Is College Worth It For Me?

Beliefs, Access to Funding, and Inequality in Higher Education

Sergio Ernesto Barrera
University of Minnesota

October 2021

Abstract

In the US the bachelor's attainment rate of White high-socioeconomic status youth is much higher than the bachelor's attainment rate of Hispanic, Black, and low-socioeconomic status youth. This is true even among students with high academic scores. For high-scorers, how much of these gaps in bachelor's attainment can be explained by differences in subjective beliefs about own academic ability? Relatedly, are targeted policies that provide information and funding to low socioeconomic status high-scorers more efficient at narrowing overall bachelor's attainment gaps than universal policies like free college for all, or a tracking system in the US? To answer these questions, I estimate a dynamic discrete choice model with heterogeneous financial support and beliefs about college ability, learning through grade revelation and credit constraints. I find that for Black high-scorers beliefs play almost no role in explaining gaps. For Hispanic and low socioeconomic status youth, differences in beliefs explain 38-49% of the gap relative White high-socioeconomic status high-scorers. I then show that the targeted policy is more efficient at closing overall inequality than free college for all, or a tracking system. However, gaps will persist with differences in early human capital investment and non pecuniary utility by race, ethnicity and socioeconomic status.

1 Introduction

In the United States there are still large gaps in bachelor's attainment by race, ethnicity, and socioeconomic background. Even among students with high academic ability, youth from lower socioeconomic backgrounds are less likely to enroll in four-year institutions and selective colleges (Hoxby and Avery 2013). Previous research has shown that information frictions concerning academic ability generates underinvestment in college for all high academic ability youth (Arcidiacono, Aucejo, Maurel, Ransom 2016). However, Hoxby and Turner 2013 demonstrated that these information frictions might be more important for youth from demographic groups with lower levels of college completion. Specifically, they suggest these youth have less information about their college ability and expected returns than their more affluent peers with similar measures of academic ability.

This paper focuses on inequality in bachelor's attainment rates by demographic group. Specifically, I focus on the difference in bachelor's attainment between a reference group with high bachelor's attainment rates to three comparison groups with low bachelor's attainment rates. The reference group is White high socioeconomic status (SES) youth. The first two comparison groups are Black, and Hispanic youth regardless of SES. The final comparison group is low-SES youth without restricting by race and ethnicity. My first question focuses on youth with high measures of academic ability and good behavior, what will be referred to as "High-Scorers". I ask How do differences in the distribution of beliefs about college ability affect inequality in bachelor's attainment for high-scorers? For this first question I focus on high-scorers because underinvestment by demographic group for high-scorers is more likely to have higher economic costs[1]. Also, because initiatives to increase college diversity have traditionally sought enroll more high-scorers from traditionally underrepresented backgrounds, what many colleges call opening pipelines into higher education.

[1] In terms of foregone benefits and even potential economic growth (Hsieh, Hurst, Jones, and Klenow 2020)

My second question is a policy question that focuses on inequality between the reference group and comparison groups, independent of being high-scorers. I ask, are policies that target low-SES high-scorers more efficient at decreasing inequality than policies that are universally applied to all students? The targeted policy under consideration provides free college and information about college ability to low-SES high-scorers. It will be referred to as the college recruiting policy like those studied in Dynarski, Michelmore, Libassi, and Owen 2019 and Hoxby and Turner 2013. The first universally targeted policy is free college for all, where tuition is set to zero for everyone. The second universal policy is the implementation of a tracking system that provides information about college ability to all. The policies will be primarily concerned with how they narrow differences in bachelor's attainment rates between the reference group and comparison groups. The second concern will be an efficiency measure. In this paper efficiency will be measured in terms of college mismatch. Mismatch takes the form of over-investment of low-scoring youth with lower returns to college, and under-investment of high-scoring youth with high returns to college. It will be measured by the amount of youth who would make different education decisions if they had complete information about their own ability.

The preferred policy will depend on three distributions by demographic group: subjective beliefs about ability type, financial assistance, and the proportion of high-scorers. If beliefs are mostly correct for everyone, then tracking and the targeted college recruiting policy are less likely to be preferred. If information frictions are important and the fraction of high-scorers enrolling is low in the comparison groups, then tracking may increase inequality. This is because increases in enrollment of high-scorers could be offset by decreases in enrollment from low-scorers in the comparison group . In this case the targeted recruiting policy will be preferred since it will only increase enrollment among high-scorers with no effect on low-scorers. In this way inequality will be closed and mismatch reduced.

To answer these questions, I estimate a dynamic discrete choice model, with credit constraints, latent college ability type, and information frictions concerning college ability. In the model agents learn about college ability, post college earnings, and non-pecuniary utility through college GPA[2]. The model will have heterogeneity of subjective beliefs about ability and heterogeneity in financial assistance. The model will allow for the distribution of beliefs to be biased[3] and differ by demographic background. The level of financial assistance from colleges, government and family will also depend on demographic group. Differences by demographic group in early childhood human capital development will enter the model through the distribution of high-scorers by demographic group. This will be how early childhood human capital generates grades and earnings outcomes.

The main benefit of the model is that it allows the role of beliefs to be disentangled from parental wealth, education, human capital, financial assistance, and non-pecuniary utility since the model makes explicit assumptions concerning the relationship between these variables. More importantly the predicted exit behavior from learning from GPA will assist in estimating the prior distribution of beliefs, along with measures of subjective beliefs from the data set.

To estimate the model, I use the National Longitudinal Study of Youth 1997. The data set is representative of US youth born in the early 1980's and over samples Black and Hispanic youth. This makes it useful for studying racial and ethnic inequality. I estimate the structural model using a two-step procedure. In the first step I estimate predicted financial assistance by demographic characteristics as well as a finite mixture model for high-scorers and low-scorers. The main items of interest are college earnings by latent type, probability of college GPA conditioned by latent type, and probability of being high-scorer that depends on demographic characteristics and measurements of human capital. For the first step I use

[2]This type of learning was shown to be important in Stinebrickner and Stinebrickner 2012, for Berea College students.

[3]Biased with respect to what would be predicted from the data by an econometrician. Agents may not actually be biased with respect to local economic conditions for their relevant reference group.

data on demographic characteristics, earnings, financial assistance, educational attainment, college grades, ASVAB math and verbal scores for cognitive human capital, and measures of adverse behavior at young ages for non-cognitive human capital. The parameters from the first step will be used in the second step.

The second step is an internal estimation procedure where I use indirect inference to estimate the parameters of interest. The parameters I estimate in this step are the tuition sticker price per decision period, the distribution of subjective beliefs, and the distribution of non-pecuniary utility. The distribution of beliefs is estimated using reported beliefs about degree attainment from the NLSY97. I identify the distribution of beliefs by targeting the OLS coefficient of measured beliefs on enrollment, and the OLS coefficients of GPA levels in college continuation while controlling for access to financial resources. The relationship between GPA and beliefs comes directly from the learning mechanism in the model and is primarily identified by the change in exit behavior by increases in GPA in college.

In answer to the first question, I find that in comparison to the reference group, differences in beliefs explains 33% of the Black-White high-scorer gap. However this estimate has a large standard error which prevents us from ruling out an effect of zero. However, differences in beliefs explains 38 percent of the low-SES high-scorer gap, and 49 percent of the Hispanic high-scorer gap. This is because, for high-scorers, the average belief of White high-SES youth is much closer to that of Black youth than it is for Hispanic and low-SES youth. For all three comparison groups, differences in financial assistance, and the distribution of non-pecuniary utility play important roles in inequality relative to the reference group.

For the policy question, I find that the targeted college recruiting policy is the most efficient policy at reducing inequality between the reference group (White high-SES youth) and the three comparison groups. This is for two reasons. The first is that it is the most effective policy in reducing gaps between the reference group and the comparison groups.

The second reason is that it also decreases mismatch by encouraging more investment from high-scorers who would otherwise not enroll. The reduction in gaps is greatest for low-SES youth, suggesting college recruiting policies can be strengthened by targeting by race and ethnicity regardless of SES.

Free college for all decreases inequality but increases mismatch by encouraging more low-scorers that are overly optimistic in enrolling in college. The policy has a much smaller effect on high-scorers. As expected, the tracking system reduces mismatch. But on the other hand, it actually increases inequality. This is because the fraction of high-scorers in the comparison groups is much lower than for White high-SES youth. Additionally White high-SES youth also have inaccurate beliefs, so enrollment increases for them with information revelation as well [4].

Although the results of the targeted recruiting policy suggest that increasing college recruiting initiatives will increase representation among Black, Hispanic, and low-SES youth, gaps will likely still persist. As shown in the decomposition exercise and proportion of high-scorers in the comparison groups, this is because e differences in early childhood human capital development, and college experience remain important in explaining differences in bachelor's attainment.

1.1 Contribution to the Literature

This paper will contribute to the structural modeling literature that focuses on the role of information frictions in higher education decisions. One strand of the literature uses nationally representative panel data to study the role of information frictions in the decision to enter the high-skilled workforce by going to college (Navarro and Zhou 2017; Heckman,

[4] Although for White high-SES high-scorers beliefs are much closer to the truth than the high-scorers from the comparison group

Cunha Navarro 2005[5]). The second strand of the literature uses panel data from a single university that includes subjective beliefs along with grades to study the roll of belief formation on dropout and major choice (Arcidiacono, Hotz, and Kang 2012; Stinebrickner and Stinebrickner 2014; Wiswall and Zafar 2015; Reuben, Wiswall, and Zafar 2015).

This paper is closest to Arcidiacono, Aucejo, Maurel, and Ransom 2016. Like this paper I bridge the two strands of the literature together by examining enrollment, non-continuation, bachelor's attainment, and efficiency in the presence of information frictions. However, my paper differs in that this is the first paper to use data on reported beliefs about higher education outcomes and predicted model behavior to estimate the distribution of prior subjective beliefs about ability instead of assuming a rational expectations prior. In my estimation beliefs are allowed to differ by demographic background and allowed to differ from the distribution of ability estimated from the data. This is what allows me to estimate the role that differences in beliefs play in generating inequality in higher education outcomes.

This paper also contributes to the empirical literature demonstrating that information campaigns can increase enrollment and completion for high achieving students from lower income backgrounds. (Dynarski, Libassi, Michelmore, Owen 2020; Hoxby and Turner 2013; Bettinger Long, Oreopoulos, Sanbonmatsu 2012). The results of my policy analysis not only validate the findings of these papers but also show that if the policies studied in these papers were enacted at the national level, then they can increase representation in higher education as well as decrease mismatch across the United States.

[5]These papers estimate information sets by conducting factor analysis on the error terms of wage regressions along with regressing education choices on factors to test if factors were known at the time of the decision.

2 Empirical Analysis and Facts

Before discussing the model, this section will show that in the data subjective beliefs are highly correlated with probability of college enrollment, continuation, and completion while holding human capital and financial resources constant. Segments of this empirical analysis will inform the structural model. Specifically, the relationship between belief measures, financial assistance, and grades with education outcomes will be used as moments to identify model parameters. In section 2.1, I will go over the data set used in this analysis. In section 2.2, I discuss empirical facts in the NLSY97. Section 2.3 summarizes the empirical findings.

2.1 Data

The dataset I use to examine the relationship between subjective beliefs about education outcomes versus actual education outcomes is the 1997 wave of the National Longitudinal Study of Youth (NLSY97). The NLSY97 is a nationally representative longitudinal data set of individuals born between 1980-1984 living in the United States. The survey was administered annually from 1997 to 2011 and then biannually from there forward. The survey also over samples Black Americans and Hispanic Americans. This makes the NLSY97 useful for studying racial and ethnic inequality. The data also allows me to control for other important factors like access to financial resources, and measures of human capital that are also correlated with beliefs.

For this paper I use data on parental education, household net worth, self reported probabilities of school enrollment and obtaining a degree by age 30[6], labor market participation, earnings, schooling activities, financial assistance, and parental transfers. Additionally I use demographic information like race, ethnicity,census region, urban/rural categorical

[6]For individuals that are missing Probability of Degree, I impute it using the quantitative model equivalent to probability of degree; probability of enrollment times probability of continuation; using consecutive year estimates of probability of enrollment.

variables, gender, as well as year of birth.

In the quantitative analysis to control for the early childhood human capital stock and to estimate whether one is a high-scorer or low-scorer I use Armed Services Vocational Aptitude Battery (ASVAB) Math and Verbal scores as measures of cognitive human capital. I also control for non cognitive human capital by using indicator variables for participation in adverse behavior such as theft, violence, and sexual intercourse before age 15 (Heckman and Hai 2017).

For the analysis and the structural model that follows, the sample is restricted to adolescents who are not missing household net worth, parental education information, earnings in later years, ASVAB test scores, self reported beliefs before age 18 and self reported adverse behavior. I primarily rely on the transcript data for GPA. I impute transcript GPA for individuals that are missing GPA information from their college transcripts. For portions of the sample that have both transcript and self reported GPA I regress transcript GPA on self reported college GPA, demographic characteristics, and human capital measures. I then use the predicted values from the portion of the sample that only includes self-reported GPA.

I drop Asian Americans, Native Americans and races marked as other due to small sample sizes. For this reason I restrict the analysis to Hispanic, White, and Black youth. In total the sample size is 2,133 individuals. All statistics, regressions, and patterns in the empirical analysis are weighted using sampling weights[7] created by the Bureau of Labor Statistics for the NLSY97.

2.2 Empirical Facts

In this section we review some empirical facts in the NLSY97. Summary statistics by parental education and by race are reported in Appendix A.1 under Table 10 and Table 11. The summary statistics in the appendix show that Black, Hispanic, and lower education background

[7] The custom sampling weights for whether individuals are in all years of the sample is used

youth have lower enrollment and bachelor's attainment rates. They also have less access to resources measured by household net worth, and family financial aid in college. They have lower measures of human capital, as well as more pessimistic beliefs.

In Table 1 we control for parental education, race, ethnicity, access to financial resources, and measured human capital to see if self reported probability of having a degree is positively related to enrollment, continuation and bachelor's attainment. Table 1 shows that holding all else constant that being more optimistic is associated with a higher likelihood of enrollment, continuation and bachelor's degree attainment. This relationship continues to hold in Column 3 even with the inclusion of GPA, and financial assistance from schools, government and family.

Another interesting finding is that holding all else constant, being Black and Hispanic are associated with higher enrollment, and that being black is associated with higher degree attainment and completion conditional on enrollment. This is despite the fact that there are large unconditional gaps in these three variables. Unconditional gaps can be explained by differences in parental wealth and education, as well as human capital measures as shown in Table 1. [8].

In Table 2, we examine the relationship between beliefs and demographic characteristics holding human capital constant. The reason for sample size differences in Table 2, is due to the fact that the probability of degree question was only asked to the older cohort while they were high school aged, and for the younger cohorts probability of enrollment was asked while they were high school aged. [9]

We see in Table 2 that parental education and household net worth holding all else

[8] In Table 14 in appendix A.1, a Oaxaca Blinder decomposition shows that the unconditional gap in enrollment and completion for Black and Hispanic youth is explained in large part due to differences in parental education, household net worth, and measures of human capital. For Hispanics probability of degree is also important.

[9] For Table 1 and the quantitative analysis a measure of probability of degree is used that is imputed from subsequent years self reported beliefs of being in school during college age. Any bias in the imputed variable is controlled by the year of birth dummies in Table 1.

constant are associated with more optimism. Number of peers with college plans is also positively associated with more optimism. This is consistent with the findings of Hoxby and Avery 2012 that low-SES youth may know less about suitability for college resulting from less adults in their social networks that have higher education experience. These low-SES youth that exhibit behavior consistent with less information also have less peers with college attendance.

Similar to the last specification being Black or Hispanic is associated with more optimism regarding enrollment. Being Black is associated with more optimism regarding completion as well. This anticipates the finding that for Black high-scorers, beliefs have almost no explanatory role for lower bachelor's attainment rates than White high-SES high-scorers. The stronger role for beliefs for Hispanics may be driven by a higher proportion of Hispanic youth among the least educated group in the sample[10].

Figure 1 and Table 3 takes a closer look at non completion by group, measured by students who enroll and do not complete a 4 year degree, including those that enroll in community college and exit upon receiving an associates degree[11]. The graph as well as the interaction terms shows differences in continuation patterns among students who receive similar grades, but differ with respect to parental education, and optimism.

In terms of unconditional college exit rates, early human capital investments may still play an important role in generating outcomes. This is because differences in human capital investment would lead to differences in probability of achieving higher grades. Differences in non continuation rates between youth with similar grades suggests mechanisms other

[10]Table 13 in Appendix A.1 shows Oaxaca Blinder decomposition results for beliefs. The analysis shows that the unconditional gap for blacks is primarily explained by parental education, hh net worth, ASVAB and peers. This is true for Hispanics too, but there is an unexplained portion coming from the effect of parental education on beliefs.

[11]A youth is enrolled if they reported being enrolled in a 2 year or 4 year college degree or reported more than 12 years of schooling. Those enrolled in 2 year degree programs are included because they have the option to transfer credits to a four year university. Also according to Hoxby and Avery 2012 it is not obvious that this is always the cheapest option for college.

than human capital differences. This appears to be the case in Table 3 where we control for measures of human capital and financial resources and still find statistically significant coefficients for beliefs interacted with GPA category. Similarly, parental education is still marginally significant when interacted with GPA category for high grades. The smaller effect for high grades interacted with beliefs means that as agents become more optimistic the difference in continuation probability between medium and low grades shrinks.

The decrease in non-continuation with higher grades and the different effects grades have holding belief constant are consistent with the role of beliefs and learning about own ability through grades. This will be consistent with the belief mechanism in the quantitative model. In the model students begin with an original prior estimate of their ability and update with information revealed through grades. If higher beliefs mean more utility from college because of higher expected post college earnings, then more optimism should lead to weakly higher probabilities of enrollment, as in Table 1. If after grade revelation the new belief is proportional to the prior[12] than receiving the same grade but having a lower prior estimate means that your new belief would be lower, as in Table 3 for medium and high grades. Hence continuation probability would be weakly lower for more optimistic youth.

This learning mechanism and these patterns are consistent with Stinebrickner & Stinebrickner 2012, in which the authors found that a substantial portion of dropout (about 45%) is from learning about ability and earnings through grades. The rate of decrease also appears to be different by parental education, consistent with the hypothesis that exposure to college through parent's or other influential adult's experiences affects students own ratings of their own ability to succeed in college.

2.3 Discussion and Summary of Empirical Facts

We can summarize the findings from the last section as follows. In the NLSY97 we have

[12] As would be using Bayes Rule

1. Holding human capital, race, ethnicity, parental background, access to financial resources constant, youth who are more optimistic about college completion are more likely to enroll, more likely to continue by grade level, and hence more likely to obtain a bachelor's degree.

2. Holding human capital constant, youth from households with more wealth and education are more optimistic about college outcomes.

3. Holding human capital constant, Black and Hispanic youth are more optimistic regarding college outcomes than similar White youth. However there is a larger unconditional difference in beliefs between Hispanic and White youth[13], than for Black and White youth [14].

What this means is that we find evidence of a connection between subjective beliefs and college outcomes like enrollment, continuation, and degree attainment. Differences in human capital, subjective beliefs, and access to financial assistance by demographic group likely play a role in generating inequality in higher education outcomes as well.[15]

When compared to comparable White youth, Black Americans are more optimistic, have similar dropout rates by grade, and have better enrollment and completion rates. For Hispanic youth, relative to White youth, the story is slightly different. They are more optimistic regarding enrollment, and tend to enroll more than similar White students, but this is not the case with completion. The difference between Black and Hispanic youth may be due to more optimism among Black youth, as well as higher levels of government and institutional financial aid. Part of the differences in outcomes may also be due to Hispanic youth comprising a much larger portion of the students from the lowest education

[13] See Table 10 in Appendix A.1, where White youth are still more optimistic on average than Black and Hispanic youth

[14] See footnote 11

[15] As shown in Figure 16 appendix A.1. there is little evidence of differences in lower returns to college for Black, Hispanic, low familial wealth, and low parental education youth in the sample.

backgrounds while Black youth have parents with slightly better education levels as seen in Table 10 and 12-13 in appendix A.1.

The positive coefficient for Black youth on beliefs may also mean that information frictions may be less relevant for black youth then other demographic groups. In section 4, we will revisit to what extent beliefs explain different bachelor's attainment rates for high-scoring youth by demographic group.

In the section that follows I will propose a theoretical model that will be calibrated to match moments from the NLSY97 to show how differences in beliefs, along with differences in human capital, financial assistance, and non-pecuniary utility generate higher education decisions. Once the model is calibrated I will also discuss to what extent there is mismatch in the higher education market and whether any of the three policies that will be discussed can decrease inequality without generating more mismatch.

3 Economic Model

In this section I will propose an economic model that serves two purposes. The first is to demonstrate how differences in beliefs, differences in net tuition (tuition net of financial aid and family assistance), and early human capital development generates inequality in higher education outcomes by race, ethnicity, and parental education. Parameters of the economic model will also be be used to examine the role of beliefs in generating inequality by demographic group and the effects of policies on mismatch and inequality. The model predicts that lower net tuition as well as more optimism regarding latent college ability will have higher levels of enrollment, persistence, and completion, holding all else constant.

The economic environment will consists of agents who live $T = 24$ periods, where each period lasts 2 years and represents an age span from 18-66. This time frame coincides with the end of post secondary education up until retirement. In each period agents can save or borrow up to a specified borrowing limit. While working agents face the natural borrowing limit, but when agents are in school they face a more constrained borrowing limit. The 24 periods will be broken up into 3 stages; enrollment stage, continuation stage, and working stage. Since the worker's problem acts as an absorbing state, the education problem is essentially a two stage problem. This is because in this model once an agent chooses to work they do not return to school.

