Is College Worth It For Me?

Beliefs, Access to Funding, and Inequality in Higher Education

Sergio Barrera
University of Minnesota

September 2021

Abstract

This paper focuses on inequality in Bachelors degree (BA) attainment by race, ethnicity, and socioeconomic status and asks if targeting information and subsidies to young adults with high measures of academic ability (High Scorers) is more effective at closing gaps by race, ethnicity, and socioeconomic status (SES) than policies providing free college or information to all. To do this I estimate a dynamic discrete choice model with information and credit frictions using US panel data from the National Longitudinal Study of Youth 1997 which includes data on subjective beliefs about higher education. I find that pessimism regarding ability and resulting returns to college contribute between 20-25 percent of the gap in BA attainment for Hispanic and low SES High Scorers compared to White High Scorers from high SES backgrounds. I also find that differences in beliefs has little role in generating BA attainment gaps for Black high scorers. In the policy analysis targeting better information and funding to low SES High Scorers is more effective at decreasing overall inequality relative to High Scoring White high SES youth. Additionally this has the benefit of increasing completion among students who would benefit more from college, while universal free college also decreases inequality but mostly through increasing BA attainment from those who benefit less from college. Overall this paper shows targeted policies can be as effective, if not more efficient and cost effective at closing higher education gaps. Furthermore they can be strengthened by targeting racial and ethnic minorities regardless of socioeconomic status. However because of differences in the distribution of High Scorers and non pecuniary utility from college by demographic group higher education gaps will likely still persist as long as these differences are not addressed.

1 Introduction and Literature Review

In the United States there are persistent gaps in Bachelors degree (BA) attainment by race, ethnicity[1] and familial background[2]. If there is a large amount of youth who are prepared for and would benefit from college then obstacles preventing these youth from completing college can be costly [3]. In addition to barriers like limited financial resources and lower quality schools, differences in BA attainment can also be driven by systematic differences in beliefs about ability and resulting returns to college.

In this paper I ask how do differences in beliefs about ability and resulting returns to college contribute to differences in BA attainment for Black, Hispanic, and Low socioeconomic status (SES)[4] "High Scoring" youth compared to their counterparts from families with high wealth and high parental education levels or from High SES backgrounds. By High Scoring youth, I mean those with high cognitive test scores and good behavior who are predicted to have higher Grade Point Averages in college and higher earnings after college.

I also ask what is the effect of race neutral targeted policies that are designed to increase college outcomes of Low SES youth. I then compare their effects to universal policies. The policies will be in the form of providing free college, information on the true probability of Scorer type, and a combination of both[5]. The first outcome of interest is the relative difference of BA attainment of Black, Hispanic, low SES youth compared to White High SES youth irrespective of type. That is how effective is targeting only High Scorers

[1] In 2020, White BA attainment was 35 percent, while for Black and Hispanic young adults it was 21 and 15 percent respectively

[2] For Cohorts born in the early 1980s, top income quartile BA attainment was 54 percent, while for those in the bottom quartile it was 9 percent (Dynarski 2011)

[3] Hsieh, Hurst, Jones, and Klenow 2019, show that ending legal segregation in jobs that require college education was responsible for as much as 20 to 60% of growth from the 60s to 2021

[4] In this paper Low SES youth are those whose parents have an average education of High School or Less, or are in the bottom tercile of the household net worth distribution

[5] Universal Subsidy Policies can be compared to Free Public College or Free College for all policies,and Tracking systems as where predicted High Scorers are placed on a college track, while Lower Scorers are placed on a vocational track. Targeted policies can be thought of as generous grants/scholarships, or recruiting efforts and better guidance counseling

from disadvantaged groups in a race neutral way in closing overall gaps between demographic groups. The second outcome is a measure of efficiency that will be referred to as mismatch, which is the amount of youth who would change education decisions if they knew their Scorer type with certainty. This mismatch is often referred to as Ex Post Regret by the information frictions literature and in this paper takes the form of High Scorers under investing in college and Lower Scorers over investing in college.

I find that for Hispanic and Low SES High Scorers differences in beliefs about ability and returns to college account for 20-25 percent of the gap in BA attainment relative to White High SES High Scorers youth. There is also a significant role for differences in financial assistance and unobserved utility from college contributing to these gaps. For Black High Scorers beliefs do not play a major role and the difference is almost entirely due to differences in financial assistance and unobserved utility, which is contrary to the hypothesis proposed by Boerma and Karabarbounis 2021, who posit that more optimism of White Americans relative Black Americans creates the Black/White wealth gap.

In the policy analysis, I find that universal policies exhibit an equity-efficiency trade off. For example "Free College for All" is effective at narrowing differences in BA attainment, but at a cost of generating a higher fraction of youth who would choose differently under perfect information. This is primarily among Lower Scorers who would choose less schooling. On the other extreme providing more accurate information to everyone as would be done in a tracking system generates more inequality, but less mismatch.

Targeted policies do not exhibit this equity-efficiency trade off. They decrease gaps without increasing mismatch, where the most effective is the combination of providing information and subsidies. However the effectiveness of targeted policies is more limited for Black youth, suggesting that additional targeting by race and ethnicity would go further in closing gaps. Taken together this shows that because of the information frictions concerning individual ability and earnings type there is significant underinvestment of High Scorers in higher education among all demographic groups. Although free college and targeted interventions do narrow some of the observed gaps in higher education, gaps will likely remain persistent as long as there are differences in early childhood human capital development and unobserved utility from college.

I answer these questions by estimating a finite horizon dynamic discrete choice model with information frictions and learning about type through grade revelation in college. I use US panel data from a nationally representative sample, the National Longitudinal Study of Youth 1997 (NLSY97), for cohorts born in the early 1980s. I estimate the model using a two step procedure, where the first step uses data on grades, average lifetime earnings, human capital measures, demographic information, and years of schooling to estimate a finite mixture model (FMM) of two latent types for High Scorers and Lower Scorers. Then using the economic model and the FMM estimates, I internally estimate the distribution of beliefs about ability and returns to college[6], tuition "Sticker Price" (i.e price before financial aid), and the distribution of unobserved utility by race, ethnicity and parental education[7]. I identify these parameters through indirect inference where I match simulated moments to data moments relating enrollment and college completion to measured beliefs, financial assistance, grades, and familial background.

This paper will contribute to the empirical literature that suggests an important role in education inequality for rising tuition (Turner 2004), financial aid (Dynarski 2004), binding credit constraints (Lochner and Monge-Naranjo 2011), early childhood human capital (Cunha and Heckman 2007), and beliefs about ability and net returns (Dynarski, Libassi, Michelmore, Owen 2020; Hoxby and Turner 2013; Bettinger Long, Oreopoulos, Sanbonmatsu 2012). The estimated model will incorporate each of these mechanisms so that they

[6]The distribution of beliefs in the model is a linear function of subjective beliefs from the data and parental education

[7]This is to control for factors such as bias in college, lack of community, or familial obligations while in college.

could affect college enrollment/completion decisions in the baseline scenario and in the policy analysis.

The preferred policy recommendation of targeted information and subsidies also further validates the findings of Dynarski, Libassi, Michelmore, Owen 2020, and Hoxby and Turner 2013, who found large effect on enrollment of providing information about elite college acceptance and reduced tuition to High Scorers. This paper suggests that scaling up their interventions across the US can narrow gaps in higher education outcomes.

This paper is also related to the structural education literature studying information frictions in education decisions. One strand of the literature uses nationally represented 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, Cunha Navarro 2005). Unlike this paper these papers have little to say about non completion and the mechanism for learning. They also do not use data on subjective beliefs to estimate information sets[8].

Another strand of the structural literature uses data on subjective beliefs along with grades and demographics to study the roll of beliefs on dropout and major choice. Unlike this paper these papers do not examine enrollment or cover the whole US population. They also don't make policy inference on the wider US population because they only use data from a single university such as Berea College or NYU (Stinebrickner and Stinebrickner 2014, Wiswall and Zafar 2015, Reuben, Wiswall, and Zafar 2015).

This paper will bridge these two strands of the structural literature together by combining the strengths of each by studying enrollment and completion, using data on subjective beliefs to estimate information sets, and a sample that is nationally representative to estimate the effect of national policies on national outcomes. Additionally my paper will be the first to study the role of information frictions in generating inequality in higher education

[8]These papers estimate information sets by conducting Factor Analysis on the error terms of wage regressions along with hypothesis tests of whether coefficients on the estimated components were known at the time of enrollment

outcomes by race, ethnicity, and socioeconomic status.

2 Empirical Analysis and Facts

This paper seeks to understand the role for differences in subjective beliefs in generating gaps between White High SES youth and Black, Hispanic, Low SES youth. It is important in the analysis that the important role of financial assistance, credit constraints, and differences in human capital are also controlled for as well. Once that is done it is important to understand how much of these gaps are due to beliefs, financial assistance, human capital, and other unobserved preference shocks like bias, lack of community, familial obligations. Then given the role of each of these mechanisms we will discuss the effect of policy on inequality and mismatch.

Before discussing the model, this session will show that in the data in addition to the role of financial resources and human capital, subjective beliefs are also highly correlated with probability of college enrollment, continuation, and completion. Segments of this empirical analysis will inform the structural model as well as the moments used 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 used in order to examine the relationship between subjective beliefs about education outcomes and education decisions while controlling for other important factors like access to financial assistance, parental income/education, and race and ethnicity is the 1997 wave of the National Longitudinal Study of Youth (NLSY97). The NLSY97 is a nationally representative longitudinal data set of people born between 1980-1984 in the Unites States of America. The survey also over samples Black Americans and Hispanic

Americans, which makes it useful for studying racial and ethnic inequality. The survey was administered annually from 1997 to 2011 and then biannually from there forward.

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[9], labor market participation, income, schooling activities, financial assistance, and parental transfers. Additionally I use demographic information like Race, Ethnicity,census region, urban/rural categorical variables, gender, as well as year of birth. In the quantitative analysis to control for early childhood human capital and estimate whether one is a High Scorer I use grades, realized earnings, Armed Services Vocational Aptitude Battery (ASVAB) scores, and participation in adverse behavior such as theft, violence, and sexual intercourse before 15 recorded at the start of the study [10]. For the empirical results that follow AFQT will be used to control for cognitive human capital as well as adverse behavior indicators to control for non cognitive human capital as suggested by Hai and Heckman 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, income 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, but impute transcript GPA from self reported college GPA while in school. Since Asian Americans, Native Americans and races marked as other were smaller in the data set we restrict the analysis to Hispanic, White, and Black youth. In total we have a sample size of about 2,133 individuals. All statistics, regressions, and patterns in the empirical analysis are weighted using the sampling weights created by the Bureau of Labor Statistics for the NLSY97.

