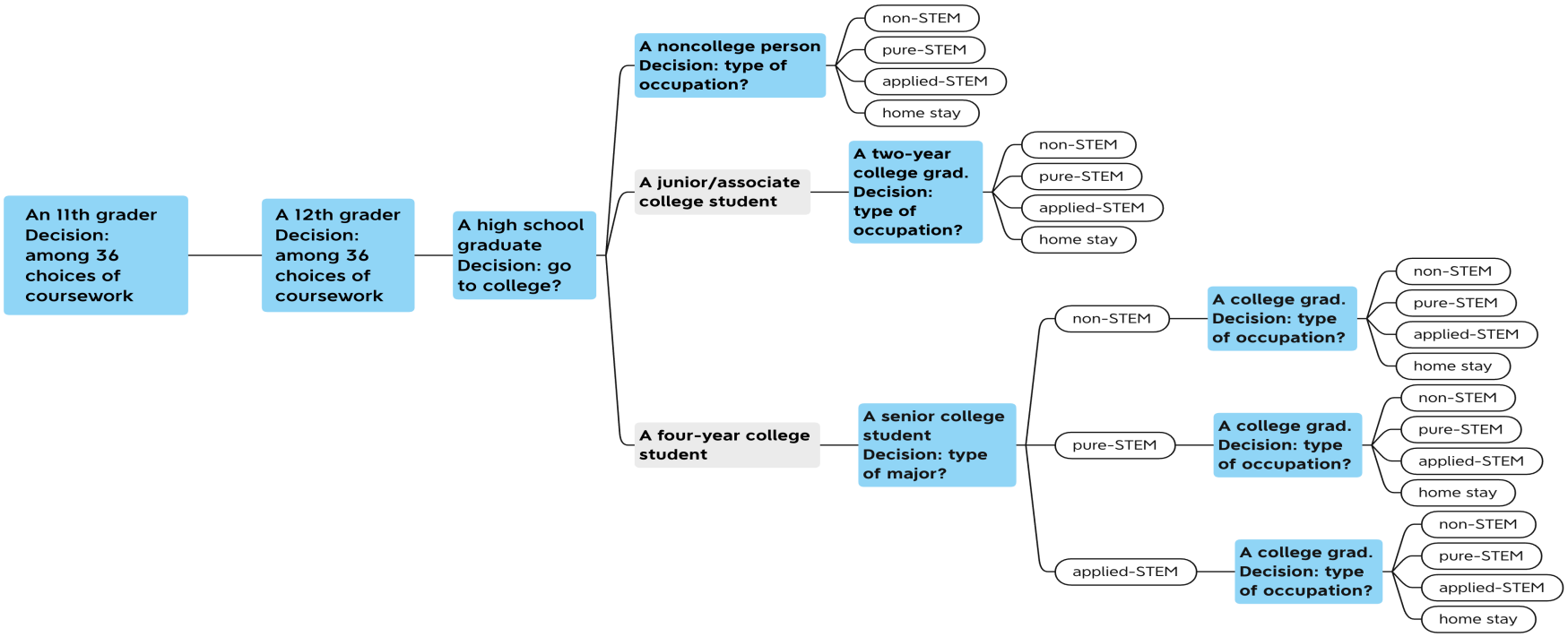


Figure 9: Dynamic Model Decision Flow