In the final work stage of the problem, post college earnings, $w_c(\tau_i)$, and non-pecuniary utility, $\mu_c(\tau_i)$, depend on an unknown type $\tau_i \in \{\tau_h, \tau_l\}$ for high-scorers and low-scorers. Where i subscript refers to agent i. The realization of τ_i depends on true probability $P_{\text{true},i}$ of being type τ_h. $P_{\text{true},i}$ depends on parental education, household net worth, race, ethnicity, sex, and measures of human capital. $P_{\text{true},i}$ will capture the positive relationship between early childhood human capital investment, educational attainment, and post college earnings. Since $P_{\text{true},i}$ also depends on demographics it can include a role for labor market discrimination as well.

In the model agents will not know $P_{\text{true},i}$ but they will have a subjective belief P_i. They then update P_i after receiving grades in college as in Stinebrickner & Stinebrickner 2012. P_i captures a broad belief about success at college for the individual, since the latent college ability type generates grades, earnings, and non-pecuniary utility from school.

A decision tree representation of the problem is shown in Figure 2. In the first stage at around age 18 agents have subjective belief P_i, asset level $b_{1,i}$, non-pecuniary utility shocks for work and school, $\vec{\varepsilon}_{1,i} = (\varepsilon_{c,1,i}, \varepsilon_{w,1,i})$.

Agents choose between enrolling in college where they pay net tuition fee $f_{1,i} = tuit_1 - Aid_{GC,i} - Aid_{Fam,i}$, or work and earn non college earnings w_n. Notice $f_{1,i}$ is equal to the sticker price $tuit_1$, net of aid from government or college $Aid_{GC,i}$ and families $Aid_{Fam,i}$.

In the second stage at around age 20 agents realize a signal for their latent type given by the GPA g_i for the previous schooling period. They then update P_i to $P'(P_i, g_i)$, and observe non-pecuniary utility shocks for school and work, $\vec{\varepsilon}_{2,i} = (\varepsilon_{c,2,i}, \varepsilon_{w,2,i})$. They then decide to continue schooling and pay net tuition $f_{2,i} = tuit_2 - Aid_{GC,i} - Aid_{Fam,i}$, or dropout to work and earn w_s for having completed some college. Like $f_{1,i}$, $f_{2,i}$ is the sticker price $tuit_2$ in period 2 net of aid from government or college $Aid_{GC,i}$ and families $Aid_{Fam,i}$. Government, college and family aid are assumed to be equal for period 1 and 2, while sticker price is allowed to be different.

If agents choose to complete school then after the next period from ages 22-66 agents work and earn earnings dependent on type, $w_c(\tau_i)$ each year as well as non monetary utility $\mu(\tau_i)$. Agents make borrowing saving decisions in all periods of the problem whether in school or in the labor force. During School the borrowing limit is $-B_s(t)$ while in the labor force it is $-B_n(w)$, with $-B_s(t) \geq -B_n(w)$ so that credit constraints are more binding while enrolled in school (Lochner and Monge-Naranjo 2012).

Heterogeneity by parental background, race, and ethnicity enters the problem through four channels. The first is through the distribution of initial subjective beliefs P_i. Second through transfers from parents, government, and institutions that lead to differences in net tuition $f_{t,i}$ for $t = 1, 2$ while in school. Third through the true probability of being type τ_h, $P_{\text{true},i}$ which determines the distribution of grade realizations and future earnings. Finally through the distribution of non-pecuniary utility shocks $\vec{\varepsilon}_{t,i}$.

In order to allow for human capital development while in school, mean earnings are such that $w_n < w_s \leq w_c(\tau_i)$ reflecting increasing mean returns to years of schooling regardless of type. Even though expected earnings increase with schooling, a binding credit constraint while in school will make college much less appealing for those with $\tau_i = \tau_l$. This is because agents will be unable to consumption smooth and face lower consumption for the first two periods of their life.

3.1 Workers Problem

At any time period t, for all three stages the workers problem is given by (1) below. Where utility depends on, assets/debt b_i, earnings w, and t since this determines how many periods agents have left in their life cycle.

$$(1) \qquad V_w(w, b_i, t) = \max_{\{b_{n,i} \geq -\tilde{B}_n(w)\}_{n=t}^{T}} \sum_{n=t}^{T} \beta^{n-t} u(w + Rb_{n,i} - b_{n+1,i})$$

Per period utility $u(\cdot)$ is given by CRRA preferences of the form

$$(2) \qquad u(c_i) = \frac{c_i^{1-\gamma} - 1}{1-\gamma}$$

For every period the borrowing constraint is the natural borrowing limit, given below. The natural borrowing constraint is determined by how much the agent can credibly pay back in the future. As a result in the final period T, agents are not allowed to borrow.

$$\tilde{B}_{T-n}(w) = \frac{w + \tilde{B}_{T-n+1}(w)}{1+r} \qquad \tilde{B}_T = 0$$

Therefore in the final period $b_{T+1} = 0$.

3.2 Enrollment Work Problem

In the first stage, corresponding to age 18, agents make the decision to enroll in school or work starting at period one until the end of the life cycle. Agents begin with initial assets $b_{1,i}$, unobserved tastes for college and work $\vec{\varepsilon}_{1,i} = (\varepsilon_{c,1,i}, \varepsilon_{w,1,i})$ a belief P_i that they are of type τ_h. The agent's stage 1 problem is thus given by (3) below.

$$(3) \qquad V_1(P_i, b_{1,i}, f_{1,i}, \vec{\varepsilon}_{1,i}) = \max\{V_w(w_n, b_{1,i}, 1) + \varepsilon_{w,1,i}, V_{c,1}(P_i, f_{1,i}, b_{1,i}) + \varepsilon_{c,1,i}\}$$

$$s.t.$$

$$V_{c,1}(P_i, f_{1,i}, b_{1,i}) = \max_{b_{2,i} \geq -\tilde{B}_{s,1}} [u(Rb_{1,i} - f_{1,i} - b_{2,i}) + \beta \mathbb{E}_{g,\varepsilon}(V_2(P'(g, P_i), f_{2,i}, b_{2,i}, \vec{\varepsilon}_{2,i})) \; |P_i \;]$$

Agents update beliefs after realizing grades using Bayes Rule according to equation (4), where the new belief $P'(g, P_i)$ is given below. Where $\pi_{k,j} = Prob(g_k|\tau = \tau_j)$.

$$(4) \quad P'(g_k, P_i) = \frac{P_i \pi_{k,h}}{P_i \pi_{k,h} + (1 - P_i)\pi_{k,l}}$$

3.3 Completion Dropout problem

In the second stage, corresponding to age 20, agents make the decision to continue and complete college or dropout and work for the remainder of the life cycle. Agents observe GPA g from the first stage then update belief P_i to $P'(g, P_i) = P_i'$. Agents also begin the second stage with debt/savings from the first stage $b_{2,i}$, and realize unobserved tastes for college and work $\vec{\varepsilon}_{2,i} = (\varepsilon_{c,2,i}, \varepsilon_{w,2})$ respectively. The agent's problem is given by

$$(5) \quad V_2(P_i', f_{2,i}, b_{2,i}, \vec{\varepsilon}_{2,i}) = \max\{V_w(w_s, b_{2,i}, 2) + \varepsilon_{w,2,i}, V_{c,2}(P_i', f_{2,i}, b_{2,i}) + \epsilon_{c,2,i}\}$$

$$s.t.$$

$$V_{c,2}(P_i', f_{2,i}, b_{2,i}) = \max_{b_{3,i} \geq -\tilde{B}_{s,2}} [u(Rb_{2,i} - f_{2,i} - b_{3,i}) + \beta(P_i'[V_w(w_c(\tau_h), b_{3,i}) + \mu_c(\tau_h)]$$
$$+ (1 - P_i')[V_w(w_c(\tau_l), b_{3,i}) + \mu_c(\tau_l)]) \;\;]$$

Grades reveal information about τ_i since their distribution depends on τ_i. But since τ_i also determines non-pecuniary utility, the information revealed in school can also include psychosocial elements of higher education that are often discussed in the sociology literature. In this model, the assumption is that this is closely tied to performance, and a bad signal in performance will likely reinforce that college will not be a good fit for the individual. Factors that are likely to be more stable between the first and second period such as distance from home community, enjoyment of school, and family obligations would be captured through a constant location parameter of non-pecuniary shocks $\vec{\varepsilon}_{t,i}$. The location parameter will be allowed to differ by demographic group.

Since in the first two stages the agent faces a discrete choice problem, the optimal decision for each agent can be described by a cutoff rule with respect to belief about type. Where if P_i is higher than a certain threshold the agent will enroll. For example in the first stage the optimal decision could be characterized by equation 6 below, where $\sigma_{d,2}, \mu_{d,2,i}$ are the normalized scale parameter and location parameters[16] for the Type I extreme value shocks.

$$\text{(6)} \qquad \text{Choice}_{t=1,i} = \begin{cases} \text{Enroll} \quad \text{if} \quad P_i > \tilde{P}_1(b_{1,i}, f_{1,i}, f_{2,i}, \vec{\varepsilon}_{1,i}, \mu_{d,2,i}, \sigma_{d,2}) \\ \text{Work} \quad \text{if} \quad P_i \leq \tilde{P}_1(b_{1,i}, f_{1,i}, f_{2,i}, \vec{\varepsilon}_{1,i}, \mu_{d,2,i}, \sigma_{d,2}) \end{cases}$$

.

Similarly in stage 2, given $\{\pi_{k,j}\}_{k,j}$ the decision to continue also follows a cutoff rule for updated belief $P'(g_k, P_i)$ after realizing $g_k, \varepsilon_{c,2,i}, \varepsilon_{w,2,i}$ and starting with P_i, given by equation (7) below.

$$\text{(7)} \qquad \text{Choice}_{t=2,i} = \begin{cases} \text{Continue} \quad \text{if} \quad P'(g_k, P_i) \geq \tilde{P}_2(b_{2,i}, f_{2,i}, \vec{\varepsilon}_{2,i}) \\ \text{Dropout} \quad \text{if} \quad P'(g_k, P_i) < \tilde{P}_2(b_{2,i}, f_{2,i}, \vec{\varepsilon}_{2,i}) \end{cases}$$

.

The cutoff rules holding non monetary utility shocks and distribution constant, are weakly increasing in $f_{1,i}, f_{2,i}$. In certain spaces of the distribution of non monetary utility

[16]Normalized with respect to the difference in Type I extreme values. Since the difference in shocks is what is identified

shocks the decision rules are strictly increasing in $f_{1,i}, f_{2,i}$. $P'(g_i, P_i)$ also increases in P_i. For $P_i \in (0,1)$, if higher grades provide a stronger signal of being $\tau_i = \tau_h$, $P'(g_i, P_i)$ will increase in g_i. Therefore depending on the underlying distribution of grades and earnings by type, as well as non monetary utility shocks, the model can reproduce some of the facts discovered in the empirical analysis. Particularly higher levels of financial assistance may allow more pessimistic students to enroll. On the other hand for students with low levels of financial assistance more optimistic initial beliefs along with better grades will allow more students to enroll and complete college.

3.4 Example for Model Prediction

Figures 3, 4, and 5 provide an example for how financial assistance, subjective beliefs of having $\tau_i = \tau_h$, and early childhood human capital investment realized through grades affect probability of enrollment, continuation, and degree attainment in the model.

In Figure 3 we see that more financial assistance through lower net tuition leads to a higher probability of enrollment at all belief levels. Because of the belief cutoff we also see that probability of enrollment is flat and then increases with subjective beliefs at all net tuition levels displayed. This means that if two youth have the same beliefs but different access to resources than probability of enrollment will still be different. Like wise if their access to resources are the same but beliefs differ than probability of enrollment can also differ.

Figure 4 shows conditional on already enrolling, how probability of continuation differs by grade revelation. This shows how learning affects the continuation decision. We can see that high grades which are a signal of being a high-scorer. After a certain belief threshold higher grades lead to an increase in the probability of continuation. The difference in continuation probability between high grades and medium grades diminishes as a youth gets more optimistic [17]. For the same initial belief, lower grades lead to a dramatic decrease in continuation except in the case where the agent is near certain they are a high or low scorer.

Figure 5 takes probability of enrollment and continuation together and shows that even though the effect of net tuition is somewhat more muted than in Figure 3, net tuition and subjective beliefs about being a high-scorer still affect the probability of degree attainment in the model.

Some discussion on the assumptions of the model is warranted. First is the fact that college decisions happen only in the first four years of life. The model can be changed to allow for switching back from work to school, that way later enrollment decisions can be allowed for as well as the saving up of assets for school. The main object of interest will be bachelor's attainment period not so much timing of enrollment. Any bias introduced by ignoring systematic differences in later enrollment would be captured in the non-pecuniary utility components that are allowed to differ by demographic group.

The second objection raised can include the fact that beliefs about probability of acceptance that also depend on τ_i is not included. The model can also be adjusted for this, where probability of acceptance could depend on latent type and agents can learn from acceptance as well. However since in the data we look at any enrollment including at non selective community colleges we do not include it. For these colleges acceptance probability is likely close to one. Even if youth enroll in community college, the possibility still exists to transfer to a four year university.

Finally uncertainty regarding tuition is not explicitly modeled. In the calibration this is indistinguishable from allowing the net tuition rate to change between periods. This is because over estimating tuition and learning it were lower after enrollment would be equivalent to $f_{1,i}$ being higher than $f_{2,i}$. However in the calibration this would be an average

[17]this is consistent with the standalone and interacted coefficients for GPA category estimated in Table 3

uncertainty in tuition not dependent on demographics, which can be adjusted in future formulations. We can also extend the model where second period net tuition depends on latent type, grade realization or human capital measures. For simplicity we will use average tuition by demographic as in Hai and Heckman 2017.

In the next section we will discuss the calibration of the model. The model will then be used to discuss how beliefs, financial assistance, and human capital investment with earnings outcomes affect gaps in educational attainment, with the focus on bachelor's attainment. Then given this role we will discuss the effect that universal free college and targeted college have on inequality and mismatch.

4 Quantitative Analysis

In this section I discuss how I identify and estimate the parameters of the structural model described in section 3. I will also describe some of the assumptions governing the distribution of earnings as well as parameters whose values will be set outside of the estimation routine. As part of this I will discuss what data moments that will be used to identify parameters related to the main mechanisms of the model.

As section 3 suggested the model will include room for differences by race, ethnicity, and parental background in financial assistance, and non-pecuniary utility. The financial assistance measure used in the model will be the predicted financial assistance given demographic information as in Heckman and Hai 2017. This will be estimated separately for financial assistance that comes from family and for financial assistance that comes from universities or government[18]. In addition to type dependent non-pecuniary utility in the final stage, the distribution of non-pecuniary utility shocks will also depend on race, ethnicity, and whether the student is a potential first generation college student[19]. This is to control

[18] Including work study

[19] That is if their parents had ever attended college

for any inherited preferences for college by parental education or for unobserved differences in the college experience for minorities.

The distribution of beliefs will be a linear function of measured beliefs about college completion from the NLSY97 and parental education. Systematic differences by race, ethnicity and socioeconomic status in measured beliefs[20] will allow for systematic differences in the estimated distribution of subjective beliefs. The estimated distribution of subjective beliefs will have an extra effect for parental education to capture information gained through parents education experiences that might not be captured in the measured beliefs. In addition to systematic differences in cognitive and non cognitive measures, the estimation of P_{true} will also allow for an extra effect to come from race, ethnicity and parental background. This can capture unobserved human capital and discrimination.

First I will explain the externally estimated parameters. Then I will explain the moments and parameters that will be estimated using indirect inference with some discussion on how the model is identified.

4.1 External Parameters

The parameters that will be set outside of the model are given in Table 4. The coefficient of relative risk aversion γ, the discount factor β, and the interest rate $(1 + r)$ are set to standardly assumed values. The college borrowing limits are set to match average student debt levels as in Abbot Gallipoli, Meghir, and Violante 2016. The first stage borrowing limit while in school is set to \$16,600 in 2017 dollars. The second period borrowing limit is set to \$31,100. Together these match average borrowing for the first two years and last two years of college respectively (Wei and Skomsvold, 2011). In total the amount students are allowed to borrow in the model is higher than the highest cumulative total that students could borrow

[20] Table 2 shows all else equal parental education is positively correlated with optimism regarding degree attainment. Table 10 shows average beliefs differ by race and ethnicity with Hispanics being the most pessimistic.

from Federal student loan programs for a bachelor's degree, \$46000, which likely reflects the use of private loans amongst some students (Lochner and Monge Naranjo 2010).

Financial assistance is estimated outside of the model. Where financial assistance is the sum of family aid and government/institution financial aid. The distribution of financial assistance is drawn from a log normal distribution, of the form below, estimated by OLS.

$$(8) \qquad \ln(f_{i,k}) = X_i\beta_{f,k,} + \epsilon_{f,k,i}$$

Where X_i includes demographic variables like race, ethnicity, gender, household net worth, parental education, year of birth, and a constant term. The subscript k indicates that Equation 9 above is estimated separately for family assistance $k = 1$ and government/institution financial assistance $k = 2$. To get total financial assistance, the sum of both predicted values for students is used. Therefore financial assistance used in the model is the predicted value given by demographic and socioeconomic variables (Hai & Heckman 2017).

The distribution of latent type τ by demographic group will be estimated using a finite mixture model (FMM). The latent variable will take two values for $\tau_i \in \{\tau_l, \tau_h\}$, respectively corresponding to low-scorers and high-scorers in the rest of the paper. The value τ_l is normalized to $\tau_l = 0$. The value of τ_h will be estimated through the estimation of the three following measurement equations in the finite mixture model.

$$(9) \qquad Z^*_{i,j} = \alpha_{z,j}1(\tau_i = \tau_h) + \eta_{z,j}X_i + \varepsilon_{z,j} \qquad j \in \{1, \ldots, J_c\}$$

$$(10) \qquad \ln w^*_{i,s} = \mu_{w,0} + \mu_{w,1}1(s \in (12, 16)) + 1(s \geq 16)(\mu_{w,2} + \mu_{w,h}1(\tau_i = \tau_h)) + \varepsilon_{w,s}$$

$$(11) \qquad \pi(g|\tau) = \frac{exp(\gamma_{g,0} + \gamma_{g,\tau}1(\tau_i = \tau_h))}{\sum_{k=l,m,h} exp(\gamma_{k,0} + \gamma_{k,\tau}1(\tau_i = \tau_h))}$$

In equation (9) $Z^*_{i,j}$ are measures of cognitive and non cognitive ability. The measures of cognitive ability are the ASVAB scores for arithmetic reasoning, paragraph comprehension, word knowledge and mathematical knowledge. The non cognitive measures are participation in adverse behavior at young ages; sex before age 15 as well as any violence and any theft greater than 50$ at the start of the survey. To incorporate both binary and continuous variables the specification below for $Z^*_{i,j}$ will be estimated in the FMM. The choice of human capital measurements and specification for $Z_{i,j}$ follows Hai and Heckman 2017.

$$Z_{i,j} = \begin{cases} Z^*_{i,j} & \text{if } Z^*_{i,j} \text{ is continuous} \\ \boldsymbol{1(Z^*_{i,j})} & \text{if } \boldsymbol{Z^*_{i,j},} \text{ is binary} \end{cases} \qquad i \in \{c, n\}$$

Log earnings dependent on years of schooling s for individual i are described in equation (10) by $\ln(w_{i,s})$. The variance of the error term is allowed to differ for whether a student has no college experience $s \leq 12$, some college experience $s \in (12, 16)$, or a bachelor's $s \geq 16$. The expected value of the exponential of $\ln(w_{i,s})$ will be used for w_n, w_s, w_l, w_h in the model[21].

Additionally the distribution of grades $g \in \{g_l, g_m, g_h\}$ for low ($GPA < 2.0$), medium ($2.0 \leq GPA < 3.0$), and high ($3.0 < GPA$), conditional on τ is estimated using equation (11) above. To describe the distribution of grades by demographic type, the distribution of type by demographic is also needed. The distribution of type will also be important for

[21]Earnings and financial assistance are set to 2017 dollars

the effect of the policies discussed later in the paper. This is described fully by $P(\tau_h; \vec{X}_i) = Prob(\tau_i = \tau_h|\vec{X}_i)$ in equation 12 below. Since $P(\tau_l; \vec{X}_i) = 1 - \ Prob(\tau_i = \tau_h|\vec{X}_i)$.

$$(12) \qquad P(\tau_h; \vec{X}_i) = Prob(\tau_i = \tau_h|\vec{X}_i) = \frac{exp(\vec{X}_i\vec{\beta}_p)}{1 + exp(\vec{X}_i\vec{\beta}_p)}$$

Using equations (9)-(12), human capital measurements, earnings, and grades from the NLSY97, the FMM can be estimated using the individual likelihood function given by $f(\vec{Z}_i, w_i, g_i; \tau_k, X_i, s)$. The objects of interest are the parameters in (9)-(12). These objects are estimated by solving for the maximum simulated likelihood given below in equation (13). The likelihood equation will control for probability of enrollment and continuation given demographic information as well. For more detailed information regarding the functional form of the likelihood function as well as the parameter results of the individual likelihood function see Appendix A.2-A.3.

$$(13) \qquad \max \quad \sum_i \ln[P(\tau_h; \vec{X}_i) f(\vec{Z}_i, w_i, g_i; \tau_h, X_i, s_i) + (1 - P(\tau_h; \vec{X}_i)) f(\vec{Z}_i, w_i, g_i; \tau_l, X_i, s_i))]$$

Once equation (8), along with the finite mixture model given by equation (9)-(13) are estimated, we use the sum of the predicted financial assistance variables for total financial assistance, predicted earnings $\ln w_{i,s}$ for w_n, w_s, w_l, w_h, and $\pi(g|\tau)$ for the conditional grade probabilities in the model. The individual probability of being a high-scorer that will used for individual $P_{\text{true},i}$ is explained below in equation (14).

$$(14) \qquad P_{\text{true},i} = Prob(\tau_i = \tau_h|\vec{X}_i, \vec{Z}_i, w_i, g_i, s_i) \propto P(\tau_h; \vec{X}_i) \times f(\vec{Z}_i, w_i, g_i; \tau_h, X_i, s)$$

$P_{\text{true},i}$ is therefore the posterior probability of being a high-scorer estimated from the finite mixture model. In this way the distribution of individual grades in college as well as the probability of completing college will depend on the individual early human capital stock given by the vector $\vec{Z}_i$. $P_{\text{true},i}$ will be used to simulate high types and low types in the quantitative model.