[9]For individuals that are missing Probability of Degree, I infer it using consecutive year reports of probability of being in college

[10]These are variables used by Heckman and Hai 2017, to control for cognitive and non-cognitive ability

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 9 and Table 10. The summary statistics in the appendix show that Black, Hispanic, and Low Education family youth have lower enrollment and BA attainment. 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 subjective beliefs reported as probability of having a degree are positively related to enrollment, continuation and degree 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. Since the Prob of Degree variable is measured between 0 and 1, an increase in probability of degree by 10 percentage points, or by 0.1, is associated with an increase in probability of enrollment by 3 percentage points, and in completing college conditional on enrollment by 2 percentage points. So in addition to other variables studied in the literature, it appears that beliefs have some explanatory power in college outcomes.

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.

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 in high school, and for the younger cohorts, probability of enrollment was asked while they were in high school. For Table 1 and the quantitative analysis a measure of Probability of Degree is used that infers from subsequent years reports of belief individual will be in school. Any bias in the inferenced question is controlled by year of birth dummies.

We see in Table 2 that Parental Education and household Net Worth holding all else constant are associated with more optimism. Similar to the last specification being Black or Hispanic is associated with more optimism regarding enrollment, and for Black youth with completion as well. Also interesting is that the number of peers who plan to go to college is also positively associated with more optimism.

Figure 1 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[11] It shows differences in continuation patterns among students who have already enrolled in higher education and who otherwise receive similar grades, but may differ with respect to parental education, race and ethnicity.

In terms of unconditional dropout rates, early human capital investments may still play an important role in generating outcomes, since differences in human capital investment would lead to differences in the composition of grade categories by demographic group[12] However conditioning on grades that are similar conditions on students with similar academic abilities. Differences in dropout rates by grade then suggests mechanisms other than human capital differences that are further leading to differences in college outcomes. In the first panel we see that for all GPA levels, Dropout rates are decreasing in parental education. When looking by Race and Ethnicity, we see that relative to White and Black students, Hispanic students also dropout at higher rates.

[11] Community College attendees are included because they have the option to transfer credits to a four year university. Also according to Hoxby and Aver 2012 it is not obvious that this is always the cheaper option.

[12] This will enter the quantitative model through the individual probability of being a High Scorer where High Scorers and Lower Scorers have different grade distributions.

The patterns of non continuation decreasing with grades are consistent with learning about own ability through performance in school. That is students may have an original prior estimate of their ability and earnings potential and update with information revealed through grades. If the new belief is proportional to the prior than receiving the same grade but having a lower prior estimate means that your new estimate would still be lower, and hence continuation would be higher at the same grade as shown in the bottom panel. Level differences could also be reflective of differences in unobserved shocks, like familial obligations and stress, lack of community in college, or bias and discrimination while in college.

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 potential earnings through grades. The rate of decrease also appears to be different between demographic groups which may be reflective of level differences in financial assistance that makes continuing college more costly relative to working for students from more disadvantaged backgrounds.

The similarity between White and Black students is surprising regarding dropout behavior, however unconditional mean beliefs and total financial assistance are closer between Black and White students than for Hispanic and White students. This may be the reason why conditional dropout rates are only higher for Hispanics and not for Black students. Hispanic students also comprise a larger share of the less than High School educated parents group than Black and White students do, which as panel one suggest exhibits the highest dropout rates.

Figure 1 suggest that differences in dropout between White and Hispanic students, as well as between students from lower education backgrounds and higher education backgrounds may be driven by differences in beliefs, in addition to access to funding[13] and human

[13] Appendix A.1 More Empirical facts shows that access to funding through parents, government and colleges differs by parental education, household net worth, race, and ethnicity

capital. Since Black students are more optimistic and receive more total funding on average, this may suggests that conditional on enrollment, increased funding and optimism decrease gaps by performance.

2.3 Discussion and Summary of Empirical Facts

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

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, less likely to dropout 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. There are also differences in beliefs between race and ethnicity.

3. Although on average Black and Hispanic youth are more pessimistic, have lower measures of human capital, receive less total financial assistance, and come from lower educated backgrounds, compared to similar White youth, they are more optimistic regarding college outcomes, more likely to receive aid from government or institutions, and more likely to enroll and Black youth are also more likely to complete college [14].

What this means is that we find evidence of a connection between subjective beliefs and college outcomes like enrollment, dropout, 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 shown in Figure 15 in appendix A.1. there is little evidence of differences in lower returns to college for Black, Hispanic, low familial wealth, and low familial education youth.

[14] See footnote 11

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, but drop out at higher rates than White students. 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 background.

The positive coefficient for Black youth on beliefs may also mean that information may be less relevant for black youth then when just looking by familial wealth and education. In section 4, we will revisit to what extent there are high achieving youth from different backgrounds that do not enroll in college and are kept out because of information frictions.

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.

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 in addition to 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. Finally, the economic model will be used to evaluate the effect on inequality and mismatch of universal and targeted policies providing free tuition, information and the combination of both. The model predicts that lower net tuition as well as more optimism regarding latent ability will have higher levels of enrollment, persistence, and completion, holding all else constant.

Additionally, higher levels of early human capital will manifest in a higher likelihood of realizing higher grades and higher earnings. This will allow final outcomes to depend on early childhood human capital, optimism, and access to resources, all three of which as the data suggests may depend on parental education, wealth, race and ethnicity. Therefore the model captures the relationships seen in the data between access to resources, optimism, parental education, wealth, human capital, enrollment, dropout, and degree attainment.

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 to coincide with the end of post secondary education up until retirement. In each period agents can save or borrow up to a specified borrowing limit, that differs while in school. The 23 periods will be broken up into 3 stages where the final stage acts as an absorbing state, which allows the education problem to essentially be a two stage problem.

Information frictions and the role for beliefs enters the problem in that post college earnings and utility depend on an unknown type $\tau \in \{\tau_h, \tau_l\}$ for High Scorers and Lower Scorers. The realization of τ depends on true probability P_{true} of being type τ_h, which depends on parental education, household net worth, race, ethnicity, sex, measures of human capital and year of birth. P_{true} therefore captures the role that early childhood human capital investment plays in generating outcomes and earnings by background, and can even include a potential role for labor market discrimination as well.

Agents do not know P_{true} but have a subjective belief P which they update after receiving grades in college as in Stinebrickner & Stinebrickner 2012. Since P is the belief over type that generates grades, earnings, and non pecuniary utility from school, P captures a broad belief about the appropriateness of school for the individual, that captures the student's ability revealed through grades, their earnings potential, as well as any non pecuniary benefits. Non pecuniary benefits that depend on type enter through the term $\mu(\tau)$. This can also capture any dis utility that is correlated with grade realizations.

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, asset level b_1, unobserved preferences for work school, $\vec{\varepsilon}_1 = (\varepsilon_{c,1}, \varepsilon_{w,1})$ respectively and choose between enrolling in college where they pay net tuition fee $f_1 = Tuit_1 - Aid_{GC} - Aid_{Fam}$, or work and earn non college earnings w_n. Notice f_1 is the sticker price $Tuit_1$ net of aid from government or college Aid_{GC} and families Aid_{Fam}.

In the second stage at around age 20 agents realize a signal for their latent type given by the GPA g for the previous schooling period. They then update P to $P'(P, g)$, and observe unobserved preferences for school and work, $\vec{\varepsilon}_2 = (\varepsilon_{c,2}, \varepsilon_{w,2})$ respectively. They then decide to continue schooling and pay net tuition $f_2 = Tuition_2 - Aid_{GC} - Aid_{Fam}$, or dropout to work and earn earnings rate w_s for having completed some college for the remaining periods of their life. Again f_2 is the sticker price $Tuit_1$ net of aid from government or college Aid_{GC} and families Aid_{Fam}.

If agents choose to complete school then after the next period from ages 22-66 agents work and earn earnings dependent on type, $w_c(\tau)$ each year as well as non monetary utility $\mu(\tau)$. 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.

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

In order to allow for human capital development while in school mean earnings are such that $\mathbb{E}[w_n] < \mathbb{E}[w_s] \leq \mathbb{E}[w_h]$ reflecting increasing mean returns to years of schooling regardless of type, although it may still be possible for comparative advantage to exist at an individual basis.

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, earnings w, and t since this determines how many periods agents have left in their life cycle.

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

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

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

For every period the borrowing constraint is the natural borrowing limit, given below which is 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, unobserved tastes for college and work $\vec{\varepsilon_1} = (\varepsilon_{c,1}, \varepsilon_{w,1})$ a belief P that they are of type τ_h. The agent's stage 1 problem is thus given by (3) below.

$$(3) \quad V_1(P, b_1, f_1, \vec{\varepsilon_1}) = \max\{V_w(w_n, b_1, 1) + \varepsilon_{w,1}, V_{c,1}(P, f_1, b_1) + \varepsilon_{c,1}\}$$

$$s.t.$$

$$V_{c,1}(P, f_1, b_1) = \max_{b_2 \geq -\tilde{B}_{s,1}} [u(Rb_1 - f_1 - b_2) + \beta \mathbb{E}_{g,\varepsilon}(V_2(P'(g, P), f_2, b_2, \vec{\varepsilon_2})) \; |P \;]$$

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

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

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 to $P'(g, P) = P'$. Agents also begin the second stage with debt/savings from the first stage b_2, and realize unobserved tastes for college and work $\vec{\varepsilon_2} = (\varepsilon_{c,2}, \varepsilon_{w,2})$ respectively. The agent's problem is given by

$$(5) \quad V_2(P', f_2, b_2, \vec{\varepsilon_2}, \epsilon_{w,2}) = \max\{V_w(w_s, b_2, 2) + \varepsilon_{w,2}, V_{c,2}(P', f_2, b_2) + \epsilon_{c,2}\}$$

$$s.t.$$

$$V_{c,2}(P', f_2, b_2) = \max_{b_3 \geq -\tilde{B}_{s,2}} [u(Rb_2 - f_2 - b_3) + \beta(P'[V_w(w_c(\tau_h), b_3) + \mu(\tau_h)]$$

$$+(1 - P')[V_w(w_c(\tau_l), b_3) + \mu(\tau_l)]) \;\;]$$

Grades primarily reflect information about τ, and their distribution depends on τ. But since τ also determines earnings and non pecuniary utility, the information revealed in school can also include psychosocial elements of higher education that are often discussed in the sociology literature. Although in the model it is realized additively with the next period value function, it can also reflect discovering this aspect of "fit" for college in utility after the first period that realizes itself in the second period or any period thereafter. In this model, the assumption is that this is closely tied to performance, and a bad signal in performance will likely reinforce psychosocially that college will not be a good fit for the individual. Concepts that are likely to be more stable between the first and second change such as community in college, distance from home community, or enjoyment of school would be captured through a constant location parameter of non pecuniary shocks $\vec{\varepsilon_t}$.