In the policy counter factual in section 5 a modified posterior will be used that only takes into account information that would be available to policy makers or school administrators prior to college, ASVAB scores and behavior indicators[22].

4.2 Internally Estimated Moments

The remaining moments to be calibrated within the model are the sticker price of tuition $tuit_1$, $tuit_2$, the distribution of subjective beliefs of being type τ_h, the non-pecuniary utility dependent on τ, $\mu_c(\tau_i)$, as well the distribution of preference shocks. The distribution of preference shocks is given by the Type I extreme values shocks whose location parameters differ for White, Black, and potential first generation students. The variance for the Type 1 extreme value shocks for the first and second period will also be allowed to differ. For a quick summary of how the distributions of the internal parameters are calibrated see Table 5.

The distribution of subjective beliefs of being high type is given by a truncated normal distribution at zero and one given by equation (15) below.

$$p_i = \gamma_{p,0} + \gamma_{p,b} ProbDegr + \gamma_{p,h} Pedu_{\text{hsg}} + \gamma_{p,s} Pedu_{\text{scol}} + \gamma_{p,b} Pedu_{\text{bach}} + \epsilon_{p,i} \tag{15}$$

The assumption used in equation (14) is that data contained in the variable Prob Degr from the NLSY97, is a noisy measurement of the subjective belief of being type τ_h. The

[22]Specifically $Prob(\tau_i = \tau_h | \vec{X}_i, \vec{Z}_i) \propto P(\tau_h; \vec{X}_i) \times f_Z(\vec{Z}_i; \tau_h, X_i)$ will be used.

measurement error is allowed to differ by parental education. This is to capture information about college that youth may receive from their parents higher education experiences. A truncated normal is used since we want to allow for 1's and 0's since these are meaningful in the model. A value of 1 or 0 means that agents are certain of their type and hence will not change their mind with grades.

The distribution of type 1 extreme value shocks, non-pecuniary utility by type $\mu_c(\tau_i)$ and the parameters in equation (15) will be internally estimated by indirect inference. Standard errors for the parameters will be estimated by bootstrapping. The moments that will be targeted in the indirect inference specification are the coefficients for the following two regressions in equation (16) and (17).

$$(16) \quad Enroll = \beta_{E,0} + \beta_{E,B} HighBelief + \beta_{E,F_2} T2(Finaid) + \beta_{E,F_3} T3(Finaid)$$

$$+\beta_{E,1G} FirstGen + \beta_{E,W} White + \beta_{E,H} Hisp + \varepsilon_{E,i}$$

$$(17) \quad Continue = \beta_{C,0} + \beta_{C,G_m} GPA_m + \beta_{C,G_h} GPA_h + \beta_{C,F_2} T2(Finaid) + \beta_{C,F_2} T3(Finaid)$$

$$+\vec{\beta}_{C,PH} Pedu_{\text{hsg}} + \vec{\beta}_{C,PS} Pedu_{\text{scol}} + \vec{\beta}_{C,PB} Pedu_{\text{bach}} + \beta_{C,W} White + \beta_{C,H} Hisp + \varepsilon_{C,i}$$

Where $FirstGen$ is an indicator for being a first generation student, $HighBelief$ is an indicator for being in the top half of belief distribution, $T2(Finaid), T3(Finaid)$ are indicators for being in the 2nd and third terciles of the total financial assistance distribution. The parameter vector Γ are those parameters that minimize the difference between the simulated regression coefficients and data regression coefficients. The specific problem that Γ solves is given below, in equation (17). $\tilde{\beta}(\Gamma)$ are the simulation coefficients given Γ, while $\vec{\beta}$

are the regression coefficients from the data. W is the weighting matrix given by the inverse of the diagonal matrix of the standard errors of the data regression coefficients.

$$(17) \qquad \min_{\Gamma}(\tilde{\beta}(\Gamma) - \vec{\beta})' W (\tilde{\beta}(\Gamma) - \vec{\beta})$$

Important parameters are reported in Table 7. For the complete list of estimates see Table 21 Appendix A.4. Using the calibrated and preset parameters we can then decompose high-scorer inequality by differences in financial aid, subjective beliefs, and non-pecuniary utility. Overall gaps would also be determined by P_{true}. We can then evaluate the effects of policies on inequality and mismatch in higher education by race, ethnicity and parental background.

4.3 Identification Discussion

Equations (16) and (17) essentially match the two main stages of the model where education choices are made. This is stage (1), the enrollment vs work choice and stage (2), the continuation vs exit and work stage. The main parameters of interest in this estimation are the distribution of beliefs about type that is given by equation (15). Estimation is aided through the external estimates of earnings by schooling choice and type as well as the conditional grade probabilities given type.

Identification of beliefs depends on two crucial features of the data. One is that enrolling is positively correlated with measured beliefs in the data as captured by $\beta_{E,B}$, controlling for access to resources. The second is the difference in college continuation by GPA category. This is given by β_{C,G_m} and β_{C,G_h}.

To aid in understanding, first abstract away from the role of financial assistance, race, ethnicity and parental education in (16), (17), and (15). Beliefs given by p_i only matter to the extent that utility from completing college for high-scorers is greater than utility from completing college for low-scorers, as well as the relative utility of non college and some college. All of these depend on $w_c(\tau), w_s, w_n$. The difference in expected utility between school and work is also determined by the mean difference of type I extreme value shocks for school and work, which is assumed to be the same in stage 1 and stage 2 of the problem. The difference in utility from college between low and high types is also determined by $\mu_c(\tau)$.

The empirical results discussed in section 2 requires that the parameter space is such that beliefs matter. This is primarily through the coefficient on measured beliefs and the effect of grades on enrollment in equations (16-17). The top panel in Figure 6 also shows that grades do provide a signal for type in the finite mixture model which is consistent with the model. Therefore in the quantitative model there should be an increase in enrollment probability with more optimism as well as a differential response to grades in the continuation stage.

If we focus on the enrollment stage and equation (16) $\gamma_{p,b}$ in (15) is primarily identified through $\beta_{E,B}$ in (16). Since this determines how important measured beliefs are which enter the model through subjective beliefs about type.

The location of beliefs given through $\gamma_{p,0}$ is identified through difference in response to GPA along with $\mu_c(\tau)$. Panel 2 in Figure 6 shows that if prior stage 1 beliefs are too optimistic or too pessimistic then there is little change in beliefs. Therefore the degree to which grades affect updating and hence continuation depends on the location of the distribution of the prior. If estimated beliefs are located near the center of $(0, 1)$ then here changes in beliefs will lead to the biggest updates and hence biggest grade response. Therefore $\gamma_{p,0}$ will be set to where this best matches equation (17) from the data. The period constant location parameter for the type I extreme value shocks is constant in period 1 and 2, and would help match the levels of enrollment and continuation independent of grades. The type dependent non-pecuniary utility helps to adjust response to grades if the response implied by the finte mixture model through earnings and $\pi(g_k, j)$ is too restrictive.

Responses to financial aid in enrollment and continuation given by $\beta_{E,F_2}, \beta_{E,F_3}, \beta_{C,F_2}, \beta_{C,F_3}$ will identify $tuit_1$ and $tuit_2$. This is because financial assistance is externally estimated and $tuit_1$ and $tuit_2$ will set net tuition rates by demographic group which also play an important role in the higher education decision given by the quantitative model.

Differences in the location parameter by race and ethnicity will be identified through the effect of race and ethnicity in equation (16), given by $\beta_{C,W}, \beta C, H$. The effect of being a first generation college student on the difference in the location parameter is identified through $\beta_{E,1G}$. The effects of parental education on beliefs are identified by $\beta_{C,PH}, \beta_{C,PS}, \beta_{C,PB}$ in equation (17). Including the variance term in (15) and the scale parameter of the type 1 extreme value shocks leads to 16 parameters that will be estimated by 16 moments.

4.4 External and Internal Estimation results

Table 6 shows the results from the wage equation of the finite mixture model. We see that regardless of type, log annual earnings increase with education. As expected enrolling and completing school will lead to higher earnings for all youth, regardless of scorer type. However high-scorers have higher earnings than low-scorers upon completing college.

If there were no non-pecuniary utility and credit constraints than all youth would choose to enroll and complete college. However in the presence of binding credit constraints the lower utility for the next two periods brought about by very low consumption may deter some youth from pursuing education, especially if they believe they will incur some non-pecuniary utility costs from being a low-scorer.

Table 7 shows several of the key parameters that were estimated in the internal calibration exercise. The factor loading on probability of degree attainment, is 0.87 with a very precise standard error estimate. This suggest that this variable does capture beliefs about being a high-scorer with $\tau_i = \tau_h$. Holding the measured belief constant as well, the higher education background a youth comes from the more optimistic they are that they are type τ_h. This is consistent with the hypothesis that youth who know more adults with college education will rate their fit for college higher.

Figures 7-9 provide a quick snapshot of how well the model matches patterns we see in the data. Figure 8 and the left side graph of Figure 7 show that the model slightly underestimates enrollment and non completion. However on balance it has a good fit with regards to BA attainment. As we can see from Figure 9 and the left side of Figure 7, this success at capturing BA attainment carries over when we condition by demographic group as well. The quantitative model matches gender and household net worth bachelor's attainment even though these were not directly targeted in the indirect inference specification. For more information regarding model fit of the indirect inference moments see Appendix A.4.

5 Quantitative Exercises: Decomposition, Mismatch, & Policy Analysis

In this section I will discuss the role beliefs, financial assistance, and non-pecuniary utility play in generating inequality in higher education among high-scorers. Then I will discuss how estimated beliefs differ from $P_{\text{true},i}$ as well as actual type to anticipate the effects of policies discussed in the next subsection. Finally I will discuss the results of the policy analysis. In the policy analysis a college recruiting policy targeting information and subsidies to low-SES predicted[23] high-scorers will be compared to free college for all, and the institution of a tracking system where estimates of type are revealed to everyone.

In the quantitative exercise high-scorers will be individuals with a simulated $\tau_i = \tau_h$ where type is simulated using the FMM posterior probability from equation (14). Policies will be evaluated primarily by how they decrease gaps in bachelor's attainment. They will also be evaluated in terms of the amount of mismatch generated or Ex-Post regret as in the

[23]Predicted using only ASVAB and behavior as in footnote 10.

information frictions literature. It is measured by the amount of youth who would make different education decisions if they knew their type with certainty. This mismatch takes the form of under investment of high-scorers, and over investment of low-scorers.

For the decomposition exercise as well as the policy analysis I will focus on the difference between the reference group White high-SES youth compared to the three reference groups; Black, Hispanic, and low-SES youth . For the decomposition exercise I will focus on differences among high-scorers only, since this is where the inefficiency is most poignant. While in the policy analysis I will focus on overall differences regardless of latent ability between the reference group and comparison groups. In essence we are asking if increasing representation of high academic achievers from disadvantaged backgrounds is enough or a better way to close gaps than policies that target everyone equally.

In this section low-SES youth, are those whose household is in the bottom tercile of the net worth distribution, or whose parents have a high school diploma or less. High-SES youth are those that come from the top tercile of the wealth distribution and whose parents have at least a bachelor's degree.

5.1 Differences in Beliefs and Inequality

Figure 10 shows the mean values of variables corresponding to the two main mechanisms generating differing education decisions between Black, Hispanic, low-SES, and White high-SES high-scorers; subjective beliefs, and financial assistance. Since these graphs are conditioning on high-scorers human capital differences will not play a role.

Figure 10 shows that White high-SES high-scorers receive much more financial assistance and are more optimistic about being high-scorers than Black, Hispanic, and low-SES high-scorers. However when considering subjective beliefs, the average prior belief for White high-SES high-scorers is much closer to that of Black high-scorers than the other groups. This suggests beliefs will not play as big of a role in explaining the gap for Black high-scorers as it will for Hispanic and low-SES high-scorers.

Next I decompose the gap in higher education backgrounds for high-corers by differences in subjective beliefs, financial assistance, and unexplained sources brought about by differences in non-pecuniary utility. The decomposition is performed by sequentially setting beliefs and financial assistance equal to the mean values of White high-SES high-Scorers.

Figure 11 provides a visual representation of the reduction in the difference between the reference group and the three comparison groups. Table 8 shows the percentage contribution of each mechanism in the college completion gap. The graph and Table 8 show that for most demographic groups the biggest portion of the gap in college completion is due to unequal financial assistance, mostly through family assistance. We see for Hispanic and low-SES youth, differences in beliefs contribute to between 38 to 49 % of the gap, while for Black youth beliefs may contribute very little given by the large standard error in parenthesis.

Financial assistance explains between 45-50% of the gap for all three comparison groups Differences in non-pecuniary utility also appear to play an important role for all three groups in generating inequality relative to White high-SES high-scorers, where they range from explaining 6% of the gap for Hispanics to close to 17% for Black and low-SES youth. Since beliefs and financial assistance, the college recruiting policy that targets information and funding to low-SES high-scorers is likely to increase completion among the three comparison groups. This is because Black and Hispanic youth comprise a significant portion of low SES youth as suggested by the summary statistics in Table 10 and 11 in Appendix A.1.

5.2 Information Frictions and Mismatch

In the model there are differences in beliefs between demographic groups, difference in beliefs relative to $P_{\text{true},i}$, as well as difference in beliefs relative to the truth or actual type of individual agents. Differences in mean beliefs of the comparison groups relative to the mean beliefs of the reference group leads to different enrollment and continuation rates as discussed in the last section. Differences relative to estimated probability from the finite mixture model $P\text{true}, i$ will determine the effect of policy providing more accurate estimates of P_{true} to individuals as in the targeted college recruiting and tracking policy. Differences in beliefs relative to the actual type of the individual will affect the measure of efficiency used in this model, mismatch. Where mismatch is under investment among high-scorers who would have higher bachelor's attainment rates under perfect information. It also includes over investment among low-scorers who would have lower BA attainment rates if they knew their types.

This subsection will briefly go over differences in mean beliefs relative to $P_{\text{true},i}$ as well as differences in mean beliefs relative to actual type.

Figure 12 shows the difference in mean subjective beliefs vs estimated $P_{\text{true},i}$ by demographic group, estimated from the data and finite mixture model for the demographic groups considered in the last section. We see that for Black, Hispanic, low-SES, and White high-SES youth subjective beliefs are on average inaccurate compared to the finite mixture model. High-scorers are too pessimistic and low-scorers are too optimistic on average. On average Black, Hispanic, and low-SES high-scorers know something about their type since their beliefs are more optimistic than low-scorers. For White high-SES youth beliefs are almost the same between low and high-scorers.

High-scorers are also closer to the truth than low-scorers. Therefore we should expect policies revealing estimates of $P_{\text{true},i}$ will have differential effects by type. Therefore tracking should lead to a huge readjustment of low-scorers beliefs and a decline in bachelor's attainment. Depending on the proportion of high-scorers and low-scorers by demographic group this may actually increase inequality.

Similarly for the targeted policy we should see that it should close gaps the most for low-SES youth, since the difference in beliefs with respect to $P_{\text{true},i}$ is larger for Black and

Hispanic high-scorers. It also only directly targets high-scorers from this group as well. Since low-scorers are overly optimistic we should see that increasing funding on its own to everyone will likely lead more low-scorers to enroll and perhaps continue. It may close inequality due to lack of financial resources but is likely to create more mismatch.

Next Figures 13-14 show what education decisions are in the baseline versus what they would be if agent's knew their type with certainty. Hence they show mismatch. The top panel in Figure 13 shows the bachelor's attainment rate of High-scorers in the baseline scenario and under complete information about type by demographic group. The first thing to notice is that there is substantial underinvestment among high-scorers, but this is less the case for White high-SES youth. This shows that there is a substantial fraction of students who would benefit from college that choose not to complete college because they do not know their true type.

The bottom panel in Figure 13 shows the bachelor's attainment rate of low-scorers in the baseline scenario and under complete information about type for the same demographic groups. Conversely there is over investment occurring in the higher education market for low-scorers from all demographic groups. This over investment is highest for White high-SES low-scorers. This shows that there are a significant amount of students who would benefit from foregoing college and working upon completing high school if they knew their college ability type with certainty.

Finally the top panel in Figure 14 shows the aggregate effect of knowing type with certainty. This aggregate effect of having complete information depends on the proportion of high-scorers within the demographic groups considered as shown in the bottom panel of Figure 14. We can see that for Black youth there is little change in bachelor's attainment, and for other youth there is increases in bachelor's attainment. The effect of complete knowledge of type on inequality depends on the proportion of high-scorers in groups[24] and how much

[24]Which likely reflects differences in early childhood human capital investment (Cunha and Heckman 2017)

bachelor's attainment increases for White high-SES youth.

Similarly for the policies discussed the role of differences in financial assistance between groups, differences in beliefs between groups and with respect to $P_{\text{true},i}$, as well as the proportion of high-scorers will be important. The next section will discuss the effects of policies on inequality and mismatch.

5.3 Policy Analysis: Effects on Inequality and Mismatch

In this section I compare the effects of three policies on mismatch and inequality. Inequality is measured by the difference between the reference group, White high-SES youth, and the three comparison groups; Black, Hispanic, and low-SES youth. Mismatch takes the form of under investment in college for high-scorers and over investment for low-scorers as discussed in the last section.

The first of the policies under consideration is a college recruiting policy that provides free college [25] and information to low-SES predicted[26] high-scorers. The last two policies target everyone regardless of SES and predicted scorer type; free college for all and the institution of a tracking policy which provides information about type to all.

In the policy analysis free college is implemented through increasing financial assistance from government and institutions to cover tuition sticker prices. Family financial assistance is kept constant. Information is provided by revealing estimated posteriors from the FMM that only incorporates information that would be available before college, ASVAB scores and behavior indicators.

If inequality was ignored than tracking would be the most efficient policy. The degree to which policies affect inequality depends on the role differences in beliefs and financial

as well as differences in K-12 experience

[25] Takes the form of increasing government and college assistance to cover sticker price of tuition. Family assistance is kept constant.

[26] Predicted high-scorers are simulated high-scorers but using the FMM posterior from only using ASVAB and behavior indicators to mimic information available while in high school.

assistance contribute to gaps in bachelor's attainment as discussed in section 5.1 for high-scorers. The effect on inequality also depends on accuracy of beliefs, as well as the proportion of high types by demographic group. Section 5.1 showed financial assistance is important for all high-scorer comparison groups in explaining gaps, and that beliefs are important for Hispanic and low-SES high-scorers. This suggests targeting will be effective at decreasing inequality and efficient if enough high types are identified.

Section 5.2 showed youth are wrong on average, with much greater inaccuracy for low-scorers. This means that free college for all is likely to generate inefficiencies since overly optimistic beliefs are not addressed. Section 5.2 also showed that low proportions of high-scorers for some demographic groups might actually increase inequality if youth had perfect information. Since tracking provides better but not perfect information it may similarly increase inequality as well.

Figure 15 shows the difference in bachelor's attainment rate between the reference group and each of the three comparison groups. We see that among universal policies free college for all decreases inequality for the groups of interest. Tracking or better information for everyone leads to more inequality. This is suggested by the bottom panel of Figure 14 that shows the proportion of high Scorers is less common among Black, Hispanic, low-SES youth than for White high-SES youth. This is probably due to disparities in early childhood human capital development .

The targeted college recruiting policy decreases inequality the most.The effect of the policy on Black youth and Hispanic youth is less effective than it is on low-SES youth in general. This is because Black and Hispanic youth benefit from the policy only to the extent that they're low-SES. In this case, this means that they are potential first generation college students or come from a family in the the bottom tercile of the household net worth distribution. This suggest that if policy makers want to further decrease racial and ethnic gaps, expanding the policy to target Black and Hispanic predicted high-scorers regardless of

SES could be effective.

In Figure 15 gaps still exists, highlighting that disparities in the early childhood human capital measures that effects the distribution of high-scorers is still an important channel as shown by Cunha and Heckman 2007. There is also room for differences in non-pecuniary utility, that may be due to bias, less community in higher education, or familial obligations generating persistent gaps in higher education. This was suggested by the decomposition exercise shown in Table 8, especially for Black youth.

Finally Table 9, shows us the amount of mismatch present in higher education and how it is distributed among high-scorers and low-scorers. What we see is that for the youth from the NLSY97, 27 % would change their college decisions if they had perfect information. The second column shows that this is primarily amongst high-scorers who would likely increase their schooling if they knew their type. When we enact Free College for all this increases to 31%, with no decrease in under investment, but a larger 3% increase in over investment. As expected the tracking system decreases mismatch the most and almost completely removes it for low-scorers. This is because it brings youth closer to the truth.

The college recruiting policy decreases overall mismatch but only through high-scorers. The percentage of the population that are mismatched high-scorers decreases by 8.1 percentage points. This suggests a significant portion of high-scorers from low-SES backgrounds under invest in college. This is consistent with Hoxby and Avery 2012, and Dynarski, Libassi, Michelmore, and Owen 2017.

Together this shows that there are substantial frictions that exists in the economy leading to mismatch. Table 9 and Figure 15 demonstrate that universal programs exhibit equity vs efficiency trade offs. Targeted policies do not have this trade off. The combination of information and subsidies when targeted to predicted high-scorers from disadvantaged backgrounds can decrease inequality and increase efficiency. It can be just as effective if not more effective than Free College for all at closing gaps. In practice providing subsidies to only a subset of students is likely much less resource intensive than subsidizing college for all youth. Many of these youth might actually already be qualified for free college, so costs may be even smaller (Hoxby and Avery 2012, and Dynarski, Libassi, Michelmore, and Owen 2017).

The analysis in this section showed the college recruiting policy is to be preferred in terms of decreasing inequality while minimizing mismatch. However, there are still gaps in bachelor's attainment. This means disparities will likely still exist as long as there are differences in early childhood human capital development as well as in higher education environments for Black, Hispanic, and low-SES students.

6 Conclusion

In this paper we investigated the role that beliefs played in generating inequality in higher education outcomes for high-scoring youth. In the NLSY97 we found that holding access to resources, demographics, and measures of human capital constant that being more optimistic regarding degree attainment is associated with higher college enrollment, continuation, and completion. We also found that controlling for human capital measures, individual beliefs about enrollment and degree attainment are highly associated with race, ethnicity, parental education, wealth, and percentage of peers with college plans.

In the quantitative analysis I showed that for high-scorers beliefs contribute between 38-49% of the bachelor's attainment gap for Hispanic and low-SES youth, relative to White high-SES high-scorers. Beliefs explain 33% of the gap for Black high-scorers. However a zero belief effect for Black high-scorers can not be ruled out. I find that universal policies like free college for all and tracking exhibit equity efficiency trade offs. Free college for all decreases inequality for Black, Hispanic and low-SES youth but at a cost of creating more over investment among low-scorers. Tracking on the other hand decreases mismatch but leads to more inequality because of differences in proportion of high-scorers. The college recruiting policy on the other hand, combining information with subsidies to low-SES high-scorers decreases mismatch and inequality.

Therefore this paper shows that information frictions do lead to less high scoring youth from all backgrounds under investing in education. These information frictions also contributes to inequality in higher education. Because of that representation in higher education can be increased through more recruiting of academic high achievers from disadvantaged backgrounds and that this is more effective than universal policies. However because of differences in early childhood human capital development and college experience, gaps are likely to still persist. In order to fully close all gaps we must see what effects improving K-12 education, household environment, and inclusion efforts have on higher education outcomes for youth from historically disadvantaged backgrounds.

7 Tables and Figures

7.1 Empirical Analysis: Tables and Figures

Table 1: College Outcomes

VARIABLES	(1) Ever Enrolled	(2) Bachelors Attained	(3) Complete College
Parent Edu	0.0292***	0.0375***	0.0427***
	(0.0048)	(0.0056)	(0.0070)
HH Net Worth ($1000s)	0.0001***	0.0002***	0.0001*
	(0.0000)	(0.0001)	(0.0001)
ASVAB AFQT	0.0055***	0.0057***	0.0035***
	(0.0004)	(0.0004)	(0.0006)
Prob Degree	0.3226***	0.2151***	0.2164***
	(0.0346)	(0.0283)	(0.0491)
Female	0.0806***	0.0847***	0.0411*
	(0.01755)	(0.0186)	(0.0237)
Hispanic	0.0812***	0.0535*	0.0525
	(0.0302)	(0.0286)	(0.0381)
Black	0.1700***	0.1487***	0.1732***
	(0.0266)	(0.0256)	(0.0350)
College GPA			0.1803***
			(0.0152)
Total Govt/Inst Aid ($1000s)			0.0058**
			(0.0027)
Total Fam Aid ($1000s)			0.0075**
			(0.0035)
Total Stud Loan ($1000s)			-0.0081**
			(0.0036)
Geography Controls	Yes	Yes	Yes
Birth Year	Yes	Yes	Yes
Non Cognitive Controls	Yes	Yes	Yes
Peer Effects	Yes	Yes	Yes
Observations	2,133	2,133	1,467
R-squared	0.3499	0.3612	0.3240

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 1: Shows that subjective beliefs about BA attainment reported before age 18 are highly associated with enrollment, continuation, and BA attainment holding parental education, race, ethnicity, access to resources and human capital constant.

Table 2: Measured Beliefs

	(1)	(2)
VARIABLES	Pct Chance Deg by 30	Prob Enroll
Parent Edu	0.0267***	0.0282***
	(0.0046)	(0.0058)
HH Net Worth ($1000s)	0.0001***	0.0001**
	(0.0000)	(0.0000)
ASVAB AFQT	0.0022***	0.0022***
	(0.0004)	(0.0004)
Peers Coll Plan About 25%	0.0812	0.1289*
	(0.0709)	(0.0766)
Peers Coll Plan About 50%	0.1110*	0.1314*
	(0.0671)	(0.0692)
Peers Coll Plan About 75%	0.1662**	0.1562**
	(0.0670)	(0.0695)
Peers Coll Plan more than 90%	0.2117***	0.1954***
	(0.0675)	(0.0691)
Female	0.0767***	0.0117
	(0.0168)	(0.0205)
Hispanic	0.0435	0.1174***
	(0.0268)	(0.0323)
Black	0.0978***	0.1071***
	(0.0246)	(0.0312)
Geography Controls	Yes	Yes
Birth Year	Yes	Yes
Non Cognitive Controls	Yes	Yes
Observations	1,143	1,139
R-squared	0.2614	0.2304

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 2: Shows that subjective beliefs reported before age 18 are highly associated with parental education, race and ethnicity, holding net worth and human capital constant.

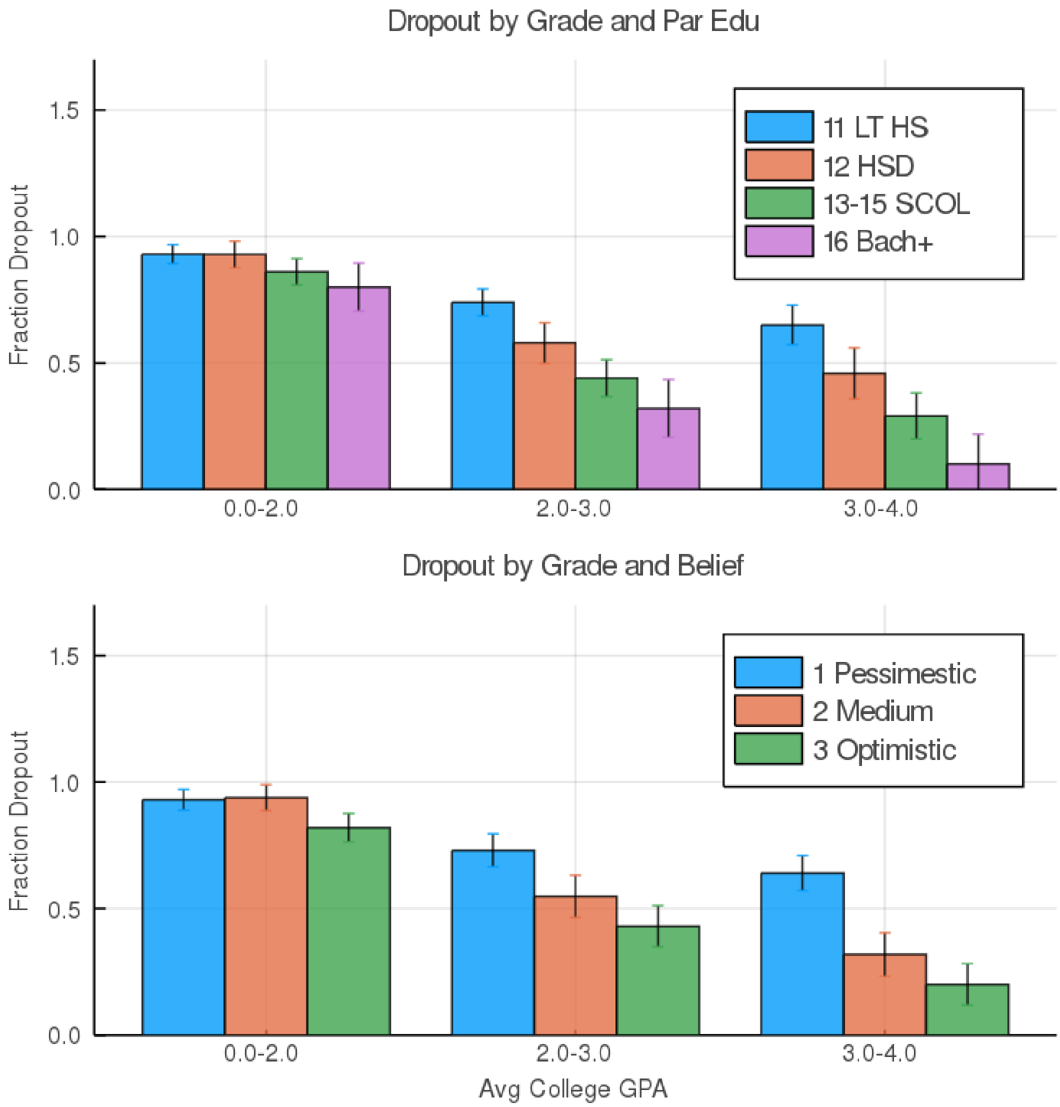

Figure 1: Non Continuation Rates Conditioned on Grades/Demographics.

Table 3: OLS: Non Continuation Interacted with GPA Categories

VARIABLES	Non Interacted	Interaction GPA 2.0-3.0	Interaction GPA > 3.0
Hispanic	-0.1210**	0.1389	-0.0003
	(0.0500)	(0.0897)	(0.0892)
Black	-0.0588	-0.0100	0.0495
	(0.0396)	(0.0671)	(0.0786)
Parent Edu	-0.0003	-0.0077	-0.0317*
	(0.0096)	(0.0161)	(0.0163)
Prob Deg	0.0677	-0.2748***	-0.2007*
	(0.0544)	(0.1032)	(0.1122)
HH Net Worth ($1000s)	-0.0000		
	(0.0001)		
Total Govt/Inst Aid ($1000s)	-0.0174***		
	(0.0043)		
Total Fam Aid ($1000s)	-0.0112		
	(0.0073)		
Total Stud Loan ($1000s)	-0.0065		
	(0.0051)		
ASVAB AFQT	-0.0009		
	(0.0007)		
GPA 2.0-3.0	-0.1130		
	(0.1185)		
GPA > 3.0	-0.2031*		
	(0.1170)		
Geography Controls	Yes		
Birth Year	Yes		
Non Cognitive Controls	Yes		
Observations	1,028		
R-squared	0.2576		

Robust standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

7.2 Model Predictions

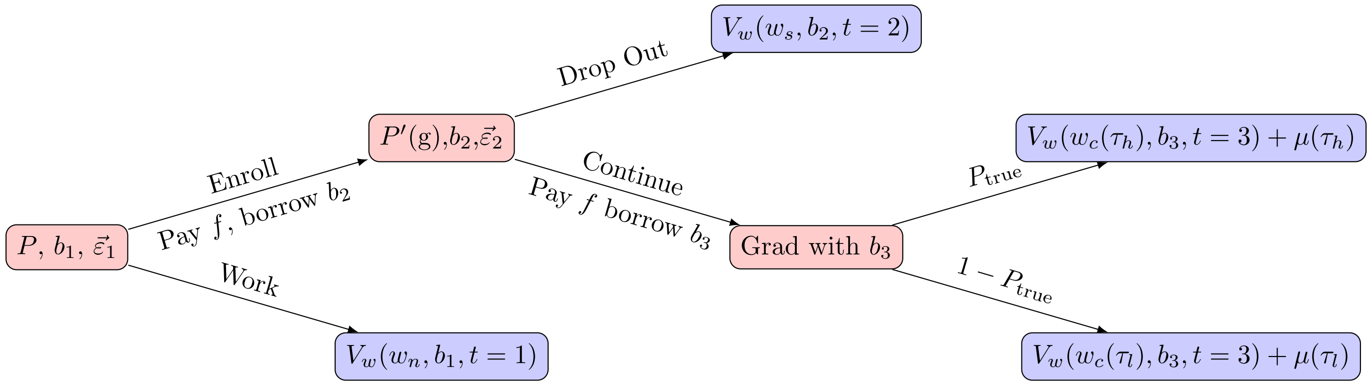

Figure 2: Decision tree representation of the quantitative model. REd nodes represent key stages of the model where decisions are made or different outcomes are realized.

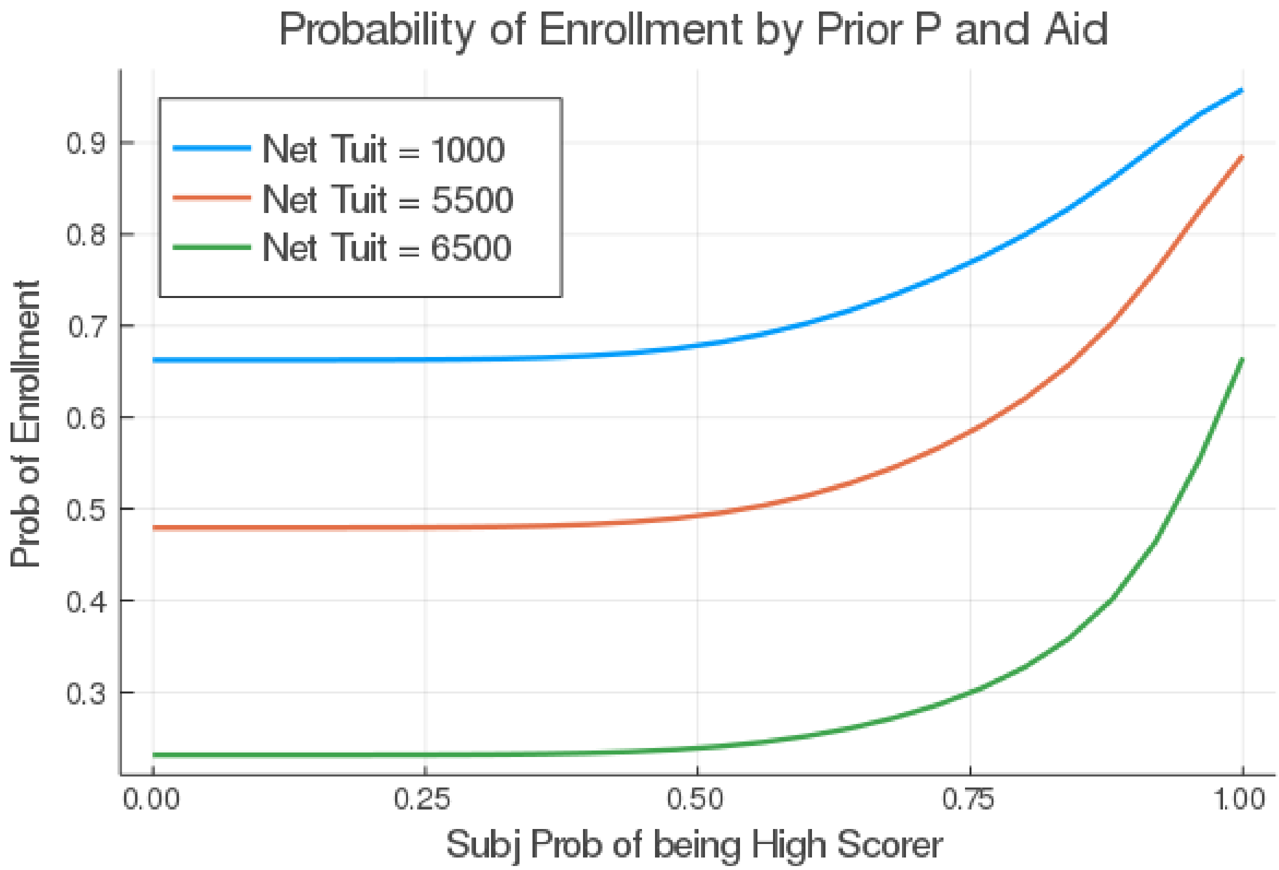

Figure 3: Model predicted probability of college enrollment by Net Tuition and Prior Belief of being high-scorer. Net Tuition is the "sticker price" of tuition minus any financial aid received from family, government or colleges.

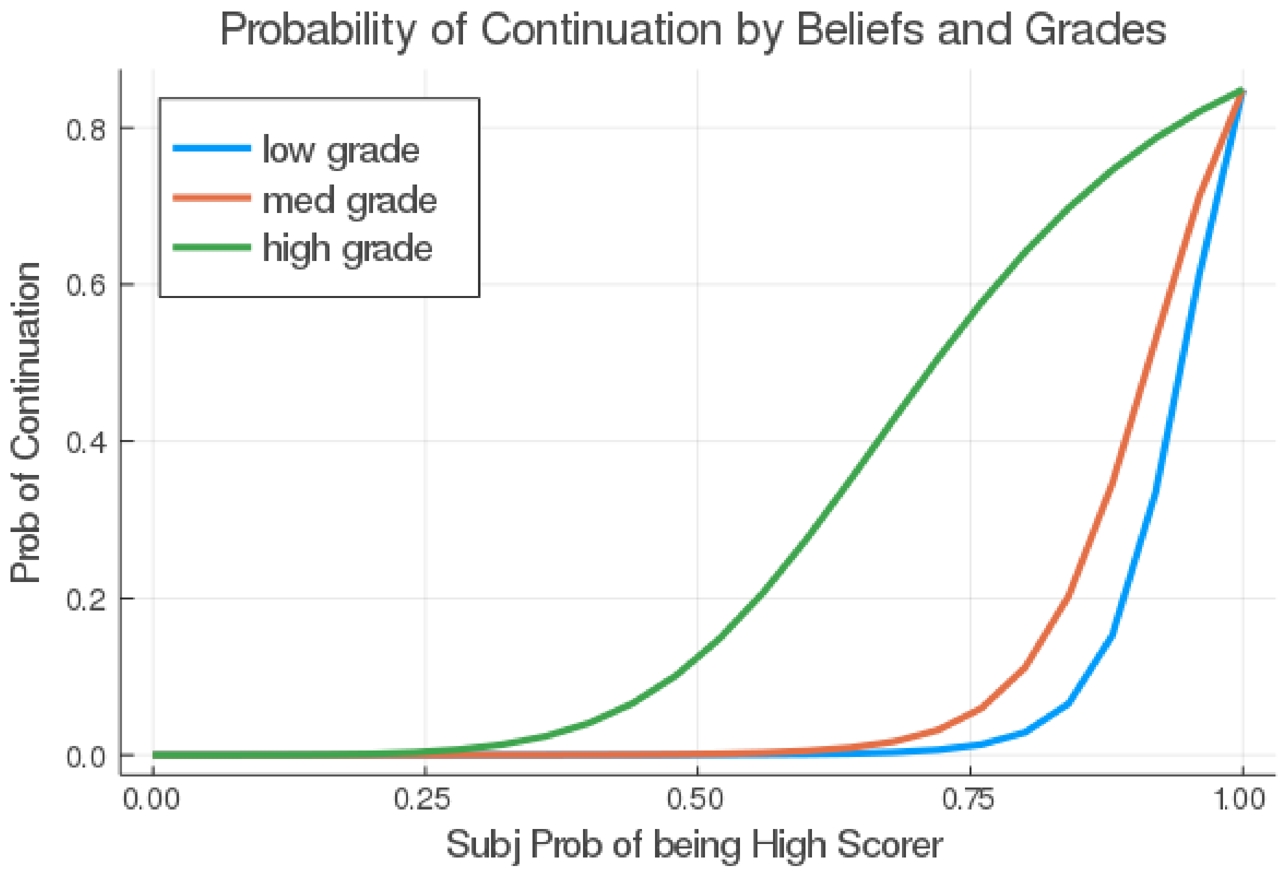

Figure 4: Model predicted probability of College Continuation by average GPA realized before the second stage after the first stage.

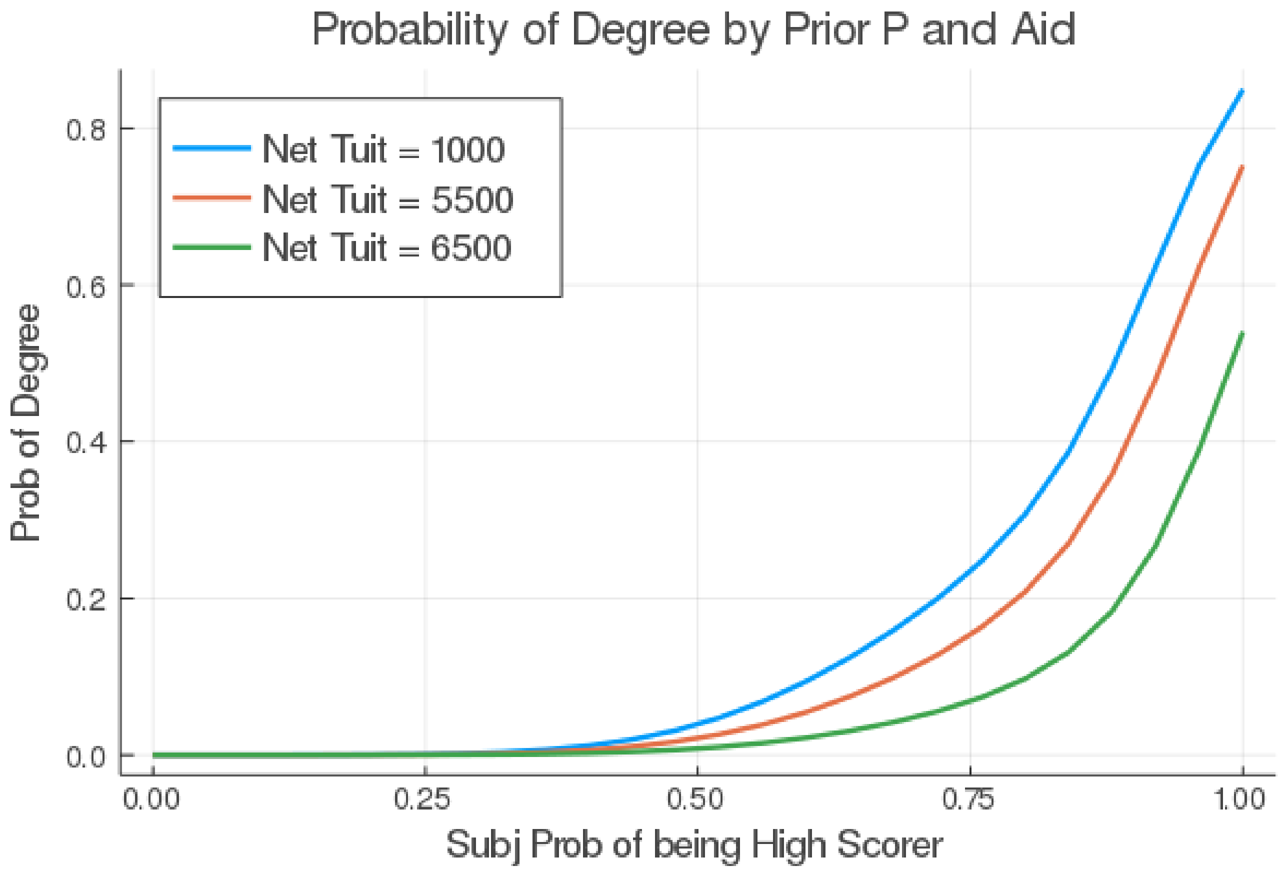

Figure 5: Model predicted probability of Bachelor's attainment, enrollment and completion, by Net Tuition and Prior Belief of being high-scorer. Net Tuition is the "sticker price" of tuition minus any financial aid received from family, government or colleges.