Since in the first two stages the agent faces a discrete choice problem, the optimal decision for each agent can described by a cutoff rule with respect to the individual subject belief of being type τ_h. For example in the first stage the optimal decision could be characterized by equation 6 below, where $\sigma_{c,2}, \mu_{c,2}$ are the scale parameter and the difference in location parameters and for the Type I extreme value shocks.

$$(6) \qquad \text{Choice}_{t=1} = \begin{cases} \text{Enroll} \quad \text{if} \quad P > \tilde{P}_1(b_1, f_1, \vec{varepsilon_2}, \mu_{c,2}, \sigma_c) \\ \text{Work} \quad \text{if} \quad P \leq \tilde{P}_1(b_1, f_1, \vec{varepsilon_2}, \mu_{c,2}, \sigma_c) \end{cases}$$

.

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

(7) below.

$$(7) \qquad \text{Choice}_2 = \begin{cases} \text{Continue} \quad \text{if} \quad P'(g,P) \geq \tilde{P}_2(b_2, f_2, \varepsilon_{c,2}, \varepsilon_{w,2},) \\ \text{Dropout} \quad \text{if} \quad P'(g,P) < \tilde{P}_2(b_2, f_2 \varepsilon_{c,2}, \varepsilon_{w,2},) \end{cases}$$

.

The cutoff rules holding non monetary utility shocks and distribution constant, are weakly increasing in f_1, f_2. In certain spaces of the distribution of non monetary utility shocks the decision rules are strictly increasing in f_1, f_2. $P'(g,P)$ also increases in P, and depending on the signal strength of grades, particularly high grades, may increase in g. 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, while for students with low levels of financial assistance they must have higher initial beliefs to enroll and higher beliefs or better grades to continue.

3.4 Example for Model Prediction

Figures 3, 4, and 5 provide an example for how financial assistance, subjective belief of being a High Scorer or having $\tau = \tau_h$, 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. These means that if two youth have the same beliefs but lower access to resources than probability of enrollment will still be different. Like wise if their access to resources are the same but beliefs differ to a certain extent.

Figure 4 shows how 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 level lead to an increase in the probability of continuation. For the same initial belief, low grades lead to a dramatic decrease in continuation except in the case where the agent is near certain they are a High 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. But this paper will focus on this formulation since most enrollment occurs before the age of 25 and increased saved assets can be captured through reductions in f_1.

The second objection raised can include the fact that probability of acceptance 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 this as well. However since in the data we look at any enrollment including at non selective community colleges we do not include it, since there the acceptance probability is likely high. 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, since over estimating tuition and learning it were lower after enrollment would be equivalent to f_1 being higher than f_2. However in the calibration this would be an average uncertainty in tuition not dependent on demographics, which may matter. 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 degree 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 2. I will also describe some of the assumptions governing the distribution of earnings as well as parameters whose values will be given outside of the estimation routine. As part of this we will discuss what data moments 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, the distribution of beliefs and non pecuniary utility. It will also allow for room for differences by race, ethnicity and parental background in human capital development, and earnings which will be embedded in P_{true}.

First I will explain the externally estimated parameters. Then I will explain the moments and parameters that will be estimated using indirect inference with 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 3. The coefficient of relative risk aversion γ, the discount factor β, and the interest rate $(1 + r)$ are set to standardly assumed values. The first stage borrowing limit while in school is set to match average borrowing limits which are close to $16,600 in 2017 dollars (Abbot et al 2016). The second period borrowing limit is set to $31,100 to match average student borrowing at 4 year public institutions (Wei & Skomsvold, 2011). In total the amount students are allowed to borrow in the model are higher than the highest cumulative total that students could borrow from Federal student loan programs for a bachelor's degree, $46000, which likely reflects the use of private loans amongst some students (Lochner & Monge Naranjo 2010).

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

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

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 aid $k = 2$. To get total financial assistance, the sum of both predicted values for students is used. Therefore financial aid used in the model is the predicted value given, demographic, and socioeconomic variables (Hai & Heckman 2017).

The distribution of latent type τ, will be estimated using a finite mixture model. The latent variable will take two values for $\tau \in \{\tau_l, \tau_h\}$,respectively corresponding to Lower

Scorers and Lower Scorers in the rest of the paper. The distribution of unobserved latent type determining wages, grades and test scores is given with probability of being a High Scorer below.

$$(9) \qquad P(\tau_h) = f_h(X_i) = \frac{\exp(\zeta_k X_i)}{1 + \exp(\zeta_k X_i)}$$

Where $P(\tau_l) = 1 - P(\tau_h)$. Normalizing $\tau_l = 0$ the values of τ_h, used for determining the effect of being type τ_h, will be estimated by estimating the following measurement equations for the finite mixture model.

$$(10) \qquad Z^*_{i,j} = \alpha_{z,j}\tau_i + \eta_{z,j}X_i + \varepsilon_{z,j} \qquad j \in \{1, \dots, J_c\}$$

$$(11) \qquad \ln w^*_{i,s} = \mu_{w,0} + \mu_{w,1}1(s \in (12, 16)) + 1(s \geq 16)(\mu_{w,2} + \mu_{w,h}\tau_i) + \varepsilon_{w,s}$$

$\ln(w_{i,s})$ are average earnings dependent on years of school s for individual i, and 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. $Z_{i,j}$ are measures of cognitive and non cognitive ability, as used in Hai & Heckman 2017.

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, like violence, theft, and sex before the age of 15. To incorporate both binary and continuous variables the following assumption is made

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

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 12 below.

$$(12) \qquad \pi(g|\tau) = \frac{exp(\gamma_{g,0} + \gamma_{g,\tau}\tau)}{\sum_{k=l,m,h} exp(\gamma_{k,0} + \gamma_{k,\tau}\tau)}$$

Using $\vec{Z}_i, w_i, g_i$ to denote the vector of $Z_{i,j}$ of human capital measurements, earnings, and grades of each individual the individual likelihood function is given by $f(\vec{Z}_i, w_i, g_i; \tau_k, X_i, s)$. The parameters for the probability of type τ_h, the distribution of earnings, conditional probability of grades, and the measurement equation for human capital are estimated by solving for the maximum simulated likelihood given below in equation (11). For specifics regarding the functional form, as well as the parameter results of the individual likelihood see the appendix.

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

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, $P(\tau_h)$ in equation (9) for P_{True}, 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.

4.2 Internally Estimated Moments

The remaining moments to be calibrated within the model are the sticker price of tuition t_1, t_2, the distribution of subjective beliefs of being type τ_h, the non pecuniary utility dependent on τ, $\mu(\tau)$, as well the distribution of the Type I extreme values shocks that differ for White, Black, and potential first generation students and a different variance for stage 1 and stage 2. For a quick summary of how the distributions of the internal parameters are calibrated see Table 4.

The distribution of subjective beliefs of being high type is given by a truncated Normal distribution at zero and one given by equation (16) below. The variance of $\epsilon_{p,i}$ will also be estimated.

$$p_i = \gamma_{p,0} + \gamma_{p,b}ProbDegr + \gamma_{p,h}HSD + \gamma_{p,s}SCol + \gamma_{p,b}Bach + \epsilon_{p,i} \tag{14}$$

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 type (τ_h). The measurement error is allowed to differ by parental education. Truncated normal is used since we want to allow for 1's and 0's since these are meaningful in the model meaning that agents are certain of their type. The parameters will be internally calibrated by indirect inference. The moments that will be targeted are the coefficients for the following two regressions in equation (15) and (16).

$$Enroll = \beta_{E,0} + \beta_{E,B}HighBelief + \beta_{E,F_2}T2(Finaid) + \beta_{E,F_2}T3(Finaid) \tag{15}$$

$$+\beta_{E,P}ParBach + \beta_{E,W}White + \beta_{E,H}Hisp + \varepsilon_{E,i}$$

$$(16) \quad Continue = \beta_{C,0} + \beta_{C,G}GP\vec{A}_{cat} + \beta_{C,F_2}T2(Finaid) + \beta_{C,F_2}T3(Finaid) + \vec{\beta}_{C,PH}ParHSD$$

$$+\vec{\beta}_{C,PS}ParSCOL + \vec{\beta}_{C,PB}ParBach + \beta_{C,W}White + \beta_{C,H}Hisp + \varepsilon_{C,i}$$

Equations (15) and (16) capture the effect that beliefs above a certain threshold play in the decision to enroll and continue, which as described in the model section follows a cutoff rule. They also include an affect of the level of net tuition through the tercile of financial assistance, as well as a separate effect holding all else constant of race, ethnicity and parental education.

The response to grades is determined by the beliefs and $\mu_c(\tau)$. The response to measured beliefs in the enrollment decision is also determined by the distribution of beliefs. The mean utility constants by demographic group will target the difference in enrollment and continuation levels after controlling for financial assistance, beliefs and grades.

Responses to financial aid in enrollment and continuation will identify t_1 and t_2. While μ_c and $\sigma_{c,t}$ capture levels of education decision rates among demographic groups holding access to funding and beliefs constant. They capture any other unexplained variation in decisions that could differ by Race, Ethnicity, and for potential first generation students. Factors captured by this are the expectation of financial strain, familial obligations, bias and discrimination, or severing community ties while in school.

The parameter vector Γ are those parameters that minimize the difference between the simulated regression coefficients and data regression coefficient. 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.

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

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.