7.3 Quantitative Figures and Tables

Table 4: Preset Parameters prior to Estimation

Parameter	Set Value	Description
β	0.94	Discount rate
γ	2.0	Coeff. of Rel Risk Aversion
$(1+r)$	β^{-1}	Int rate
T	24	Number of periods representing two years
$B_{c,1}$	\$16,600	College Borrowing limits pd 1
$B_{c,2}$	\$35,600	College Borrowing limits pd2

Table 3: Discount rate, coefficient of relative risk aversion, interest rate are set to values similar to other papers. T is intended to capture lifespan from 18-66 or working life since each period lasts two years. College borrowing limits are set to average student loan levels in the first two years and last two years of college.

Table 5: Identification Strategy for Internally Estimated Parameters

Parameters	Target	Description
$\vec{\gamma}_p, \sigma_p, \mu_c(\tau)$	Dropout by grade; Enrollment by belief, Par EDU	Dist of subj belief Non-pecuniary utility by type
$tuit_1, tuit_2$	Enrollment & Dropout by financial aid level	Tuition period 1, and period 2
$\mu_e, \sigma_{c,t}$	Coefficient of parant edu, race, ethnicity on enrollment and completion levels	non-pecuniary util by race, ethnicity, parent edu scale parameters for taste shocks

Table 4: Description of Internally Estimated Parameters and moment targets identifying parameters.

Table 6: External Estimation Results: Average Earnings

Parameter	Estimated Annual Value	Description
w_n	$\$29,584$	Non College Earnings
w_s	$\$45,026$	Some College Earnings
$w_s(\tau_l)$	$\$51,277$	Low type college earnings
$w_s(\tau_h)$	$\$65,841$	High type college earnings

Table 5: Expected value of earnings from Finite Mixture Model by education realization.

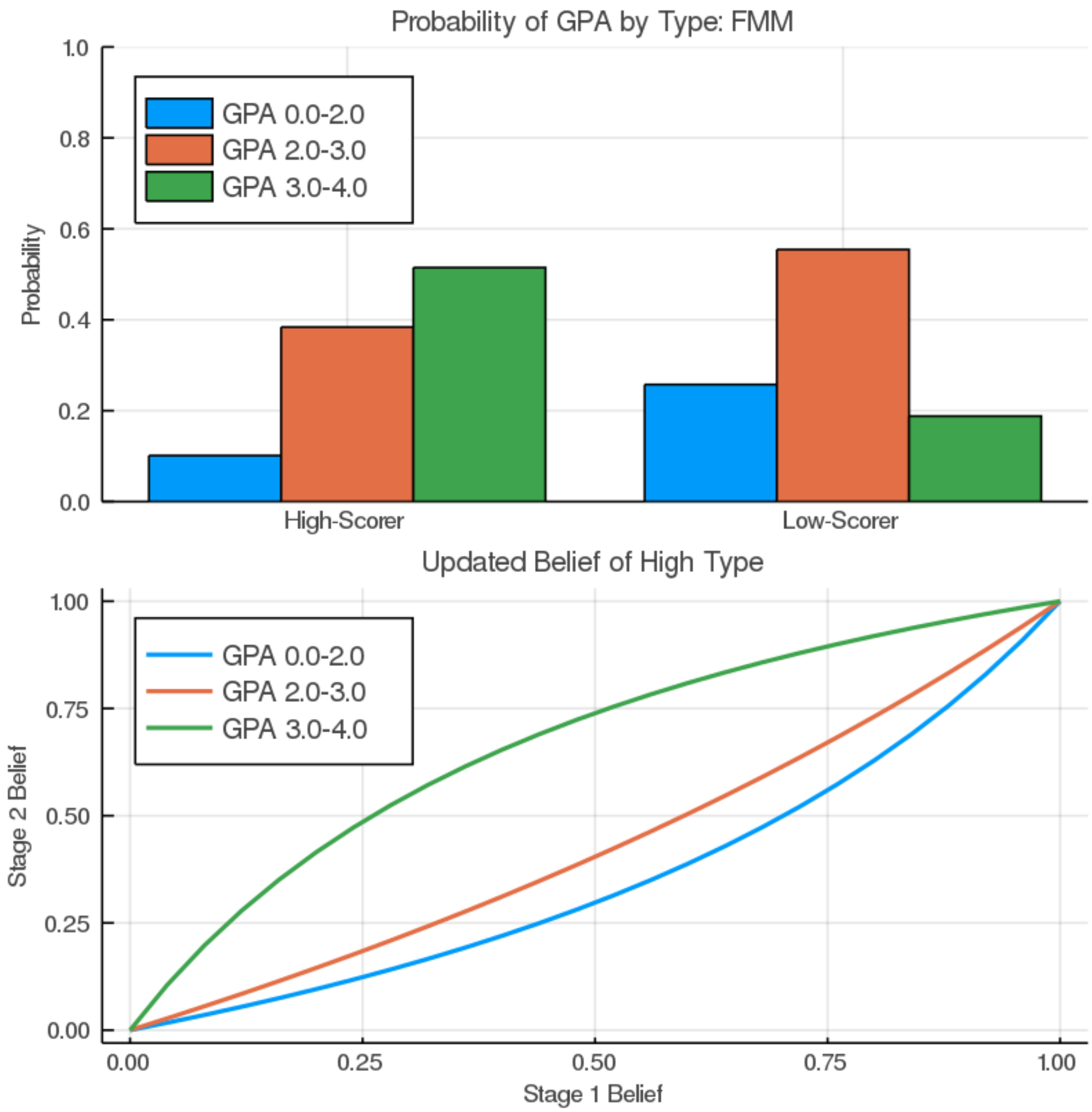

Figure 6: The top panel shows grade probability condition on being a high-scorer or low-scorer. While the bottom panel shows how beliefs are updated to the posterior belief on the y-axis (Stage 2 belief) following grades as a function of the prior given by the x-axis (Stage 1 belief).

Table 7: Key Internal Parameter Results

Parameter	Description	Estimate
$\gamma_{p,0}$	Belief Constant	0.0134
		(0.0127)
$\gamma_{p,b}$	Belief: Meas Belief	0.869
		(0.0092)
$\gamma_{p,h}$	Belief: P-Edu HSD	0.034
		(0.0118)
$\gamma_{p,s}$	Belief: P-Edu SCOL	0.030
		(0.0097)
$\gamma_{p,c}$	Belief: P-Edu Bach	0.059
		(0.0118)
$\mu_{e,0}$	Non Pecun Util: Black 1st Gen Col Stud	-0.000088
		(0.000041)
$\mu_{e,C}$	Non Pecun Util: Col Edu Parents	0.000039
		(0.000032)
$\mu_{e,W}$	Non Pecun Util: White	0.000051
		(0.00003)
$\mu_{e,H}$	Non Pecun Util: Hispanic	0.000014
		(0.00003)
$\mu_c(\tau_h)$	Non Pecun Util high	0.00053
		(0.000066)
$\mu_c(\tau_l)$	Non Pecun Util high	-0.0031
		(0.000278)
$tuit_1$	Tuition Pd 1	\$7430
		(63.36)
$tuit_2$	Tuiton Pd 2	\$6946
		(60.84)

Table 7: Description of key internally estimated parameter results from indirect inference estimation. Standard errors are bootstrapped standard errors from 300 draws with replacement from the data.

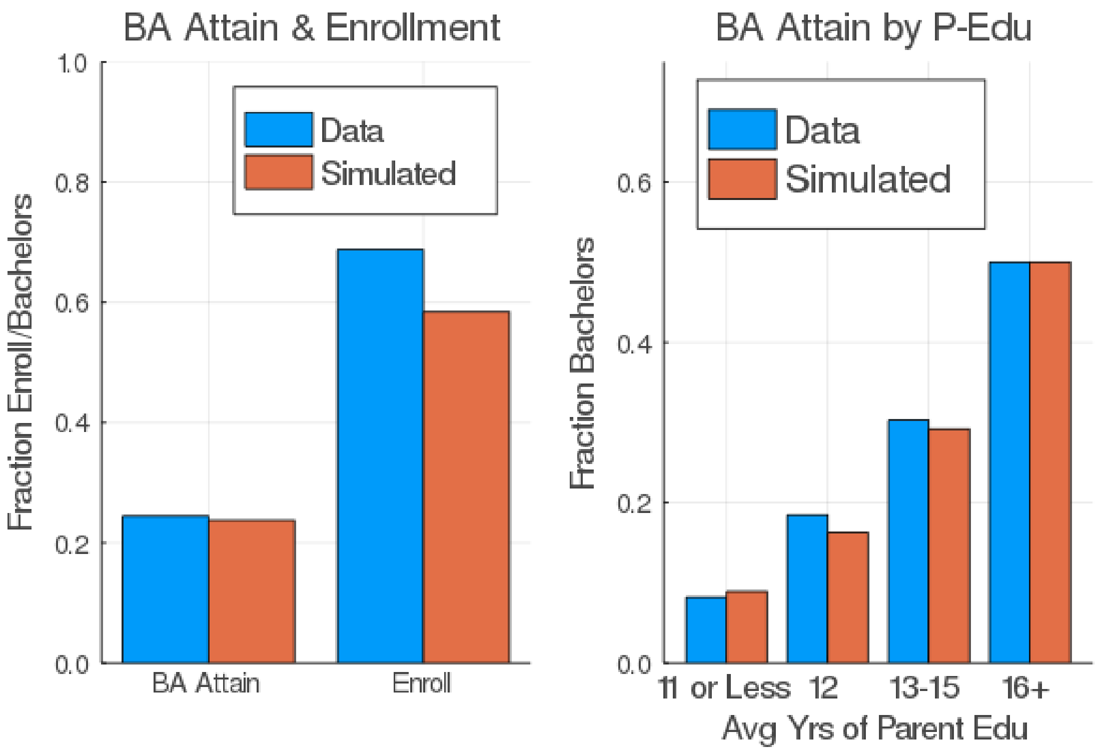

Figure 7: Fit of the Estimated Model: Enrollment, BA attainment, where Blue comes from the NLSY97 and Orange is simulated from the estimated quantitative model.

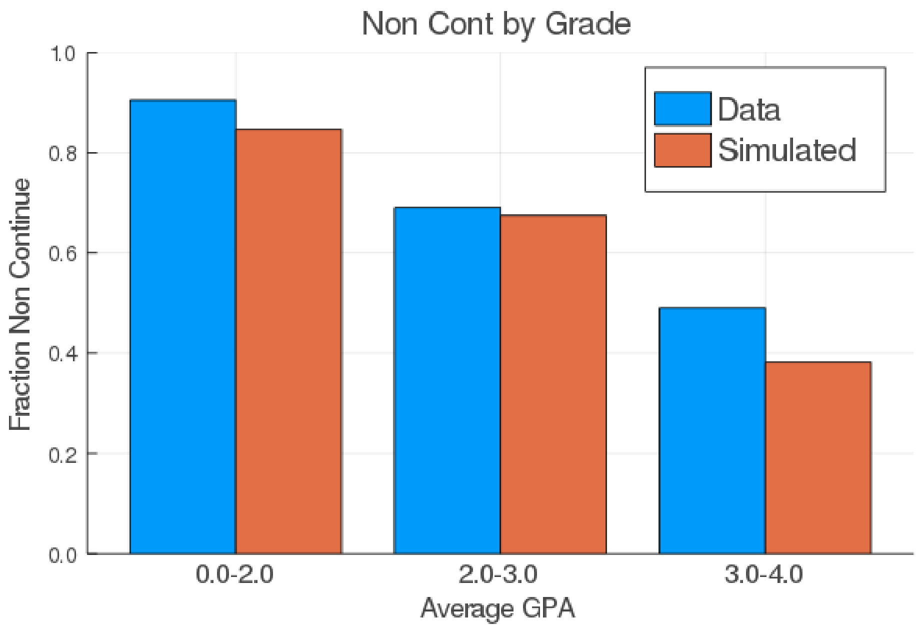

Figure 8: Fit of the Estimated Model: Non Continuation by GPA level, where Blue comes from the NLSY97 and Orange is simulated from the estimated quantitative model.

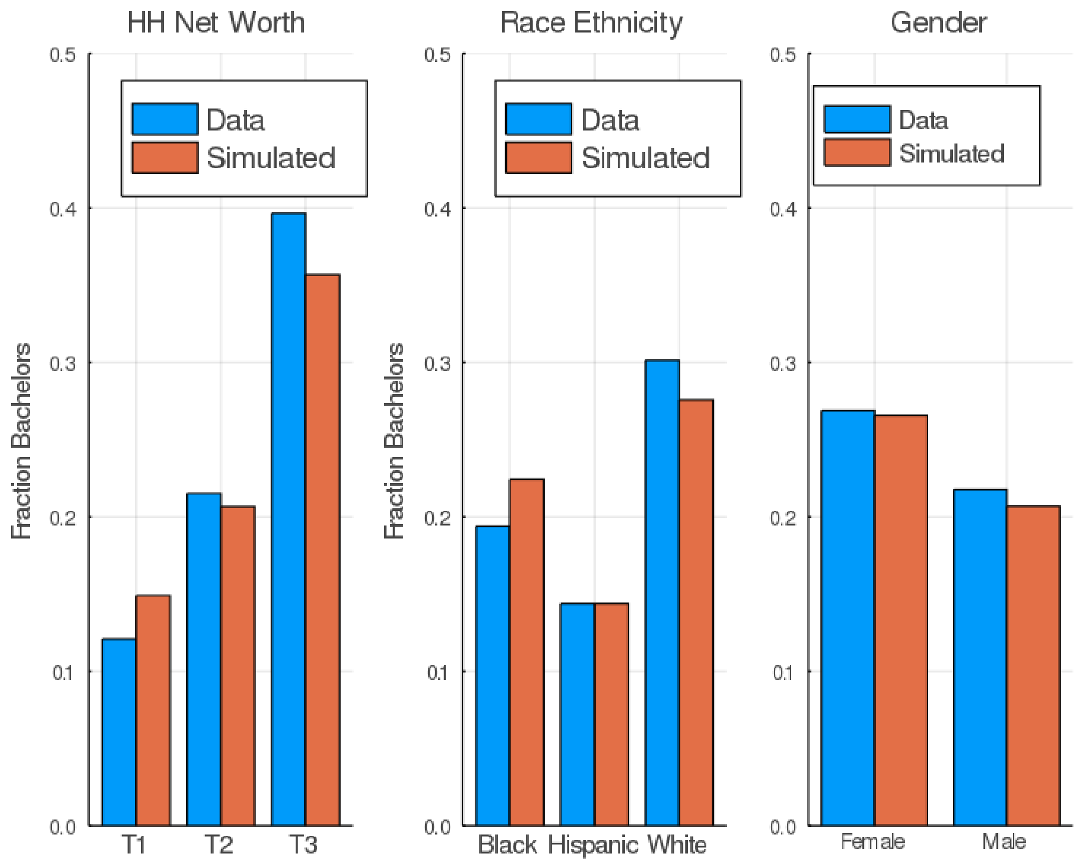

Figure 9: Fit of the Estimated Model: BA attainment by demographics, where Blue comes from the NLSY97 and Orange is simulated from the estimated quantitative model.

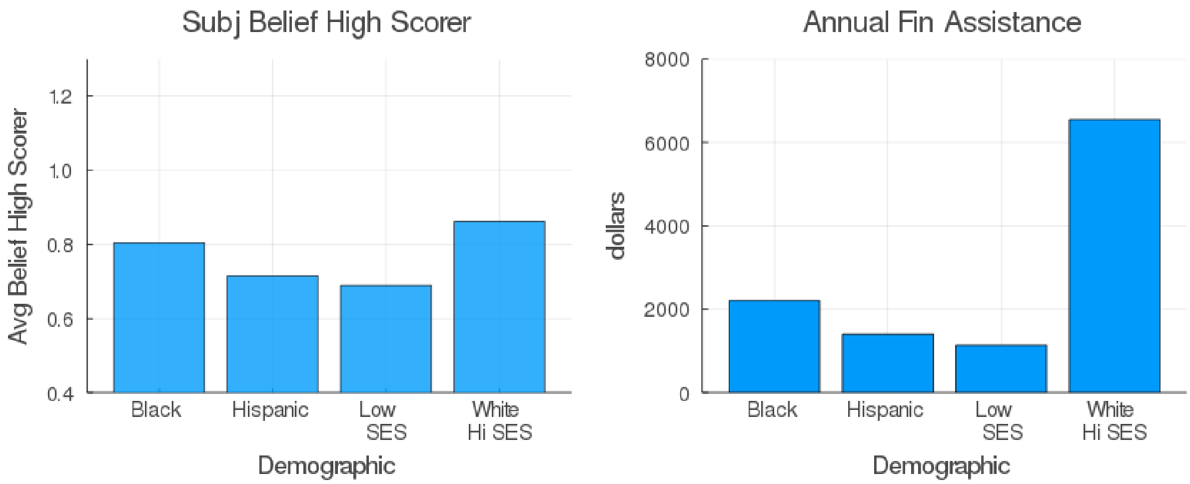

Figure 10: Shows the estimated variables relating to causal mechanism by demographic group, the average subjective belief of being a high-scorer which is a function of measured beliefs in the NLSY97 and parental education, and predicted total financial assistance by demographics which is the sum of family assistance and govt/college aid.

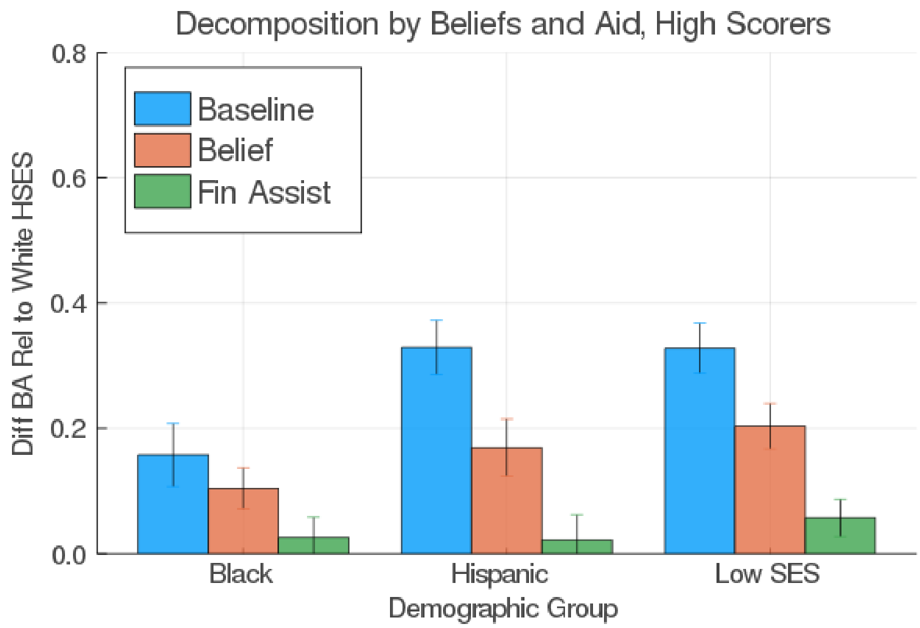

Figure 11: Difference in bachelors attainment relative to White high-SES high-scorers after sequentially equalizing variables. Std errors are bootstrapped std errors.

Table 8: Mechanism Decomposition

Demographic	Total Gap (ppt)	% Beliefs	%Fin Assis	%Other
Black	16	33 %	50%	17%
		(20.4)	(11.22)	
Hispanic	33	49 %	45%	6 %
		(13.67)	(6.34)	
Low SES	33	38%	45%	17 %
		(10.97)	(6.17)	
White High SES Bachelor's attainment	56%			

Table 8: Shows the percentage of the gap relative White high-SES high-scorers explained by each mechanism for each demographic group.

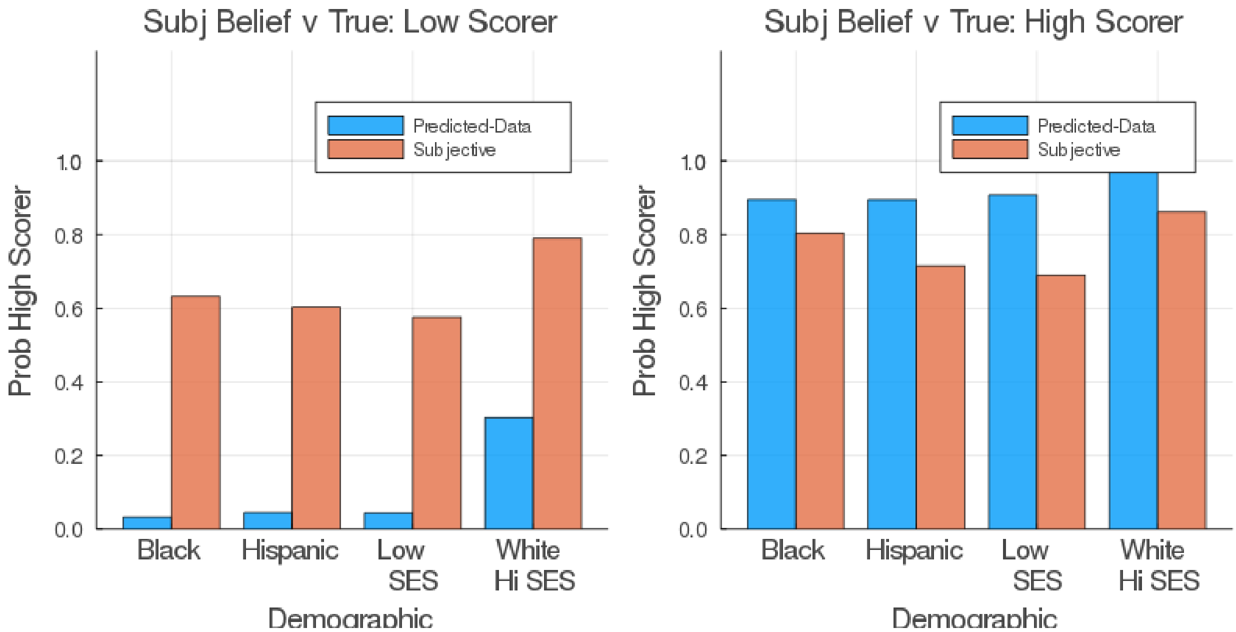

Figure 12: Compares the mean FMM estimate of Prob high-scorer vs the mean subjective belief of being a high-scorer by scorer type. The left hand side is only predicted low-scorers, and the right hand side is only high-scorers.

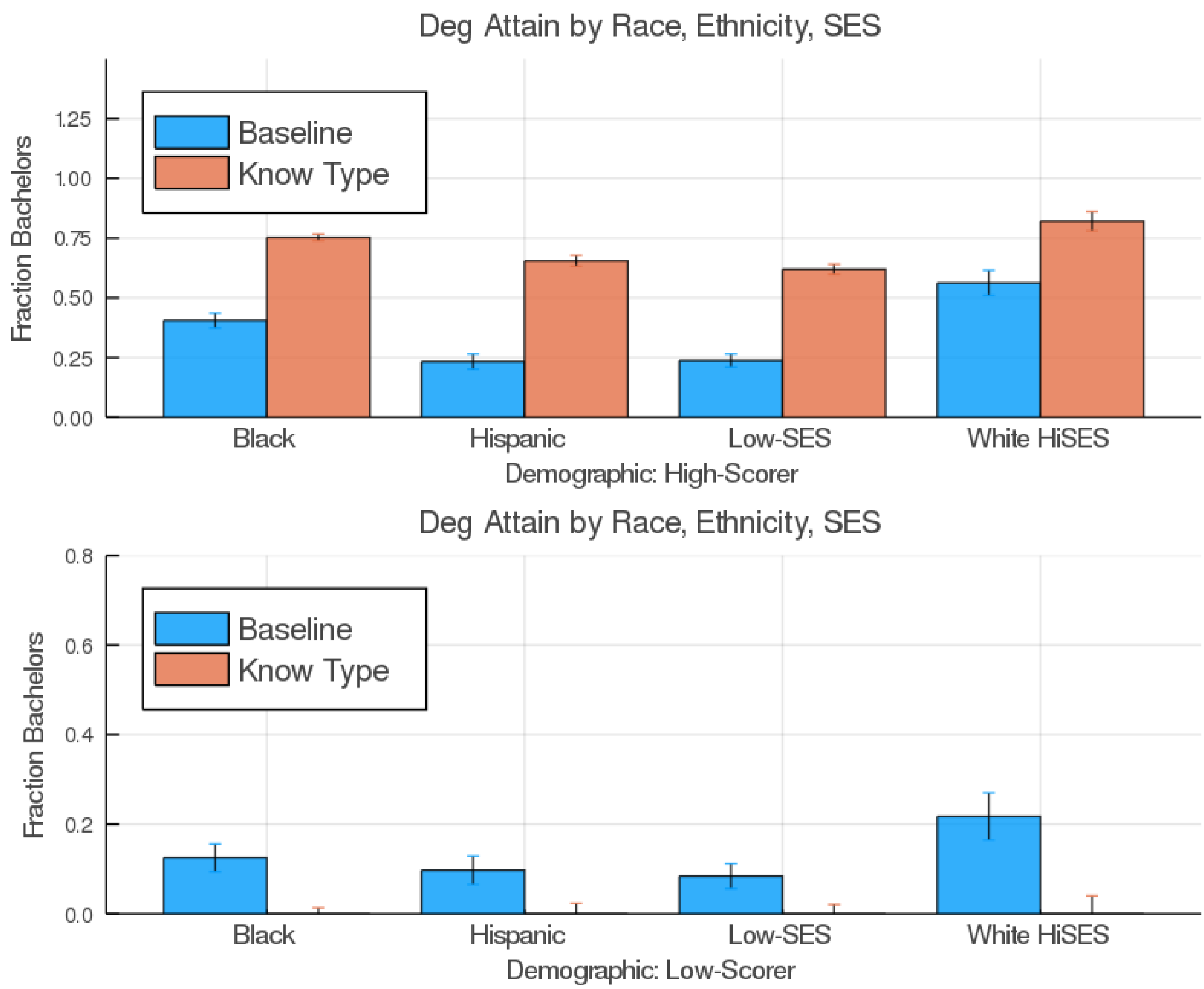

Figure 13: Shows difference in BA attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at predicted high-scorers by demographic group in the top panel, and low-scorers by group in the bottom panel.

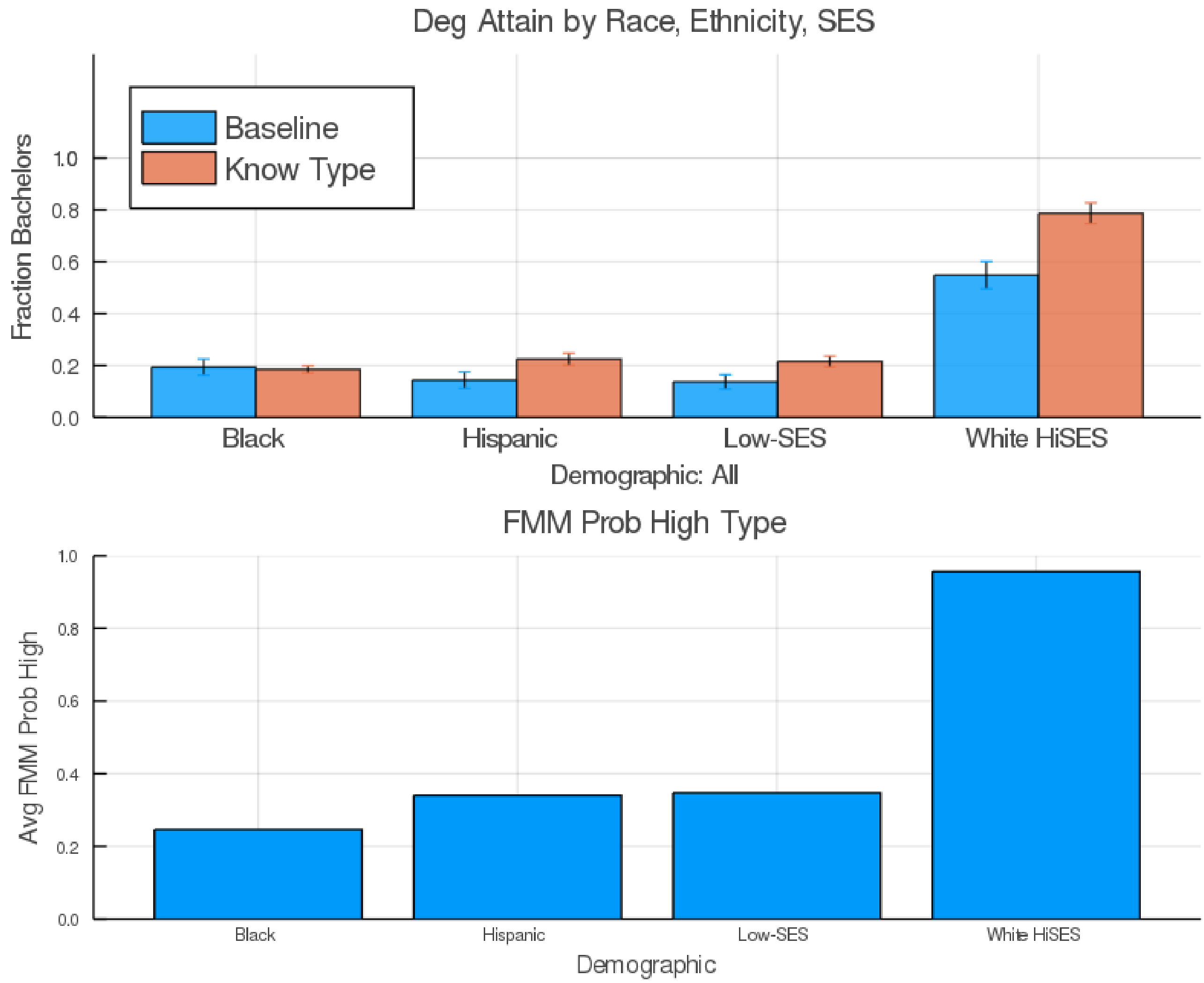

Figure 14: Top panel shows difference in BA attainment regardless of type under baseline model and under scenario where youth know their true type with certainty. The bottom panel shows the proportion of high-scorers by demographic group in order to explain aggregate result.

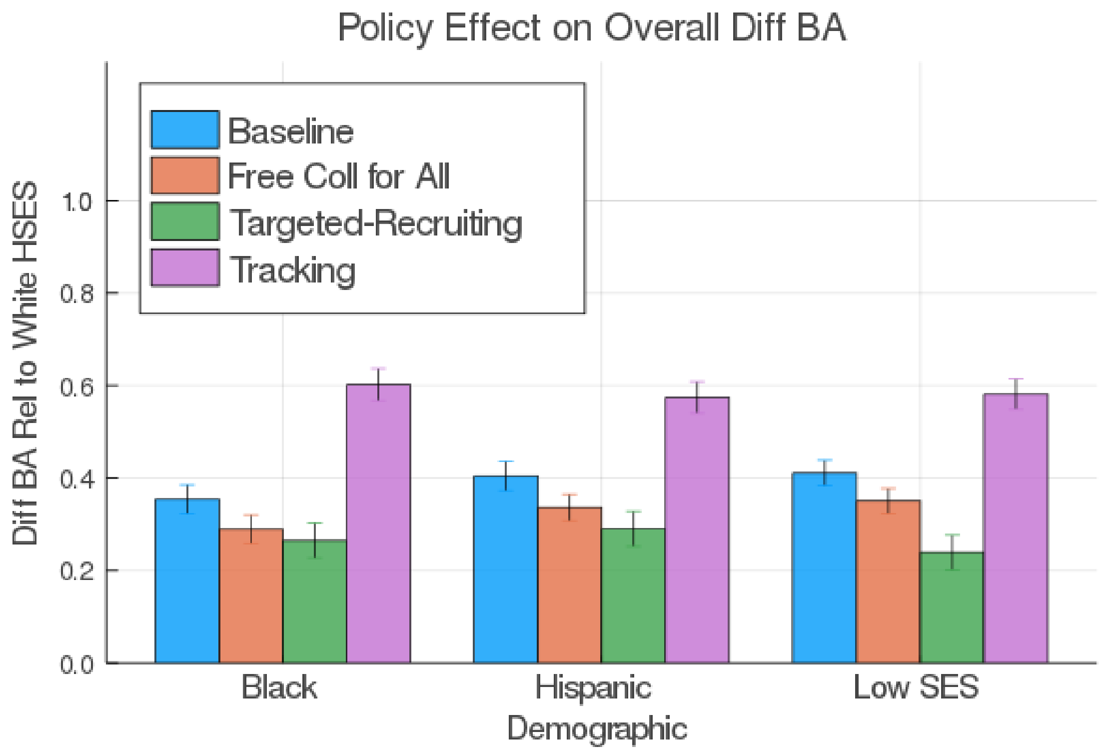

Figure 15: Difference in bachelors attainment relative to White high-SES high-scorers after Policy Enaction. Std errors are bootstrapped std errors.

Table 9: Mismatch: Percentage of Population Switch with Type Knowledge

Policy	% Pop Mismatched Overall	% Pop Mismatched High-Scorer	% Pop Mismatched Low-Scorer
Baseline	27.1 %	21.3 %	5.8 %
Free College For All	30.5%	21.5 %	9.1 %
Tracking: Info for All	4.4 %	4.1 %	0.3 %
Targeted: Recruiting	19.1%	13.3 %	5.9%

Table 9: Shows the percentage of the population in the simulations that would change their mind if they knew their type. Second and third Columns sum to the first since they are percentage of population that are mismatched and high-scorer and percentage of population that are mismatched and Lower Scorers.

8 Bibliography

Abbott B, Gallipoli G, Meghir C, Violante GL. 2016. "Education policy and intergenerational transfers in equilibrium." Journal of Political Economy. Under revision.

Autor, Katz, Kearney. 2008. "Trends in U.S. Wage Inequality: Revising the Revisionists."Review of Economics and Statistics.

Angrist J, Krueger A. 1991."Does Compulsory School Attendance Affect Schooling and Earnings." Quarterly Journal of Economics, 106, 979-1014.

Arcidiacono, Hotz, and Kang. 2012. "Modeling College Major Choices Using Elicited Measures of Expectations and Counterfactuals. Journal of Econometrics.

Arcidiacono, Aucejo, Maurel,Ransom. 2016. "College Attrition and the Dynamics of Information Revelation." Working Paper.

Antman F, Cortes K. 2021. "Long Run Impacts of Mexican American School Desegregation." NBER Working Paper.

Bettinger, Long, Oreopoulos, Sanbonmatsu. 2012. "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block FAFSA Experiment."The Quarterly Journal of Economics. August 2012.

Burdett, Vishwanath. "Declining Reservation Wages and Learning."The Review of Economic Studies. Oct 1988.

Bailey and Dynarski. 2011. "Gains and Gaps: Changing Inequality in US College Entry and Completion." EPI Working Paper.

Boerma and Karabarbounis. 2021. "Reparations and Persistent Racial Wealth Gaps." Working Paper.

Card. 2001. "Estimating the Returns to Schooling: Progress on Some Persistent Econometric Problems." Econometrica.

Card D. 1995 "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," in Aspects of Labour Market Behavior: Essays in Honour of John Vanderkamp,University of Toronto Press, 201-222.

Card D. 1999."The Causal Effect of Education on Earnings," in Handbook of Labor Economics.

Caucutt, Lochner, Park. 2015. "Correlation, Consumption, Confusion, or Constraints: Why Do Poor Children Perform so Poorly?." NBER

Chetty, et al. 2018. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." NBER.

Chetty, et al. 2017 "Mobility Report Cards: The Role of Colleges in Intergenerational Mobility." NBER.

Chetty, et al. 2019 "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." QJE.

Conlon, Pilossoph, Wiswall, Zafar. 2016 "Labor Market Search with Imperfect Information and Learning." AER.

Cunha, Heckman. 2007. "The Technology of Skill Formation." NBER.

Cunha, Heckman, & Navarro. 2005. "Seperating Uncertainty in Life Cycle Earnings." NBER.

Dynarski. 2003. "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion." American Economic Review.

Dynarski. 2011. "Gains and Gaps: Changing Inequality in College Going and Completion." 2011.

Dynarski S, Libassi C, Michelmore K, Owen S. 2020. "Closing the Gap: The Effect of a Targeted, Tuition-Free Promise on College Choices of High Achieving, Low-Income Students." AER.

Fogli, Veldkamp. 2011 "Nature or Nurture: Learning and the Geography of Female Labor Force Participation." Econometrica.

Fryer. 2016 "Information, Non-Financial Incentives, and Student Achievement: Evidence from a Text Messaging Experiment." Journal of Public Economics.

Gonzalez and Shi. 2010. "An Equilibrium Theory of Learning, Search and Wages." Econometrica.

Gittens. 1979. "Bandit Processes and Dynamic Allocation Indices."Journal of the Royal Statistical Society.

Hai and Heckman. 2017. "Inequality in Human Capital and Endogenous Credit Constraints." Review of Economic Dynamics 25 (2017).

Horn, Chen, Chapman. 2003. "Getting Ready to Pay for College: What Students and Their Parents Know about the Cost of College Tuition and What They Are Doing to Find Out." National Center for Education Statistics Report No. 2003030.

Hoxby and Avery. 2004. "Do and Should Financial Aid Packages Affect Student's College Choices." NBER: College Choices: The Economics of Where to Go, When to Go, and How to Pay For It.

Hoxby and Avery. 2013. ”The Missing One Offs: The Hidden Supply of High Achieving Low-Income Students.” Brookings Papers on Economic Activity.

Hoxby and Turner. 2013. ”Expanding College Opportunities for High-Achieving Low Income Students.” Stanford Institute for Economic Policy Research.

Heckman and Kautz. 2014. “Fostering and Measuring Skills: Interventions that Improve Character and Cognition.” The Myth of Achievement Tests : The GED and the Role of Character in American Life. University of Chicago Press.

Hsieh, Hurst, Jones, and Klenow. 2019. ”The Allocation of Talent and US Economic Growth.” Econometrica. 87-5.

Jovanovic. October 1979 “Job Matching and the Theory of Turnover.” JPE.

Jovanovic, Nyarko. 1996. “Learning by Doing and the Choice of Technology.” Econometrica.

Keane Wolpin. 1997. ”The Career Decisions of Young Men.” JPE.

Lise and Postel-Vinay. 2020. ”Multidimensional Skills, Sorting, and Human Capital Accumulation.” AER.

Lochner and Moretti. 2004. “The Effect of Education on Crime: Evidence from Prison Inmates, Arrests, and Self-Reports.” American Economic Review, 94(1),155–189.

Lochner and Monge Naranjo. 2012. “Credit Constraints in Education.” Annual Review of Economics.

Lochner. 2011. ”Nonproduction BenefitsofEducation: Crime, Health,and Good Citizenship.” Handbook of Economics of Education, Vol 4.

Miller. 1984 ”Job Matching and Occupational Choice.” JPE.

National Center for Education Statistics. ”College Enrollment Rates.” May 2020

Navarro and Zhou. 2017. ”Identifying Agent’s Information Sets: An Application to a Lifecycle Model of Schooling, Consumption and Labor Supply. Review of Economic Dynamics.
Papageorgiou, Lopes De-Melo. 2016. ”Occupational Choice, Human Capital and Learning: A Multi-Armed Bandit Approach.” Working Paper.

Reuben E, Wiswall M, Zafar B. 2015. ”Preferences and Biases in Educational Choices and Labour Market Expectations: Shrinking the Black Box of Gender.” The Economic Journal.

Stinebrickner T, Stinebrickner R. 2012. ”Academic Performance and College Dropout:Using Longitudinal Expectations Data to Estimate a Learning Model. NBER

Stinebrickner T, Stinebrickner R. 2014. ”A Major in Science? Initial Beliefs and Final Outcomes for College Major and Dropout.” Review of Economic Studies.

Turner, S. 2004. ”Going to College and Finishing College: Explaining Different Educational Outcomes.” College Choices: The Economics of Where to Go, When to Go, and How to Pay For It. NBER.

Waldfogel. 2015. ”First Degree Price Discrimination Goes to School.” Journal of Industrial Economics.

Wei C, Skomsvold P. 2011. ”Borrowing at the Maximum: Undergraduate Stafford Loan Borrowers in 2007–08.” Stats in Brief. US Department of Education.

Wiswall, Zafar. 2015. ”How do College Students Respond to Public Information about Earnings.” Journal of Human Capital.

Whittle. 1980. ”Multi-Armed Bandits and the Gittins Index.” Journal of the Royal Statistical Society.