5 Estimation Results & Policy Analysis

In this section I will present the results from the calibration exercise and then discuss the decomposition of how beliefs, financial assistance, and non pecuniary utility generate inequality in higher education among High Scorers. In the quantitative exercise High Scorers are those predicted to have τ_h given ASVAB scores and adverse behavior which are reflective of human capital measures revealed prior to completing college. Then I will discuss how information frictions create inefficiency in the form of Ex Post Regret, which includes Under Investment of High Scorers, and Over Investment of Lower Scorers. Finally I discuss the effects of universal vs race neutral targeted policies designed to increase access to High Scoring youth from low wealth and education backgrounds.

The policies I will examine are in the form of subsidies, information campaigns revealing true individual probability of being high type, and the combination of the two. The Universal Subsidy can correspond to Free College For All, as proposed in the 2020 Democratic primary, while the information for all can correspond to a tracking system as in place in many European countries, where continued high scoring leads to youth being placed in a track to go to college and others in a vocational track. The targeted subsidies can be thought of as scholarships and grants, while targeted information can correspond to highly effective recruiting and information campaigns as in Hoxby & Turner 2012, and Dynarski, Libassi, Michelmore, Owen 2020.

5.1 External and Internal Estimation results

Table 5 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 Scoring type. However High Scorers have higher earnings than Lower 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 and receive lower returns by being a Lower Scorer.

Table 6 shows several of the key parameters that were estimated in the internal calibration exercise. What is important to notice is that a percentage point increase in the measured belief from the data, probability of degree attainment, translates to about a 0.87 percentage point increase in belief of being a High Scorer with τ_h. Holding the measured belief constant as well, we see that the higher education background a youth comes from the more optimistic they are that they are type τ_h.

Figures 6-8 provide a quick snapshot of how well the model matches patterns we see in the data. Figure 7 and the first panel of Figure 6 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 8 and the second panel of Figure 6, this success at capturing BA attainment carries over when we condition by demographic group as well, where gender and household net worth were not directly targeted. For more information regarding model fit of the indirect inference moments see Appendix A.3.

Figure 9 shows the mean values of variables corresponding to the three main mechanisms generating education decisions, subjective beliefs, financial assistance, and early childhood human capital in the form of the true probability of being a High Scorer. The true probability of being a High Scorer is calculated by using the posterior of the likelihood function from the finite mixture model, and incorporates measures of cognitive human capital, non cognitive human capital, grades, and earnings.

Figure 9 shows that all High Scorers under consideration underestimate the true probability of their type. This pessimism is worse for Hispanic and low SES youth than it is for Black youth. We also see that White High SES youth receive much more financial assistance than Black, Hispanic, and Low SES youth as well. This means relative to White High SES High Scorers we should expect differences in subjective beliefs to play a bigger role in generating gaps for Hispanic and Low SES High Scorers, than for Black High Scorers. Funding should play an important role in generating gaps for all groups under consideration as well.

5.2 Decomposition of Mechanism

Next I decompose the gap in higher education backgrounds for High Scorers by differences in subjective beliefs, financial assistance, and unexplained sources brought about by differences in Non Pecuniary utility. We compare each group to a reference group with high college completion rates, White High SES High Scorers. For the decomposition and the later policy analysis by Low SES youth, we mean those whose family is in the bottom tercile of the wealth distribution, or whose parents have a high school diploma or less. We perform the decomposition by equalizing beliefs and financial assistance to the mean of White High SES High Scorers.

Figure 10 provides a visual representation of this, while table 7 provides the numerical value for the size of the college completion gap, and the portion explained by each mechanism.

Figure 10 shows the completion rate of each demographic group divided by the completion rate of the reference group. The graph and Table 7 show that for most demographic groups the biggest portion of the gap in college completion is due to unequal financial assistance, most likely resulting from family assistance. We see for Hispanic and Low SES youth gaps in beliefs contribute to between 20 to 23 percent of the gap, while for Black youth beliefs contribute very little. Other reasons also contribute a large portion of the gap ranging between 24% for Hispanic youth to as high as 44% for Black youth.

5.3 Ex Post Regret in the Baseline Scenario

Before discussing the policy analysis, it is important to understand how information frictions in the model generate mismatch in the higher education market. In this analysis mismatch is taken to mean under investment among High Scorers who would have higher BA attainment rates under perfect information. It also includes over investment among Lower Scorers who would have lower BA attainment rates if they knew their types.

Figure 11 shows the BA attainment rate of High Scorers in the baseline scenario and under perfect information about type by household Net Worth, Parental Education, and Race of ethnicity. The first thing to notice is that there is substantial underinvestment among High Scorers. That is there is a substantial fraction of students who would benefit from college that choose not to do so, because they do not know the true probability of their type as shown by the orange bars.

Figure 12 shows the BA attainment rate of Lower Scorers in the baseline scenario and under perfect information by the same demographic groups. Conversely there is substantial over investment occurring in the higher education market for the three demographic groups and appears to increase with parental education and wealth.

Finally Figure 13 shows the aggregate effect of true beliefs about the probability of Scoring Type. Because of differences in the distribution of High Scorers and Lower Scorers between demographic groups, we can see that for some groups there is a decrease in BA attainment, little positive increase in BA attainment, and for youth primarily from more socially advantageous backgrounds big increases in BA attainment.

5.4 Policy Analysis

In this section we will discuss the implications of several policies on overall inequality regardless of type in higher education outcomes and economic efficiency measured by mismatch in education. This is the percent of youth who would have made different education decisions if they had known their type. Examples of this are Under Investment for High Scorers and Over Investment for Lower Scorers. In the Information Frictions literature this concept is also known as ex-post regret. For these policies we are looking at inequality by demographic group irrespective of Scorer type, so persistent gaps will likely reflect differences in early childhood human capital generating different proportions of High Scorers by demographic group, as well as differences in non pecuniary utility.

The first set of policies we evaluate are universal policies that affect everyone. The policies considered are setting tuition equal to zero, or Free College for All. The second policy provides information on what everyone's true probability of being a High Scorer is. Real world policies that could function like this are tracking systems, where based off of human capital measures students are tracked into a college route or vocational route. The third universal policy is just a combination of the two previously mentioned.

The top panel of Figure 14 shows the Degree attainment rate for the group overall relative to Degree attainment rate for White High SES youth irrespective of type. We see that among universal policies Free College for all is the only policy that decreases inequality for the groups of interest, while both of the policies providing better information actually lead to more inequality. This is probably due to the fact the High Scorers are less common among Black, Hispanic, Low SES youth than for White High SES youth reflective of disparities in early childhood human capital development.

The second set of policies considered are race neutral policies targeting High Scorers from low wealth and familial education backgrounds. We see that the effects of the policies on Black youth and to a certain extent Hispanic youth are less effective than on Low SES youth in general. This is because Black and Hispanic youth benefit from the policy only to the extent that their parental education is high school or less, or that they're in the bottom tercile of the household net worth distribution. For the most part it appears that the combination of information and financial assistance is more effective at closing inequality. It appears more effective than even Free College for all.

In both panels gaps still exists, highlighting that the early childhood human capital measure 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.

Finally Table 8, shows us the amount of mismatch present in the economy and how it is distributed among High Scorers and Lower Scorers. What we see is that for the youth of the same cohort as those in the NLSY97, 28 % 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 a slight decrease in High Scorer mismatch, but an increase in Lower Scorer mismatch. However targeting subsidies has almost no perceivable increase or effect on mismatch. Targeting funding has a slight increase in mismatch, due probably in part to Lower Scorers who are marginally similar to High Scorers that benefit from the intervention. Whether information revelation is targeted or universal, Mismatch decreases even if the beliefs are only 10 percent closer to the truth. This is also primarily among High Scorers. Since Lower Scorers are not the target of information campaigns their mismatch does not decrease under the targeted program but does substantially under the universal program.

Together this shows that there are substantial frictions that exists in the economy leading to mismatch.Table 8 and Figure 14 demonstrate that Universal programs exhibit Equity vs Efficiency trade offs. Targeted policies do not have this trade off. The combination of information and subsidies can be just as effective as Free College for all at closing gaps. However targeted policies are likely to be more effective at closing racial ethnic gaps if they are no longer race neutral. In practice providing subsidies to only a subset of students is likely much less resource intensive than subsidizing college for all. There is still a large persistent gap in BA 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 first generation College Students.

6 Conclusion

In this paper we investigated the role that beliefs played in generating inequality in higher education outcomes for High Scoring youth while controlling for financial resources. 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 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 we showed that relative to White High SES High Scorers, beliefs contribute between 20-25 percent of the BA attainment gap for Hispanic and Low SES High Scorers. Beliefs play a very small role for Black High Scorers. I find that universal policies providing financial assistance or information that can be compared to Free College For All, or Tracking Systems 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 Lower Scorers. More information on the other hand decreases mismatch but leads to more inequality because of differences in early childhood human capital investment. Race Neutral targeted policies on the other hand, especially those combining information with subsidies to low SES High Scorers create little additional mismatch and have bigger effects on closing gaps, especially among Hispanic and Low SES youth. To have bigger effects targeting by race and ethnicity may also be necessary.

Therefore this paper shows that information frictions do lead to less High Scoring youth from all backgrounds under investing in education. We can begin to close gaps with more information and subsidies to Low SES High Scoring youth without necessarily providing free college to everyone that would be more resource intensive and have potentially negative effects on Lower Scorers. Targeted interventions can take the form of increases in scholarships and grants as well as more recruitment, better guidance counseling of High Scorers from disadvantaged backgrounds.

However because of differences in early childhood human capital development, overall gaps are likely to persist even if High Scorers from disadvantaged groups enroll at equal rates of White High SES High Scorers. Additionally Black, Hispanic, and Low SES youth are likely to continue enrolling and continuing college at lower rates due to differences in unobserved non pecuniary utility. This non pecuniary utility could be due to familial obligations, bias and discrimination, or thinner social networks while in college.

This paper shows that there is room for improvement in the higher education system by working on belief and financial resource gaps, but in order for all gaps to be closed we must also see what effects improving K-12 education, household environment, and inclusion efforts in higher ed have on alleviating early childhood human capital gaps and difference in non pecuniary utility.