A Appendix

A.1 More Empirical Facts

Table 10: Summary Statistics by Race Ethnicity

VARIABLES	(1) All	(2) White	(3) Hispanic	(4) Black
Enrolled in College	0.717	0.740	0.626	0.670
Bachelors or More	0.301	0.336	0.171	0.222
Parent Edu Lt 12	0.220	0.158	0.541	0.288
Parent Edu 12	0.216	0.202	0.176	0.313
Parent Edu 13-15	0.388	0.434	0.200	0.302
Parent Edu 16+	0.176	0.205	0.083	0.098
Avg Parent Edu	13.02	13.43	11.15	12.37
HH Net Worth ($1000s)	185.8	226.4	80.68	56.04
Pct Peers ColPlan	66.5	68.7	60.8	68.5
Prob Enroll	0.751	0.758	0.734	0.732
Prob Degree	0.777	0.793	0.679	0.767
College GPA	2.65	2.79	2.41	2.14
Total Govt/Inst Aid ($1000s)	2.3	1.96	1.65	2.71
Total Fam Aid ($1000s)	1.64	1.92	0.96	0.60
ASVAB AFQT	54.73	61.20	40.32	32.15
Ever Stole	0.0671	0.0608	0.0943	0.0779
Ever Violence	0.161	0.141	0.165	0.265
Ever Sex before 15	0.182	0.145	0.186	0.375
Sample Size	2133	1188	404	541

Table 11: Summary Statistics by Parent Education

VARIABLES	(1) All	(2) Lt 12	(3) 12	(4) 13-15	(5) 16 +
Enrolled in College	0.717	0.447	0.614	0.814	0.944
Bachelors or More	0.301	0.0787	0.208	0.359	0.544
Hispanic	0.116	0.285	0.092	0.062	0.056
Black	0.146	0.191	0.212	0.114	0.082
Avg Parent Edu	13.02	10.10	12.00	13.77	16.00
HH Net Worth ($1000s)	185.8	53.53	123.8	201.7	375.8
Pct Peers ColPlan	66.5	58.2	62.3	69.7	75.2
Prob Enroll	0.751	0.572	0.713	0.812	0.882
Prob Degree	0.777	0.633	0.691	0.840	0.917
College GPA	2.65	2.21	2.62	2.68	2.98
Total Govt/Inst Aid ($1000s)	2.3	2.40	1.68	1.93	2.29
Total Fam Aid ($1000s)	1.64	0.42	0.85	1.64	3.01
ASVAB AFQT	54.73	32.47	49.53	60.13	75.08
Ever Stole	0.0671	0.0928	0.0492	0.0750	0.0422
Ever Violence	0.161	0.233	0.176	0.147	0.0903
Ever_Sex before 15	0.182	0.295	0.210	0.152	0.0845
Sample Size	2133	586	493	736	318

Table 12: Financial Assistance

VARIABLES	(1) Any Family Aid	(2) Total Fam Aid	(3) Any Govt/Inst Aid	(4) Total Govt/Inst Aid
Parent Edu	0.0346***	0.1854***	-0.0006	-0.0793
	(0.0072)	(0.0607)	(0.0078)	(0.0751)
HH Net Worth	0.0003***	0.0050***	-0.0002***	0.0001
	(0.0001)	(0.0009)	(0.0001)	(0.0007)
ASVAB AFQT	0.0030***	0.0114**	0.0022***	0.0216***
	(0.0006)	(0.0045)	(0.0006)	(0.0067)
Female	0.0322	-0.0604	0.0574**	0.2054
	(0.0249)	(0.2464)	(0.0276)	(0.3452)
Hispanic	0.0198	0.5455*	0.0995**	-0.5875
	(0.0403)	(0.3057)	(0.0441)	(0.5116)
Black	-0.0134	0.0212	0.1932***	0.9796**
	(0.0393)	(0.2425)	(0.0386)	(0.4450)
Geography Controls	Yes	Yes	Yes	Yes
Birth Year	Yes	Yes	Yes	Yes
Non Cognitive Controls	Yes	Yes	Yes	Yes
Robust Standard Errors	Yes	Yes	Yes	Yes
Peer Effects	Yes	Yes	Yes	Yes
Observations	1,467	929	1,467	940
R-squared	0.1478	0.2416	0.0503	0.0379

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Oaxaca-Blinder Decomp: Subj Prob Degree: White vs Hispanic/Black

VARIABLES	White Hisp overall	White Hisp explained	White Hisp unexplained	White Black overall	White Black explained	White Black unexplained
Parent Edu		0.0604***	0.0822**		0.0317***	0.0582
		(0.0105)	(0.0326)		(0.0056)	(0.0473)
HH Net Worth (1000$s)		0.0139***	0.0092		0.0158***	-0.0032
		(0.0043)	(0.0084)		(0.0050)	(0.0079)
ASVAB AFQT		0.0537***	-0.0218		0.0682***	-0.0317
		(0.0083)	(0.0306)		(0.0094)	(0.0276)
Female		-0.0002	-0.0161		-0.0033**	0.0070
		(0.0014)	(0.0176)		(0.0016)	(0.0202)
% Peers College Plan		0.0128***	-0.0505		0.0127***	-0.0168
		(0.0036)	(0.0600)		(0.0035)	(0.0581)
Ever Stole more $50		0.0002	0.0007		0.0000	-0.0060
		(0.0005)	(0.0053)		(0.0002)	(0.0044)
Ever Violence		0.0014	-0.0090		0.0038*	-0.0106
		(0.0013)	(0.0086)		(0.0021)	(0.0088)
Ever Sex bf15		0.0051**	-0.0194*		0.0191***	-0.0113
		(0.0024)	(0.0103)		(0.0051)	(0.0123)
Ref Mean (White)	0.7659***			0.7659***		
	(0.0093)			(0.0093)		
Comp Mean	0.7053***			0.7375***		
	(0.0162)			(0.0154)		
difference	0.0606***			0.0285		
	(0.0187)			(0.0180)		
explained	0.1470***			0.1477***		
	(0.0124)			(0.0115)		
unexplained	-0.0864***			-0.1192***		
	(0.0208)			(0.0194)		
Constant			34.6180			5.9500
			(23.3366)			(21.4124)
Observations	1,592	1,592	1,592	1,716	1,716	1,716
N Comparison	404	404	404	528	528	528
N Reference (White)	1188	1188	1188	1188	1188	1188

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 13: Shows Oaxaca Blinder decomposition results explaining unconditional differences in self reported probability of degree mean of Hispanic/Black youth compared to White youth by differences in explanatory variables(explained column) and coefficient results in race/ethnicity seperate regressions.

Table 14: Oaxaca-Blinder Decomp: Enroll: White vs Hispanic/Black

VARIABLES	White Hisp overall	White Hisp explained	White Hisp unexplained	White Black overall	White Black explained	White Black unexplained
Parent Edu		0.0674***	0.0634		0.0333***	0.0559
		(0.0139)	(0.0448)		(0.0069)	(0.0588)
HH Net Worth ($1000s)		0.0152***	-0.0030		0.0163**	0.0021
		(0.0055)	(0.0133)		(0.0063)	(0.0134)
ASVAB AFQT		0.1317***	-0.0324		0.1740***	-0.1048***
		(0.0132)	(0.0427)		(0.0142)	(0.0354)
Belief Var		0.0198***	-0.0254		0.0081	0.0591
		(0.0065)	(0.0627)		(0.0052)	(0.0532)
Female		-0.0003	0.0191		-0.0085***	-0.0506**
		(0.0017)	(0.0244)		(0.0030)	(0.0242)
% Peers College Plan		0.0052	0.0246		-0.0005	0.1512**
		(0.0035)	(0.0820)		(0.0035)	(0.0699)
Ever Stole more $50		0.0002	-0.0033		-0.0000	-0.0052
		(0.0005)	(0.0071)		(0.0001)	(0.0050)
Ever Violence		0.0011	-0.0189		0.0055**	-0.0037
		(0.0012)	(0.0116)		(0.0028)	(0.0111)
Ever Sex bf15		0.0029	-0.0061		0.0106*	-0.0107
		(0.0021)	(0.0132)		(0.0055)	(0.0143)
Ref Mean (White)	0.7239***			0.7239***		
	(0.0130)			(0.0130)		
Comp Mean	0.5743***			0.6534***		
	(0.0246)			(0.0207)		
difference	0.1496***			0.0705***		
	(0.0278)			(0.0244)		
explained	0.2432***			0.2388***		
	(0.0190)			(0.0179)		
unexplained	-0.0936***			-0.1683***		
	(0.0269)			(0.0240)		
Constant			3.9612			19.0688
			(31.3443)			(25.9906)
Observations	1,592	1,592	1,592	1,716	1,716	1,716
N Comparison	404	404	404	528	528	528
N Reference (White)	1188	1188	1188	1188	1188	1188

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 14: Shows Oaxaca Blinder decomposition results explaining unconditional differences in college enrollment mean of Hispanic/Black youth compared to White youth by differences in explanatory variables(explained column) and coefficient results in race/ethnicity seperate regressions.

Table 15: Oaxaca-Blinder Decomp: College Cont: White vs Hispanic/Black

VARIABLES	White Hisp overall	White Hisp explained	White Hisp unexplained	White Black overall	White Black explained	White Black unexplained
Parent Edu		0.0905***	0.1009		0.0514***	0.0910
		(0.0178)	(0.0655)		(0.0102)	(0.0828)
HH Net Worth (1000$s)		0.0167*	-0.0197		0.0189*	-0.0249**
		(0.0091)	(0.0255)		(0.0105)	(0.0122)
ASVAB AFQT		0.0675***	0.0653		0.1078***	-0.0541
		(0.0146)	(0.0695)		(0.0175)	(0.0612)
Belief Var		0.0139**	-0.0010		0.0082*	0.0484
		(0.0057)	(0.1128)		(0.0044)	(0.0809)
College Avg GPA		0.0602***	-0.0533		0.1141***	-0.0520
		(0.0118)	(0.0934)		(0.0132)	(0.0865)
Total Govt/Inst Aid		0.0013	0.0155		-0.0052*	0.0065
		(0.0013)	(0.0214)		(0.0029)	(0.0159)
Total Fam Aid		0.0086**	-0.0143		0.0115**	-0.0258**
		(0.0041)	(0.0152)		(0.0051)	(0.0115)
College Stud Loan		-0.0035	-0.0004		-0.0001	-0.0175
		(0.0022)	(0.0126)		(0.0009)	(0.0179)
Female		0.0002	0.0087		-0.0034	0.0261
		(0.0011)	(0.0330)		(0.0031)	(0.0351)
% Peers College Plan		0.0039	0.0725		0.0049	0.0541
		(0.0048)	(0.1214)		(0.0060)	(0.0924)
Ever Stole more $50		0.0003	0.0007		0.0015	0.0086
		(0.0013)	(0.0085)		(0.0016)	(0.0068)
Ever Violence		0.0008	-0.0019		0.0033	0.0064
		(0.0028)	(0.0140)		(0.0037)	(0.0111)
Ever Sex bf15		0.0090**	-0.0453***		0.0246***	-0.0397***
		(0.0045)	(0.0147)		(0.0080)	(0.0150)
Ref Mean (White)	0.5790***			0.5790***		
	(0.0168)			(0.0168)		
Comp Mean	0.3586***			0.4124***		
	(0.0312)			(0.0262)		
difference	0.2204***			0.1666***		
	(0.0354)			(0.0311)		
explained	0.2695***			0.3373***		
	(0.0250)			(0.0239)		
unexplained	-0.0491			-0.1708***		
	(0.0356)			(0.0322)		
Constant			31.0493			-21.1310
			(41.3287)			(34.1851)
Observations	1,104	1,104	1,104	1,221	1,221	1,221
N Comparison	237	237	237	354	354	354
N Reference (White)	867	867	867	867	867	867

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 15: Shows Oaxaca Blinder decomposition results explaining unconditional differences in college continuation mean of Hispanic/Black youth compared to White youth by differences in explanatory variables(explained column) and coefficient results in race/ethnicity seperate regressions. This specification is conditional on enrollment

Table 16: Average Log Earnings

VARIABLES	(1) HS or Less	(2) Some Coll	(3) Bach Deg or More	(4) Returns SCol	(5) Returns Bach
Parent Edu	0.0133	-0.0010	-0.0271*	-0.0143	-0.0404
	(0.0196)	(0.0155)	(0.0136)	(0.0281)	(0.0268)
HH Net Worth	0.0010***	0.0002	0.0003**	-0.0008**	-0.0007**
	(0.0003)	(0.0002)	(0.0001)	(0.0003)	(0.0003)
Prob Deg	0.2397**	0.2016*	0.1355	-0.0380	-0.1042
	(0.1022)	(0.1058)	(0.1085)	(0.1561)	(0.1703)
ASVAB AFQT	0.0048**	0.0007	0.0059***	-0.0041*	0.0011
	(0.0018)	(0.0011)	(0.0013)	(0.0022)	(0.0024)
Female	-0.7265***	-0.4011***	-0.3544***	0.3254***	0.3722***
	(0.0751)	(0.0656)	(0.0558)	(0.0996)	(0.0935)
Hispanic	-0.0803	0.2513***	0.0649	0.3316***	0.1452
	(0.0954)	(0.0800)	(0.0938)	(0.1244)	(0.1338)
Black	-0.4046***	-0.2088**	0.1860*	0.1959	0.5907***
	(0.0995)	(0.0844)	(0.1019)	(0.1303)	(0.1424)
Constant	9.9542***	10.2503***	10.7313**	0.2961	0.7771*
	(0.2779)	(0.3658)	(0.2925)	(0.4697)	(0.4246)
Observations	666	696	771	2,133	2,133
R-squared	0.2594	0.1254	0.1258	0.2738	0.2738

Robust standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 16: Shows the OLS results from regressing log average earnings (where 1 is added to include zeros) on education seperately (first three columns) then interacted with all variables (Last two columns).

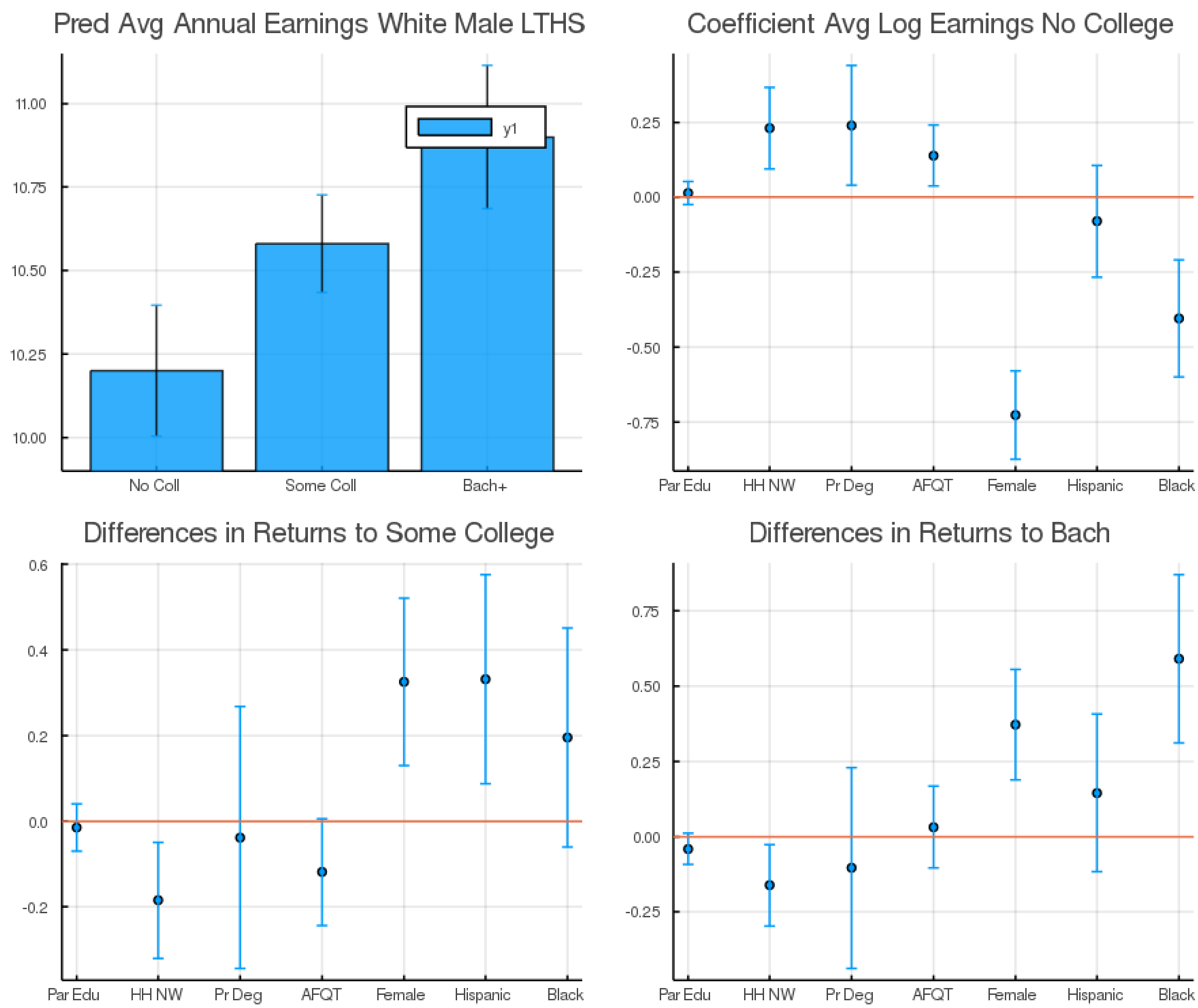

Figure 16: Simulated Update Rule

A.2 Likelihood Function: Finite Mixture Model

In this section, I briefly go over the likelihood function used to estimate the finite mixture model. The finite mixture model, uses the four continuous ASVAB test scores (Arithmetic Reasoning, Mathematical Knowledge, Paragraph Comprehension, and Word Knowledge), the three discrete adverse behavior measures (Sex before age 15, ever committed violence at start of survey, and ever stole greater than \$50 at start of the survey), discrete college GPA categories (0.0-2.0, 2.0-3.0, 3.0-4.0), and earnings as measurement equations. These measurement equations are functions of the latent type τ for scorer type. The finite mixture model also controls for demographic selection in enrollment and college continuation. The probability that $\tau = \tau_h$ is also allowed to differ by demographic group.

Equation (a.1) shows the full likelihood function. The first line is the product of the likelihood contribution of all four of the cognitive ability measures, the ASVAB test scores, which are observed for the whole sample. In the likelihood function $\phi(\cdot)$ is the pdf for the standard normal distribution, where the first argument is normalized subtracting its mean and dividing by the standard deviation.

The second line of (a.1) is the product of the likelihood contribution of observing the three discrete non cognitive ability measures. $\Phi(\cdot)$ is the CDF of the standard normal distribution where Z^*_{i,j_n} is normalized by subtracting its mean and dividing by the standard deviation.

The third line of (a.1) is the probability that an individual has less than or equal to 12 years of schooling, mutiplied by the pdf of observing log earnings $\ln w_{i,s}$, where log earnings are assumed to be normally distributed. The fourth and fifth lines are similar to line three of (a.1) in that we multiply the probabilty of observing schooling type, by the likelihood of earnings given schooling type. Lines three and four differ in that we also multiply by the likelihood of observing GPA $g = g_k$, since this information is only seen if agents enroll.

Notice type τ_k enters earnings for college graduates, grade probabilities, and cognitive ability measurements. Demographic information X_i enters probability of being high type, as well as probability of enrollment then non completion Prob($s \in (12,16)$) and probability of having a bachelor's degree Prob($s \geq 16$).

$$(a.1) \qquad f(\vec{Z}_i, w_i, g_i; \tau_k, X_i, s) = \Pi_{j_c}\phi(Z^*_{i,j_c};\tau_k)\times$$

$$\Pi_{j_n}\Phi(Z^*_{i,j_n};\tau_k)^{1(Z^*_{i,jn})}\times(1-\Phi(Z^*_{i,j_n};\tau_k))^{1-1(Z^*_{i,jn})}$$

$$\times[\text{Prob}(s\leq 12|X_i))\phi(\ln w_{i,s})]^{1(s<12)}$$

$$\times[\text{Prob}(s\in(12,16)|X_i)\Pi_{g_k}\pi(g_k|\tau_k)^{1(g=g_k)}\phi(\ln w_{i,s})]^{1(s\in(12,16))}$$

$$\times[\text{Prob}(s\geq 16)|X_i)\Pi_{g_k}\pi(g_k|\tau_k)^{1(g=g_k)}\phi(\ln w_{i,s};\tau_k)]^{1(s\geq 16)}$$

A.3 Finite Mixture Model Results

Table 17: Prob by Demographic: FMM

VARIABLES	Logit Prob High Type	Logit Prob Enroll	Logit Prob Continue
Intercept	-1.029	-0.991	-3.367
	(0.306)	(0.163)	(0.333)
Parent HS	0.930	0.610	0.460
	(0.286)	(0.132)	(0.212)
Parent Some Coll	1.296	1.407	0.756
	(0.341)	(0.151)	(0.204)
Parent Bach	2.635	2.58	1.159
	(0.663)	(0.272)	(0.217)
HH Net Worth Tercile 2	0.358	0.396	0.337
	(0.185)	(0.129)	(0.172)
HH Net Worth Tercile 3	1.044	1.063	0.637
	(0.348)	(0.169)	(0.185)
Black	-0.655	0.307	-0.040
	(0.201)	(0.145)	(0.189)
Hispanic	-1.488	0.441	0.354
	(0.467)	(0.139)	(0.164)
Female	0.224	0.629	0.043
	(0.249)	(0.105)	(0.119)
GPA Med			2.167
			(0.240)
GPA High			1.475
			(0.239)

Table 17: Shows the results from the finite mixture model that estimates the proportion of high types by demographic group in the first column and predicted enrollment and continuation by demographic group in the second and third column. Standard errors are bootstrapped with 250 simulations

Table 18: Cognitive and Non Cognitive Measurement: FMM

VARIABLES	Linear ASVAB Math Knowledge	Linear ASVAB Arithmetic Reasoning	Linear ASVAB Word Knowledge	Linear ASVAB Paragraph Comprehension
Intercept	-9.048	-11.077	-12.970	-10.231
	(1.176)	(1.097)	(1.104)	(1.149)
High Type	14.877	13.710	13.968	14.449
	(2.295)	(2.126)	(2.155)	(2.228)
Variance	6.988	7.05	6.479	6.077
	(0.503)	(0.428)	(0.470)	(0517)

	Probit Ever Sex bf 15	Probit Ever Violence	Probit Ever Stole gt 50	
Intercept	-0.488	-0.864	-1.454	
	(0.204)	(0.142)	(0.115)	
High Type	-0.646	-0.209	-0.128	
	(0.400)	(0.260)	(0.206)	

Table 18: Shows the results from the finite mixture model for human capital variables. High Type is a binary variable if agent is high type For earnings, expected log non college earnings are given by the intercept. Standard errors are bootstrapped with 250 simulations

Table 19: Grades and Earnings: FMM

VARIABLES	Logit Prob GPA (2.0-3.0)	Logit Prob GPA (3.0-4.0)
Intercept	0.767	-0.315
	(0.110)	(0.225)
High Type	0.565	1.939
	(0.177)	(0.352)
	Linear Earnings	
Intercept	9.879	
	(0.038)	
Ever Enrolled	0.423	
	(0.043)	
Bachelors	0.124	
	(0.067)	
Bachelor*High	0.256	
	(0.075)	

Table 19: Shows the results from the finite mixture model. High Type in expected GPA is a binary variable if agent is high type For earnings, expected log non college earnings are given by the intercept. Expected log some college earnings are the intercept added to the Ever Enrolled coefficient. For expected log college earnings add the bachelor's coefficient to expected log some college earnings if low type, if high also add Bachelor*High coefficient. These predicted values are used in the quantitative model. Standard errors are bootstrapped with 250 simulations

A.4 Indirect Inference: Targeted vs Simulated Moments

Table 20: Indirect Inference OLS Targets

VARIABLES	(1) Enrolled Data	(2) Enrolled Sim	(3) Continue Data	(4) Continue Sim
Intercept	0.376	0.287	-0.068	-0.012
	(0.033)	(0.065)	(0.0502)	(0.032)
High Belief	0.215	0.201		
	(0.019)	(0.027)		
Fin Assist T2	0.150	0.154	0.072	0.075
	(0.024)	(0.027)	(0.034)	(0.009)
Fin Assist T3	0.297	0.301	0.095	0.135
	(0.026)	(0.035)	(0.0403)	(0.014)
First Gen	-0.129	-0.034		
	(0.021)	(0.017)		
Parent HSD			0.077	0.061
			(0.0390)	(0.021)
Parent SCOL			0.128	0.150
			(0.0379)	(0.028)
Parent Bach			0.216	0.235
	(0.031)	(0.015)	(0.0478)	(0.029)
White	0.116	0.067	0.015	0.034
	(0.026)	(0.038)	(0.036)	(0.018)
Hispanic	0.107	0.036	-0.016	0.018
	(0.031)	(0.045)	(0.044)	(0.021)
GPA Med			0.214	0.159
			(0.0348)	(0.015)
GPA High			0.3724	0.424
			(0.0371)	(0.025)

Table 20: Shows the exact moments targeted via indirect inference, the regression coefficients from Enrollment on the covariants and regression coefficients from Continuation on covariates. Columns 2 and 4 show the simulated moments as well as bootstrapped standard errors of the coefficients.