7 Tables and Figures

7.1 Empirical 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***
	(0.0280)	(0.0283)	(0.0491)
Female	0.0831***	0.0847***	0.0411*
	(0.0164)	(0.0186)	(0.0237)
Hispanic	0.0812***	0.0535*	0.0525
	(0.0286)	(0.0286)	(0.0381)
Black	0.1700***	0.1487***	0.1732***
	(0.0261)	(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
Robust Standard Errors	Yes	Yes	Yes
Peer Effects	Yes	Yes	Yes
Observations	2,133	2,133	1,467
R-squared	0.3499	0.3612	0.3240

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	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
Robust Standard Errors	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.

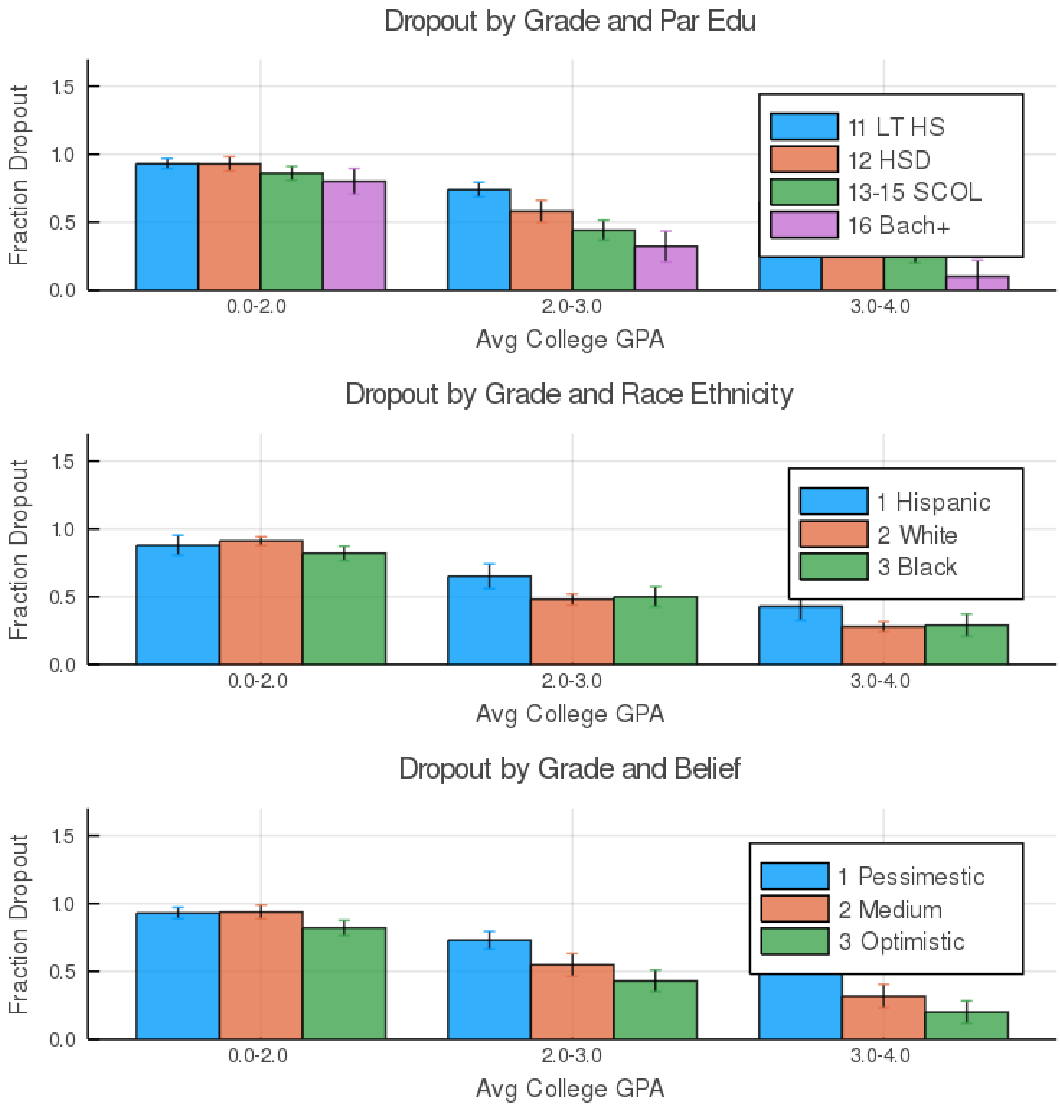

Figure 1: Non Continuation Rates Conditioned on Grades/Demographics.

7.2 Model Predictions

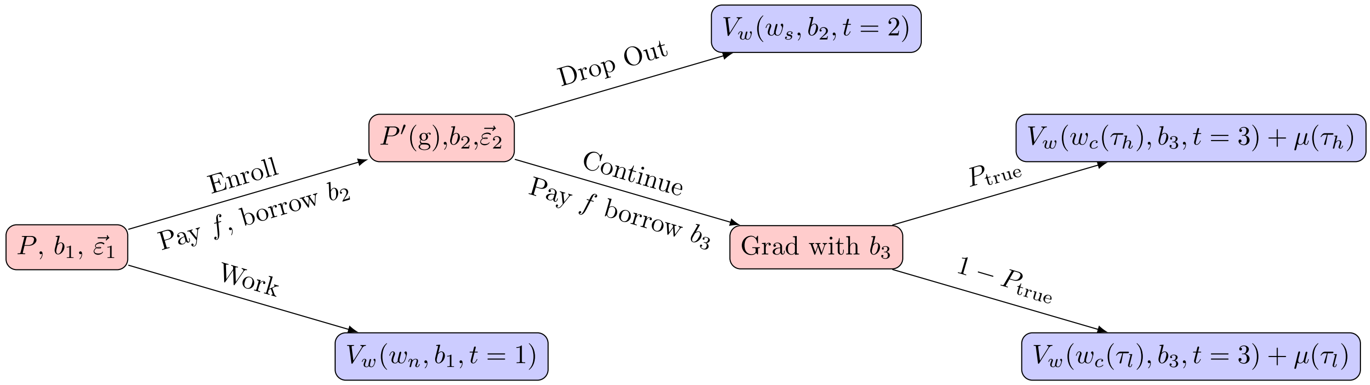

Figure 2: Decision tree representation of the quantitative model. REd nodes represent key stages of the model where decisions are made or different outcomes are realized.

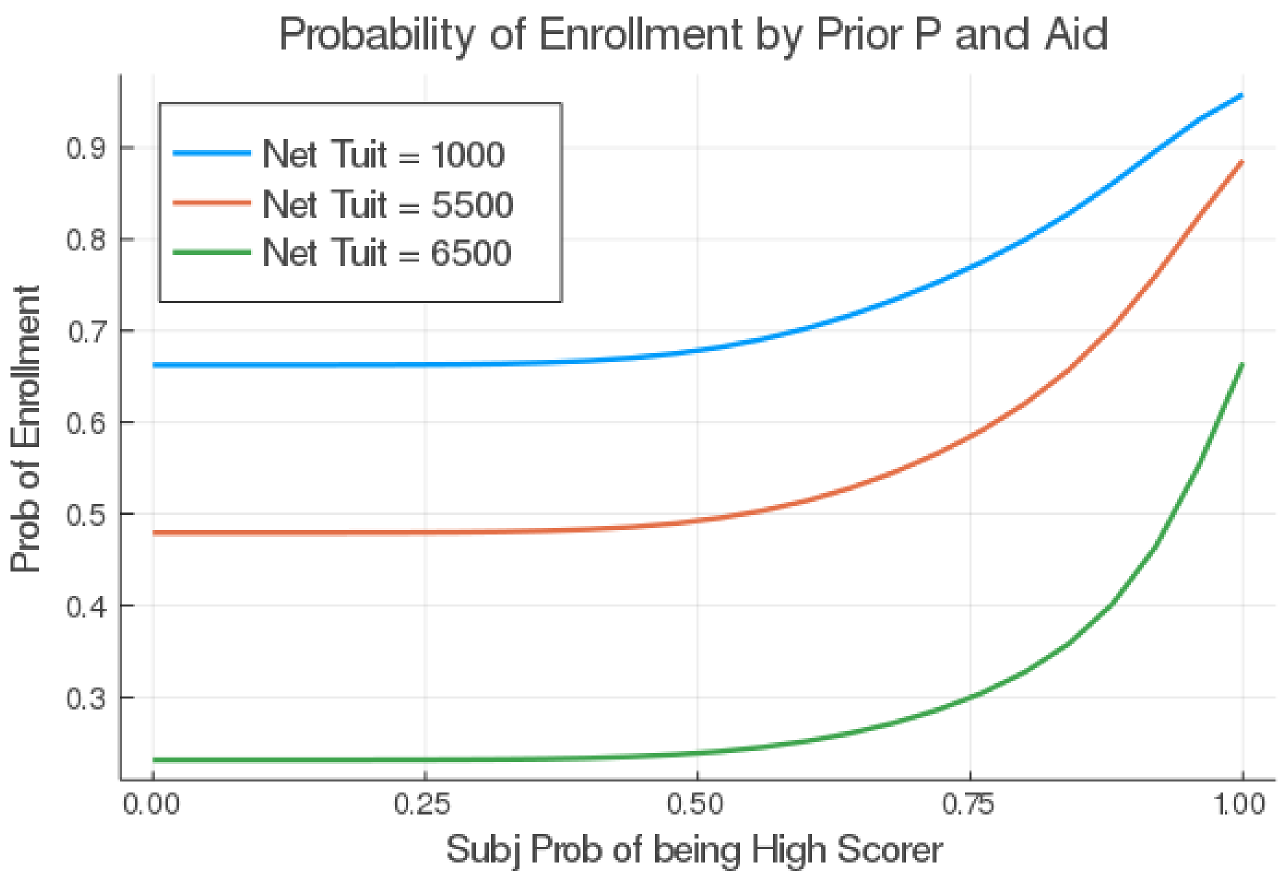

Figure 3: Model predicted probability of college enrollment by Net Tuition and Prior Belief of being High Scorer. Net Tuition is the "sticker price" of tuition minus any financial aid received from family, government or colleges.

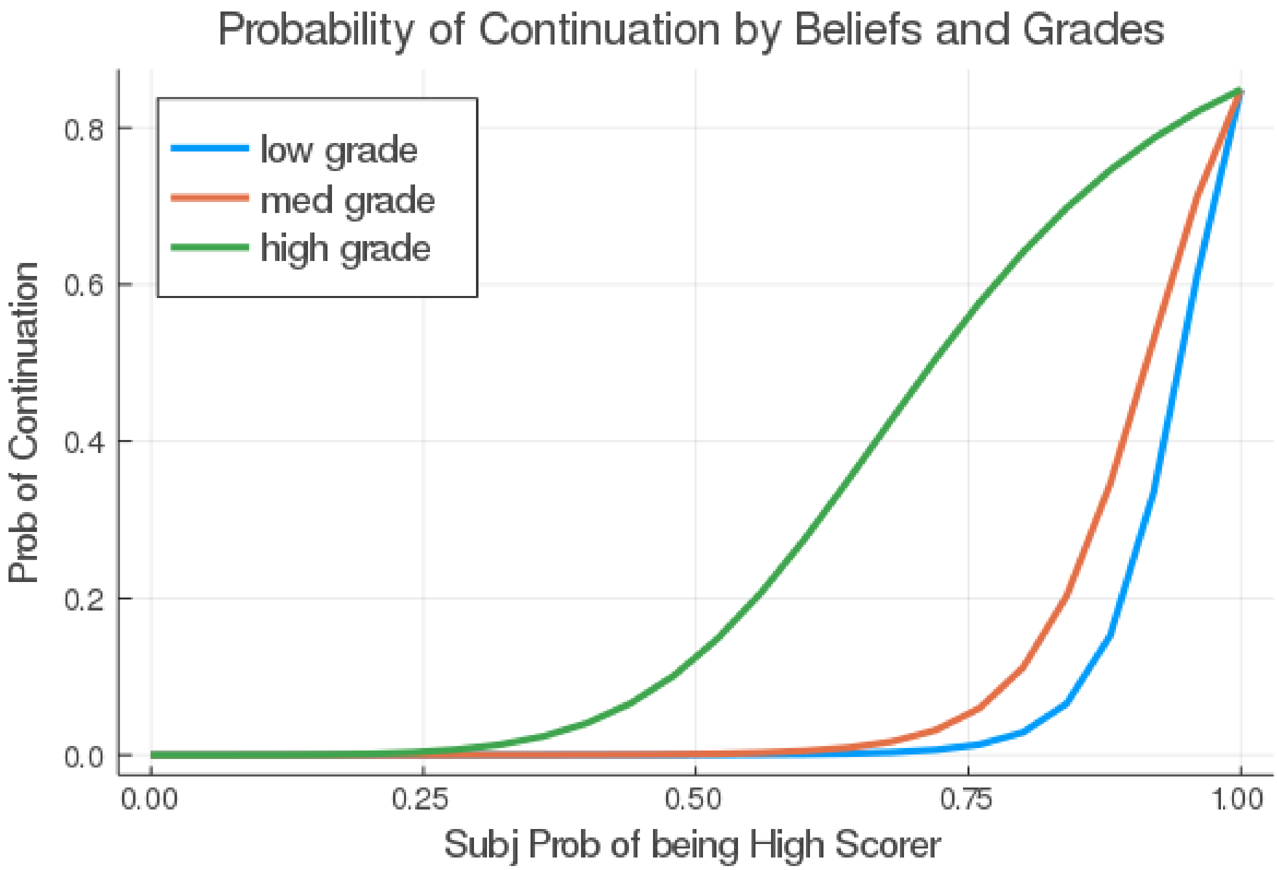

Figure 4: Model predicted probability of College Continuation by average GPA realized before the second stage after the first stage.

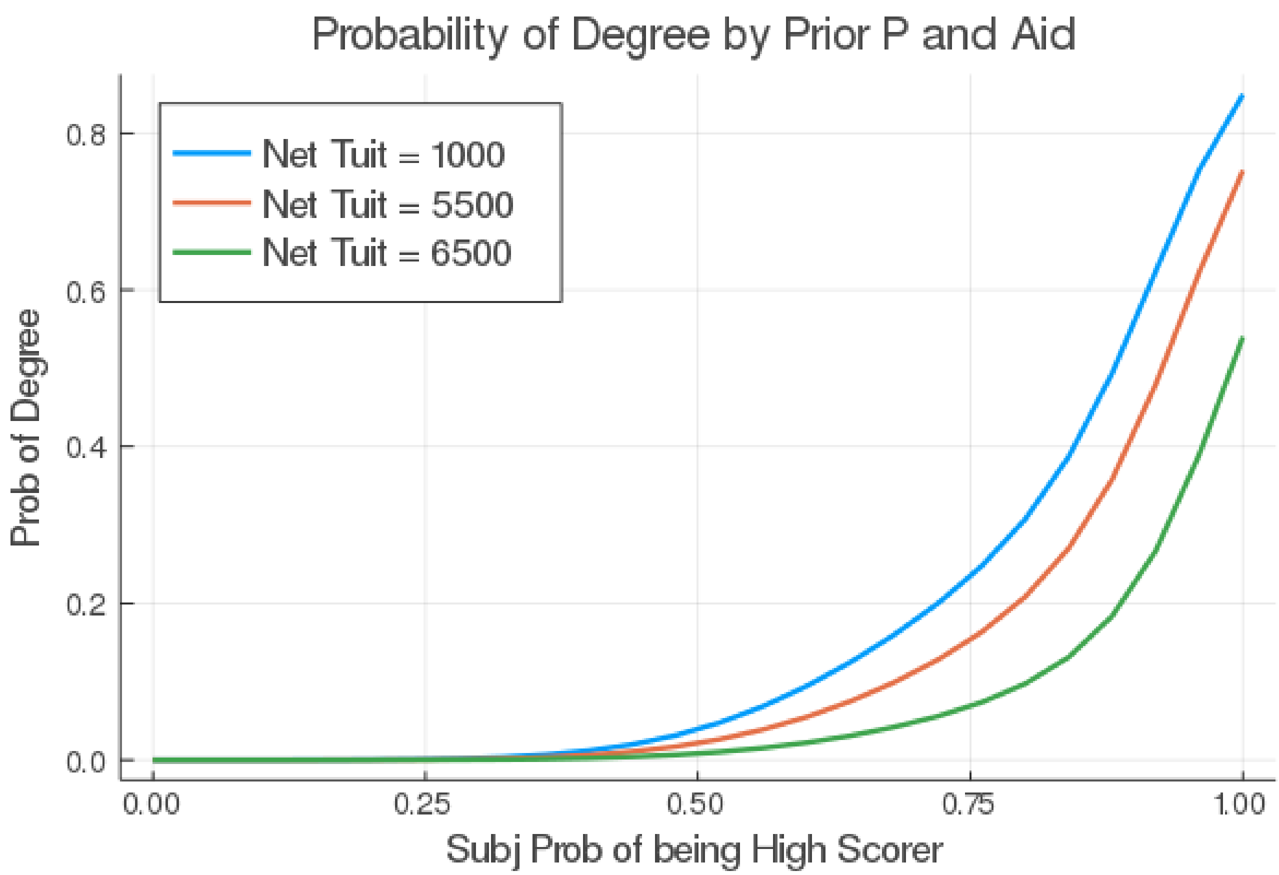

Figure 5: Model predicted probability of Bachelor's attainment, enrollment and completion, by Net Tuition and Prior Belief of being High Scorer. Net Tuition is the "sticker price" of tuition minus any financial aid received from family, government or colleges.

7.3 Quantitative Figures and Tables

Table 3: 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 4: 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
t_1, t_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, completion	Mean by race, ethnicity, parent edu scale parameters taste for taste shocks

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

Table 5: 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.

Table 6: 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)
t_1	Tuition Pd 1	\$7430
		(63.36)
t_2	Tuiton Pd 2	\$6946
		(60.84)

Table 6: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.

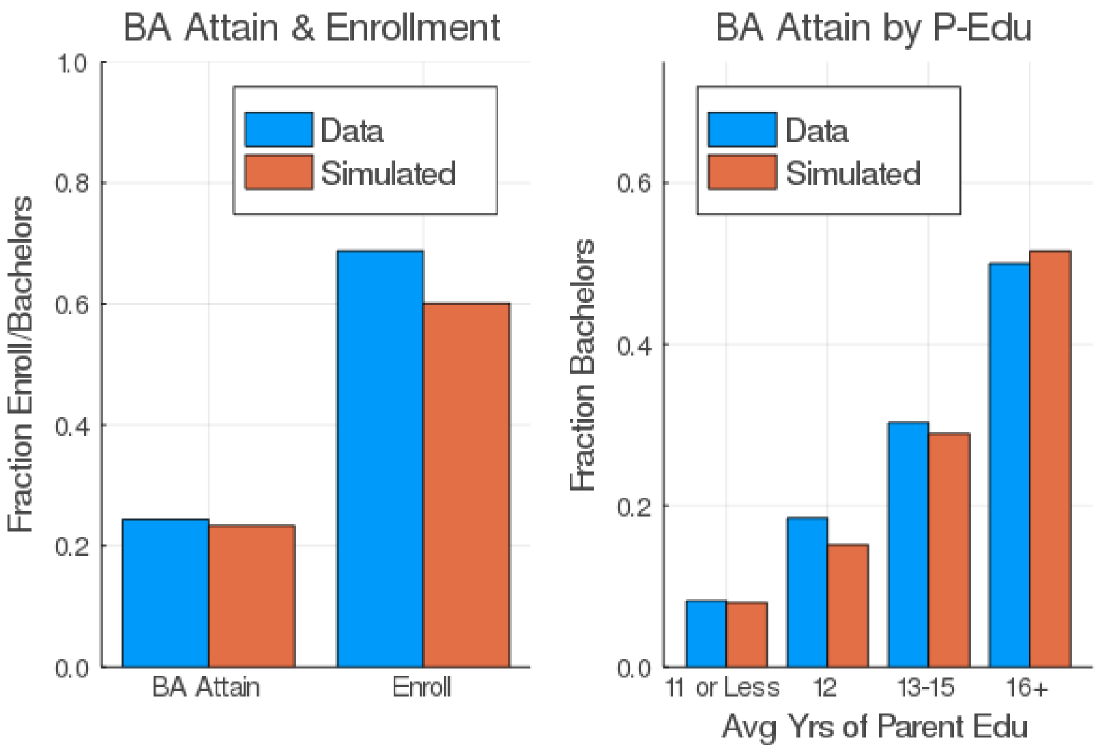

Figure 6: Fit of the Estimated Model: Enrollment, BA attainment, where Blue comes from the NLSY97 and Orange is simulated from the estimated quantitative model.