Table 21: Key Internal Parameter Results

Parameter	Description	Estimate
$\gamma_{p,0}$	Belief Constant	0.0134
		(0.0127)
$\gamma_{p,b}$	Belief: Meas Belief	0.869
		(0.0092)
$\gamma_{p,h}$	Belief: P-Edu HSD	0.034
		(0.0118)
$\gamma_{p,s}$	Belief: P-Edu SCOL	0.030
		(0.0097)
$\gamma_{p,c}$	Belief: P-Edu Bach	0.059
		(0.0118)
σ_p	Belief: Var Error	0.00018
		(0.000043)
$\mu_{d,0}$	Non Pecun Util: Black 1st Gen Col Stud	-0.000088
		(0.000041)
$\mu_{d,C}$	Non Pecun Util: Col Edu Parents	0.000039
		(0.000032)
$\mu_{d,W}$	Non Pecun Util: White	0.000051
		(0.00003)
$\mu_{d,H}$	Non Pecun Util: Hispanic	0.000014
		(0.00003)
$\sigma_{c,1}$	Non Pecun Util Scale pd 1	0.000043
		(0.000066)
$\sigma_{c,2}$	Non Pecun Util Scale pd 2	0.000027
		(0.000066)
$\mu_c(\tau_h)$	Non Pecun Util high	0.00053
		(0.000066)
$\mu_c(\tau_l)$	Non Pecun Util high	-0.0031
		(0.000278)
$tuit_1$	Tuition Pd 1	$7430
		(63.36)
$tuit_2$	Tuiton Pd 2	$6946
		(60.84)

A.5 More Counterfactuals

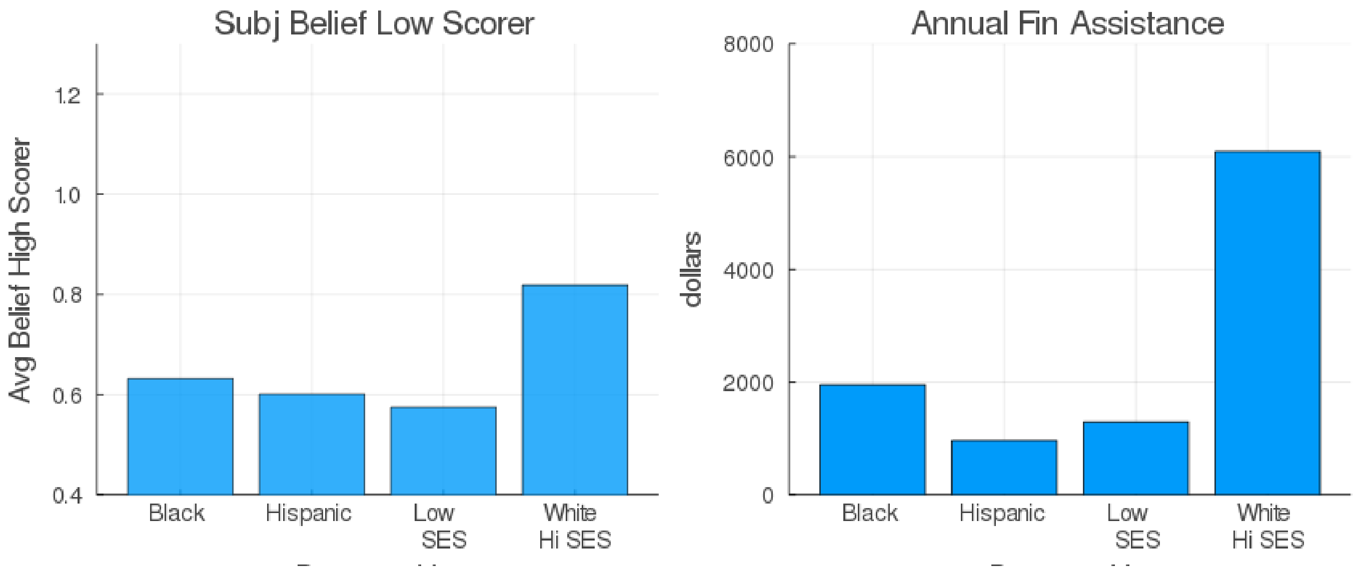

Figure 17: Financial Assistance by Demographic

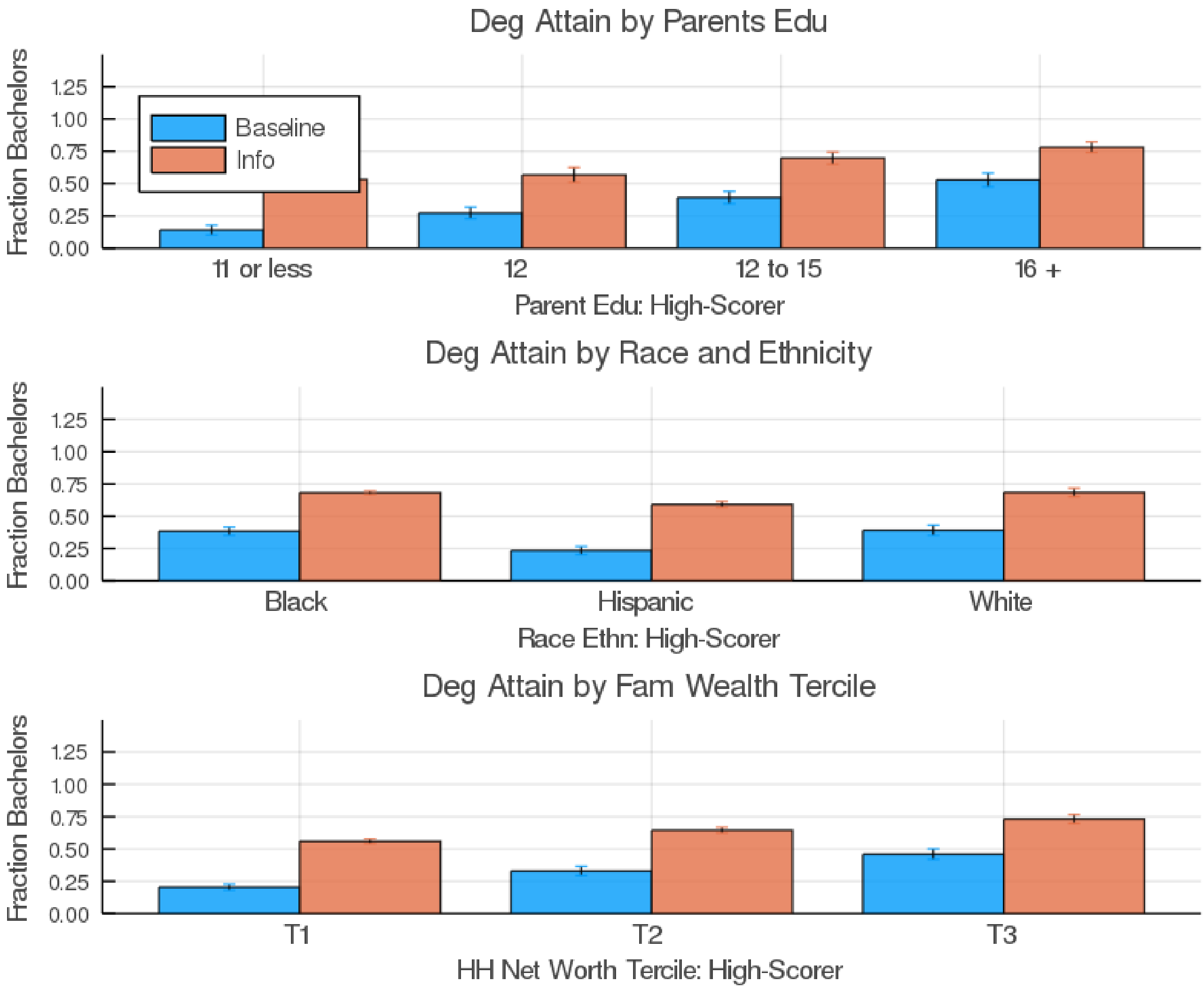

Figure 18: Shows difference in BA attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at predicted High Scorers by demographic group.

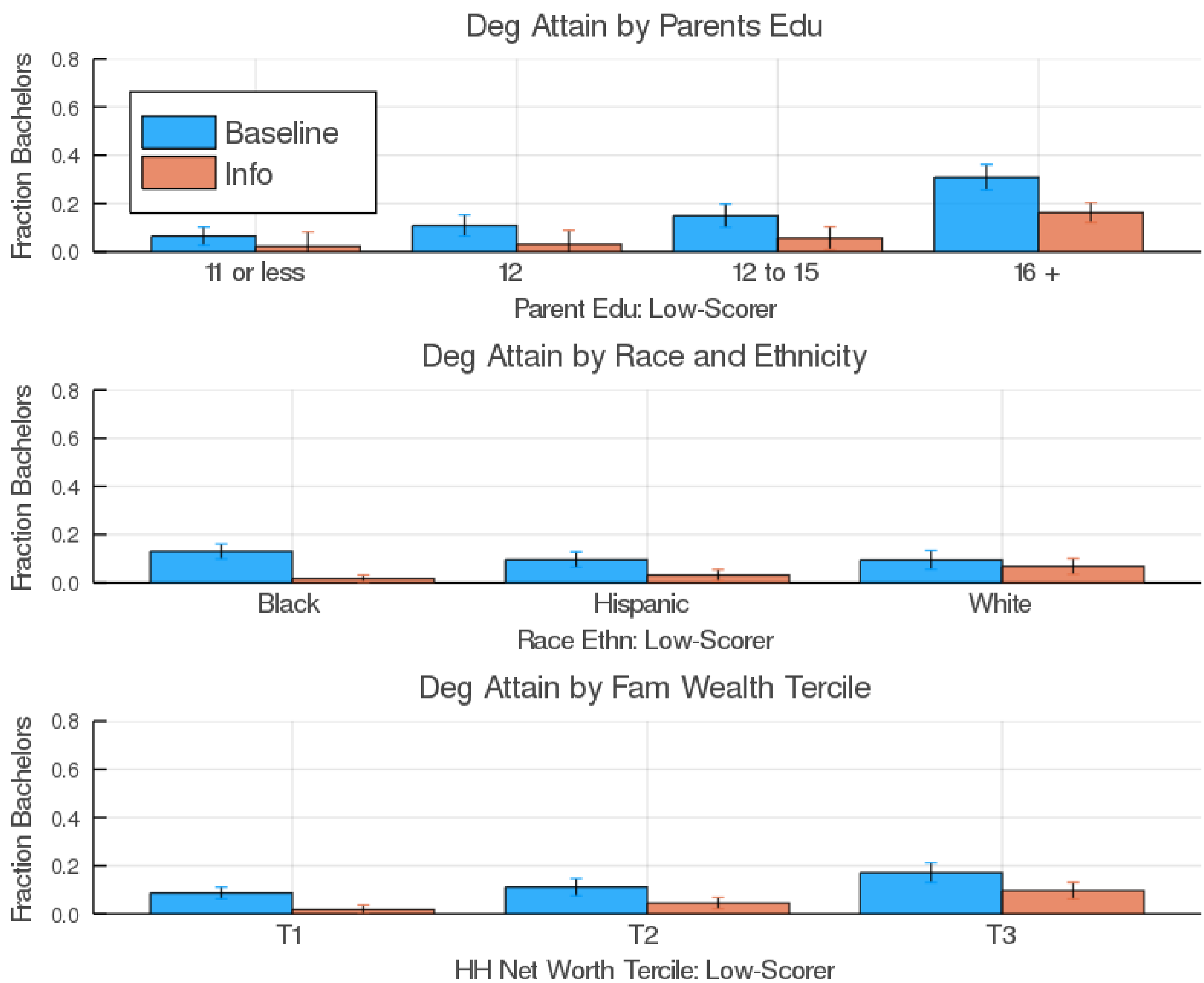

Figure 19: Shows difference in BA attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at predicted Lower Scorers by demographic group.

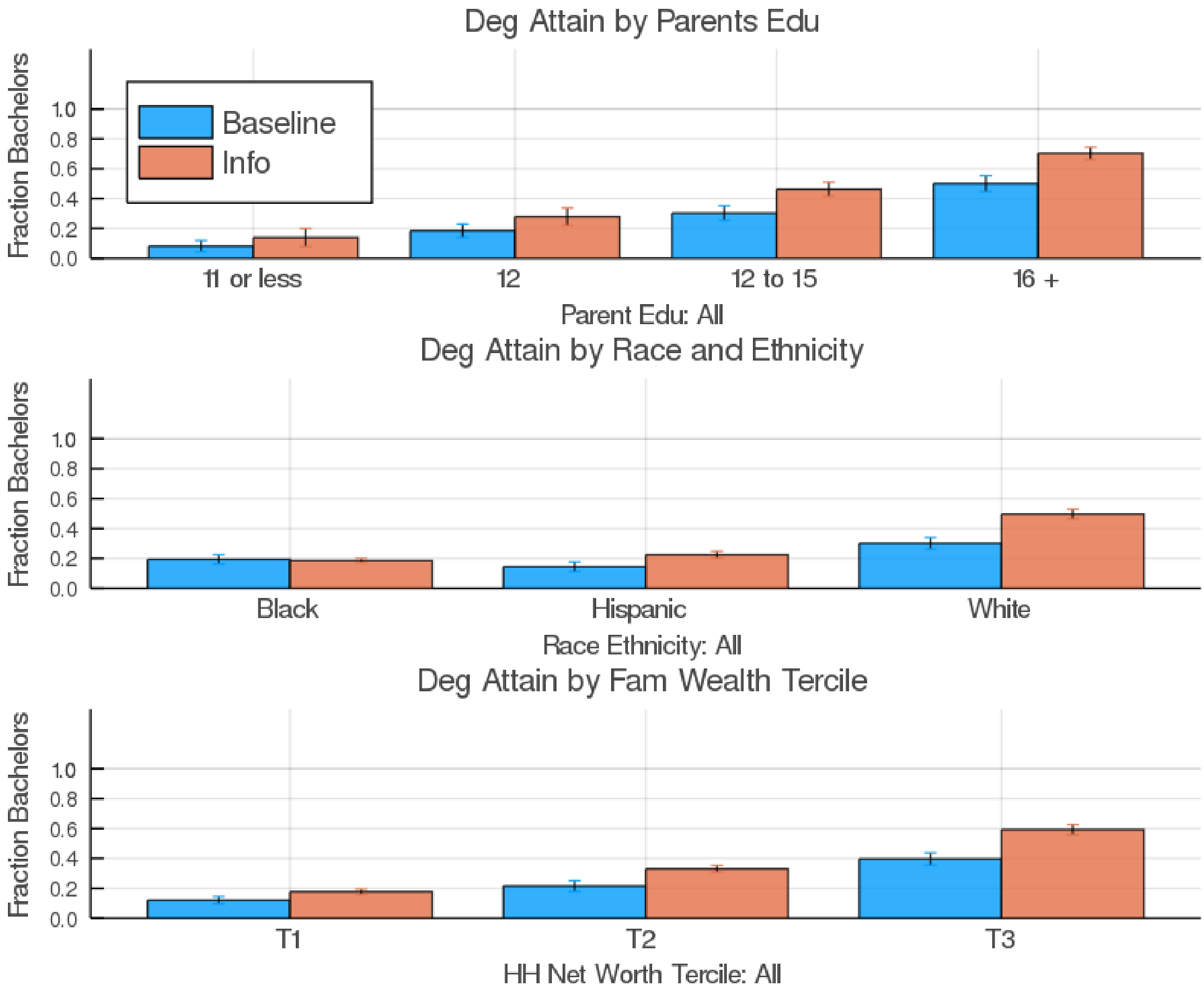

Figure 20: Shows difference in BA attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at all youth regardless of scoring type by demographic group.

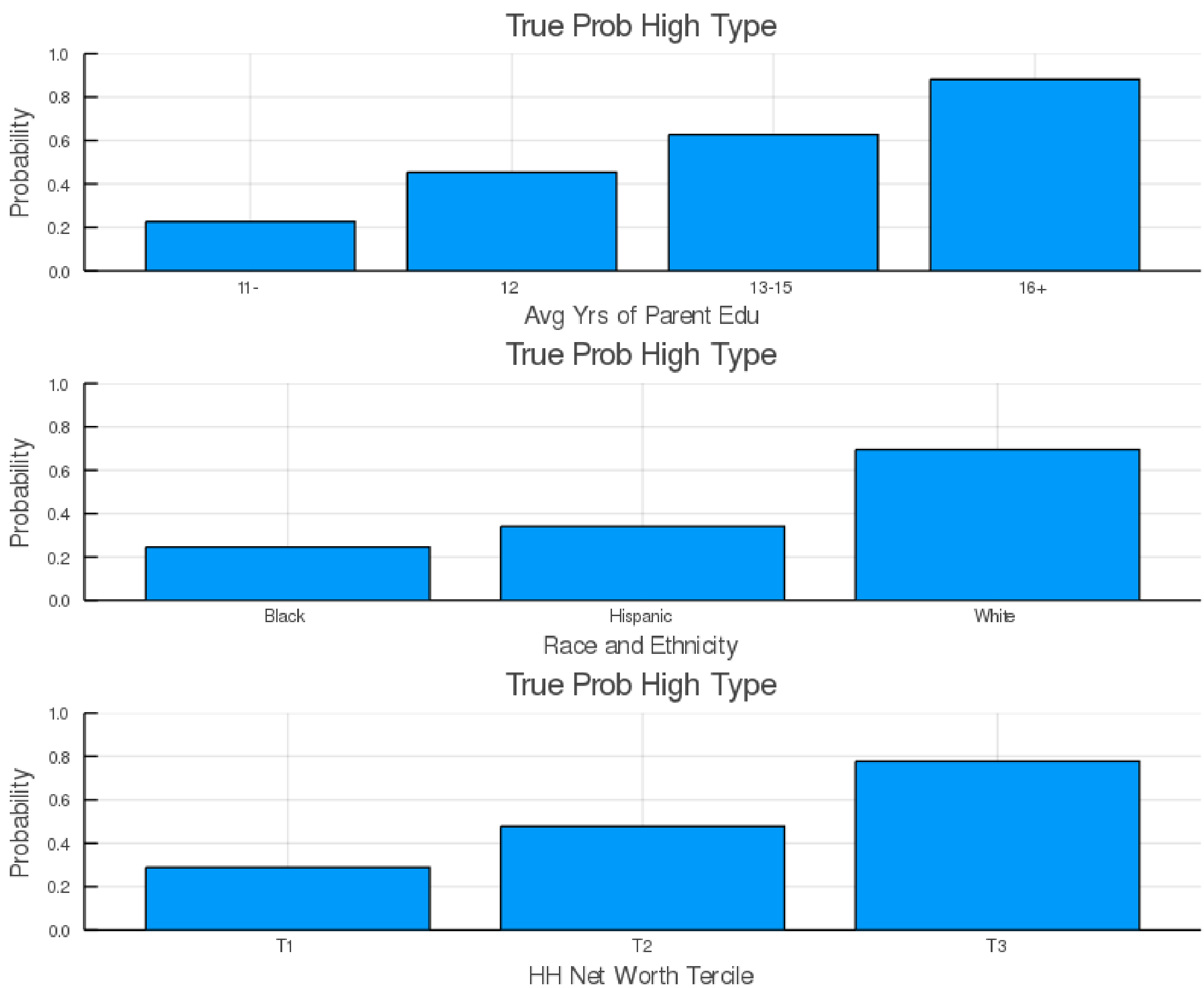

Figure 21: Shows the estimated fraction of high-scorers by demographic background from the finite mixture model.