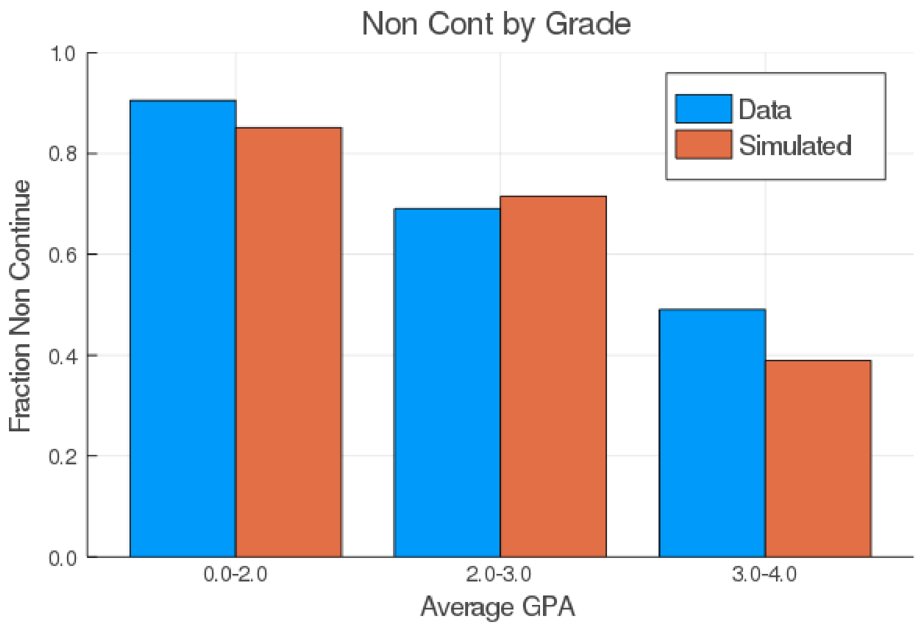

Figure 7: Fit of the Estimated Model: Non Continuation by GPA level, where Blue comes from the NLSY97 and Orange is simulated from the estimated quantitative model.

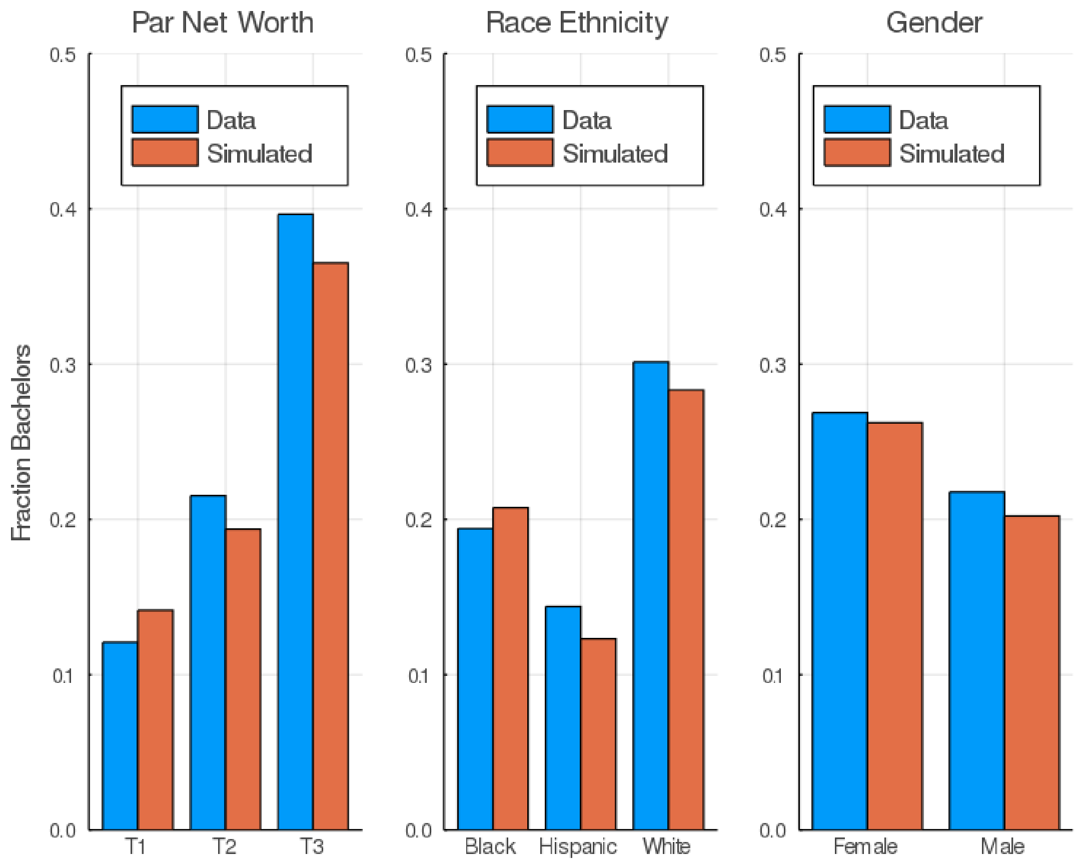

Figure 8: Fit of the Estimated Model: BA attainment by demographics, where Blue comes from the NLSY97 and Orange is simulated from the estimated quantitative model.

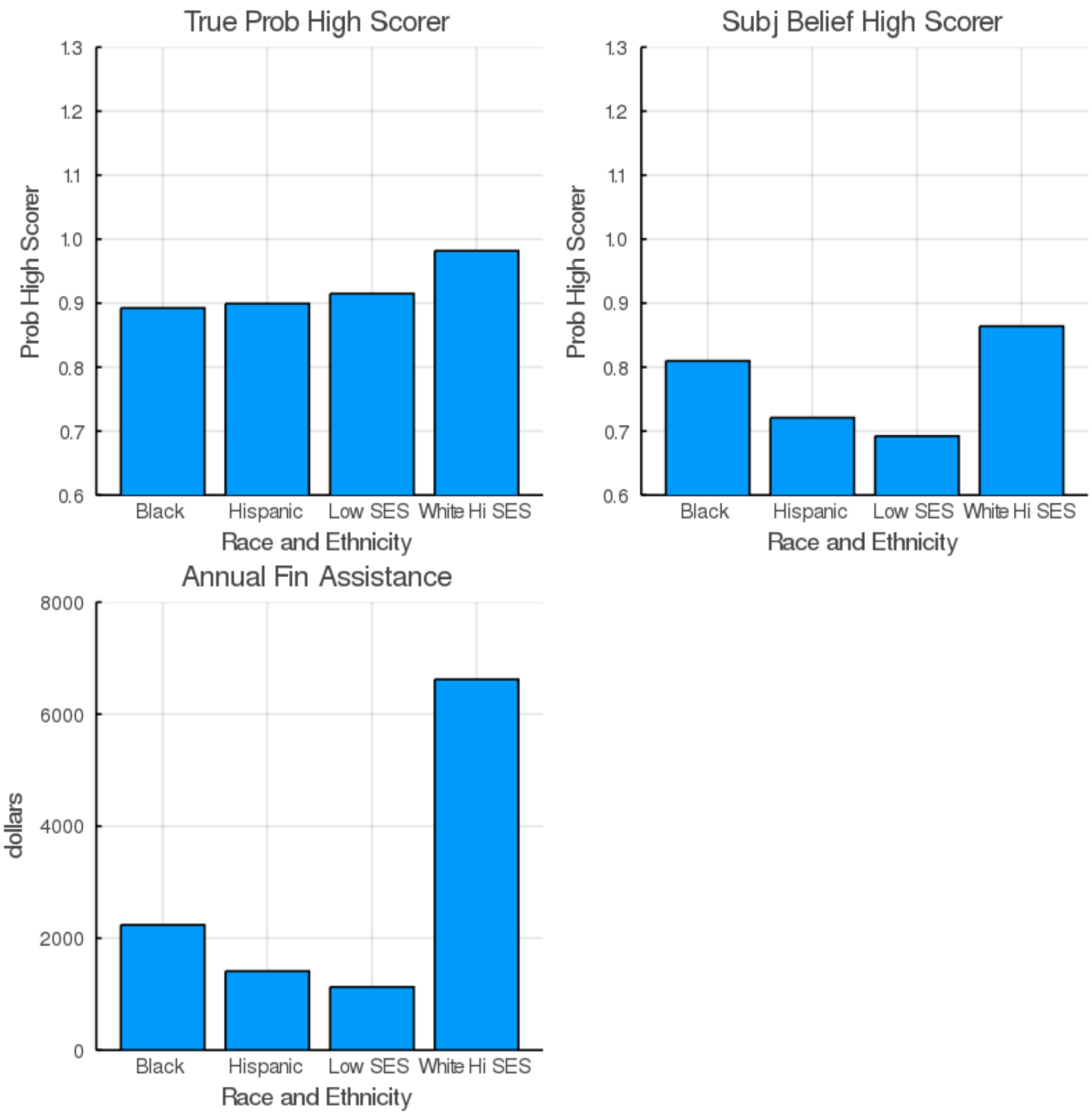

Figure 9: Shows the estimated variables relating to causal mechanism by demographic group, the average true probability of being a High Scorer determined by human capital and demographics, the average subjective belief of being a High Scorer which is a function of measured beliefs in the NLSY97 and parental education, and predicted total financial assistance by demographics which is the sum of family assistance and govt/college aid. Low SES refers to youth whose parents average years of schooling are high school or less or are in the bottom tercile of the household net worth distribution. Hi SES refers to youth in the top tercile of the household net worth distribution and whose parents have an average education of a Bachelors degree or more.

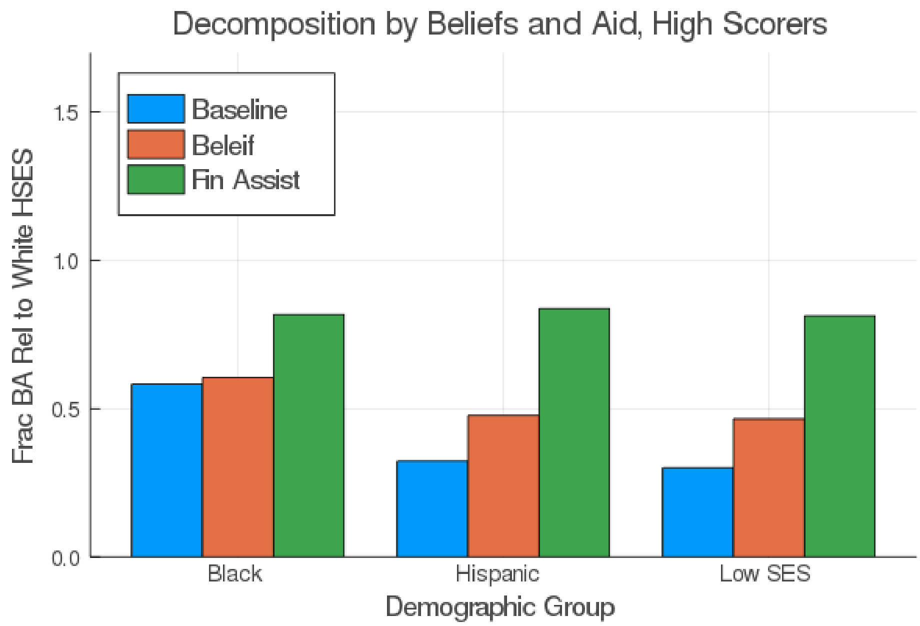

Figure 10: Shows relative BA attainment of Black, Hispanic, Low SES High Scorers relative to White High SES High Scorers after sequentially setting beliefs to mean beliefs of White High SES High Scorers Scorers and Financial Assistance to mean Financial Assistance of White High SES High Scorers. Low SES refers to youth whose parents average years of schooling are high school or less or are in the bottom tercile of the household net worth distribution. Hi SES refers to youth in the top tercile of the household net worth distribution and whose parents have an average education of a Bachelors degree or more.

Table 7: Mechanism Decomposition

Demographic	Total Gap	% Beliefs	%Fin Assis	% Other
Black	41.7 %	5.1 %	50.9 %	44.0 %
Hispanic	67.6 %	22.6 %	53.4 %	24.0 %
Low SES	69.9 %	20.9 %	51.3 %	27.8 %

Table 7: Shows the percentage of the gap relative White High SES High Scorers explained by each mechanism for each demographic group. Decomposition is performed by sequentially setting beliefs, and financial assistance to the mean of White High SES High Scorers for all High Scorers.

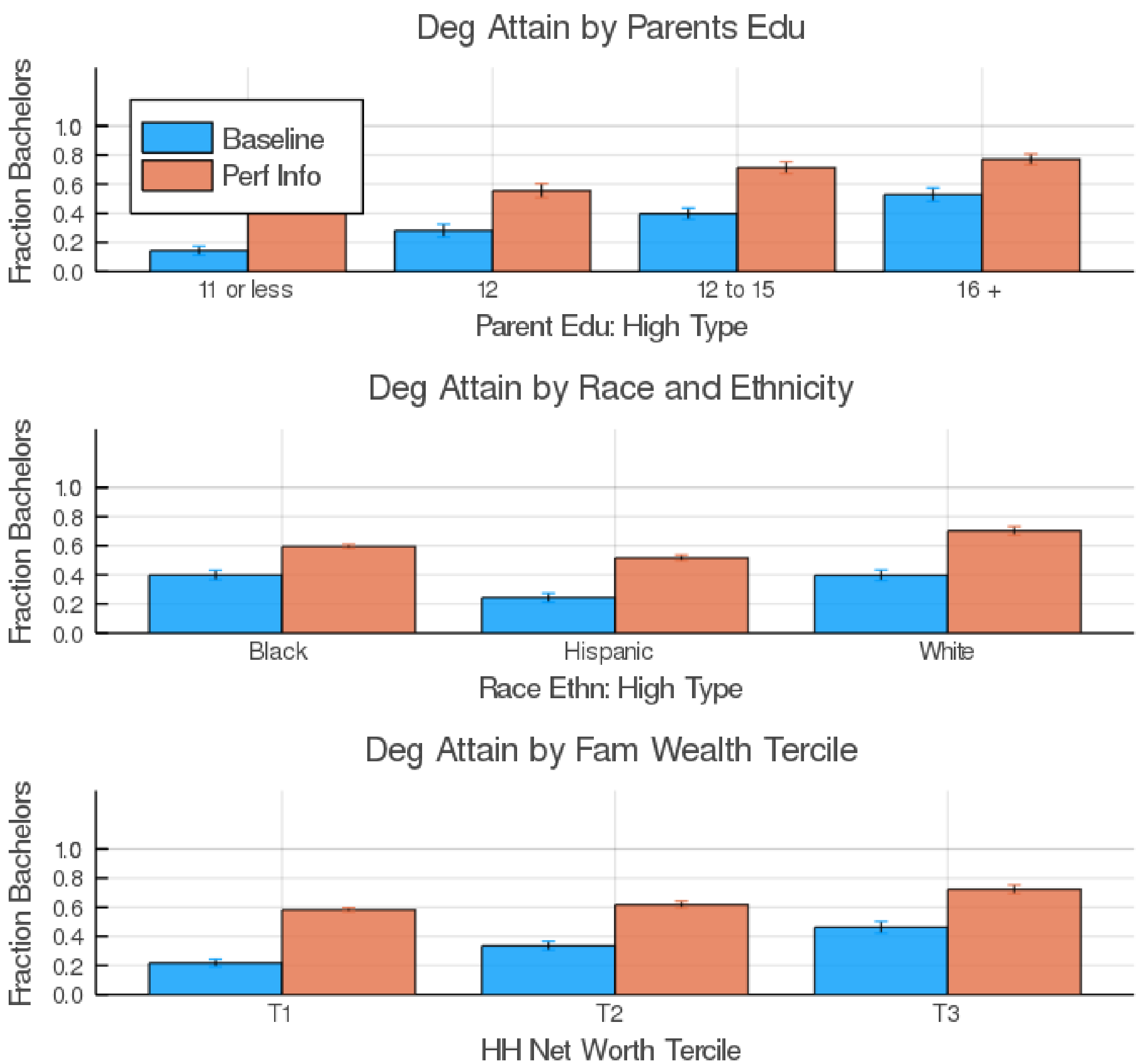

Figure 11: Shows difference in BA attainment under baseline model and under scenario where youth know their true probability of being a High Scorer. This graph looks at predicted High Scorers by demographic group.

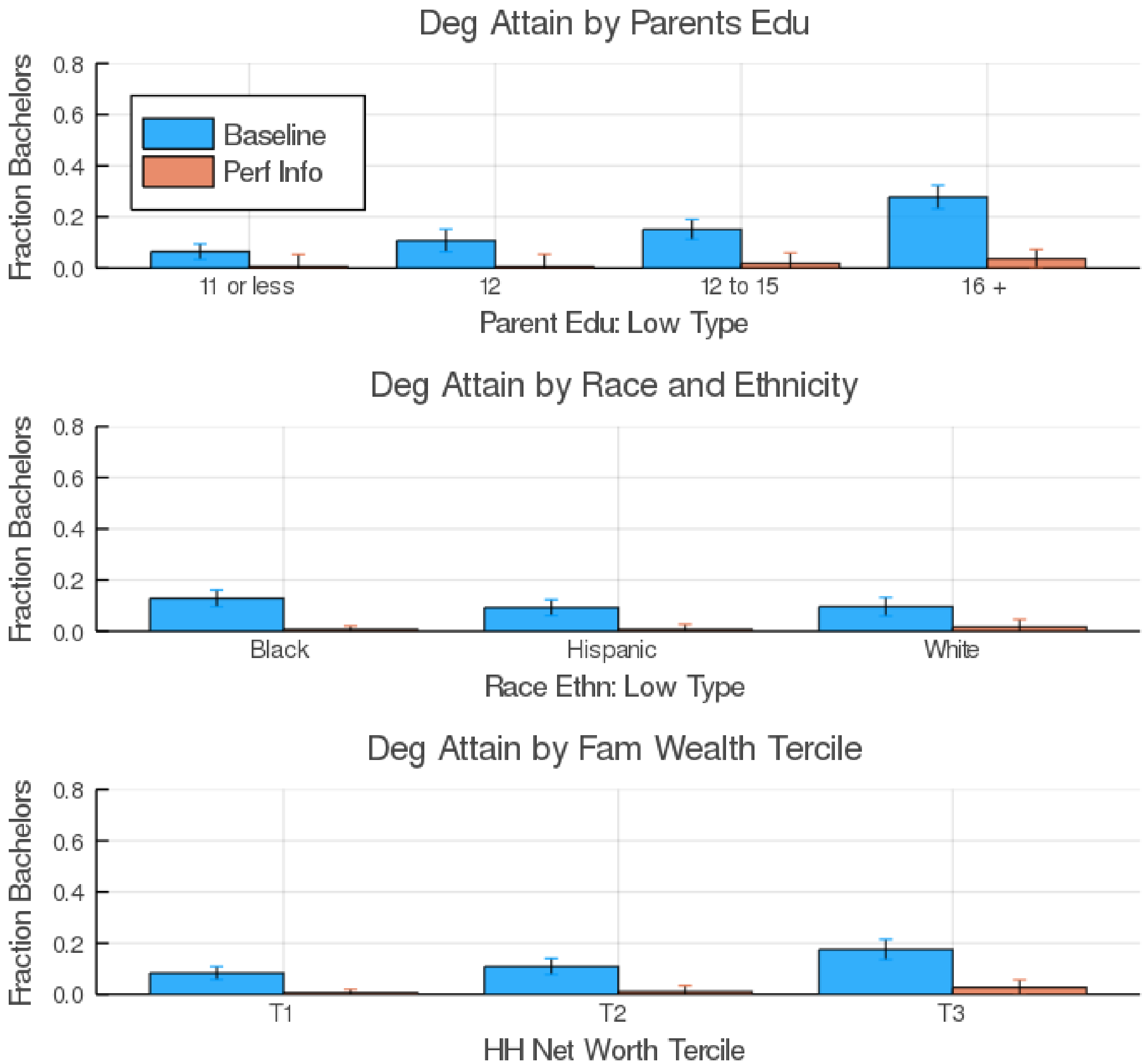

Figure 12: Shows difference in BA attainment under baseline model and under scenario where youth know their true probability of being a High Scorer. This graph looks at predicted Lower Scorers by demographic group.

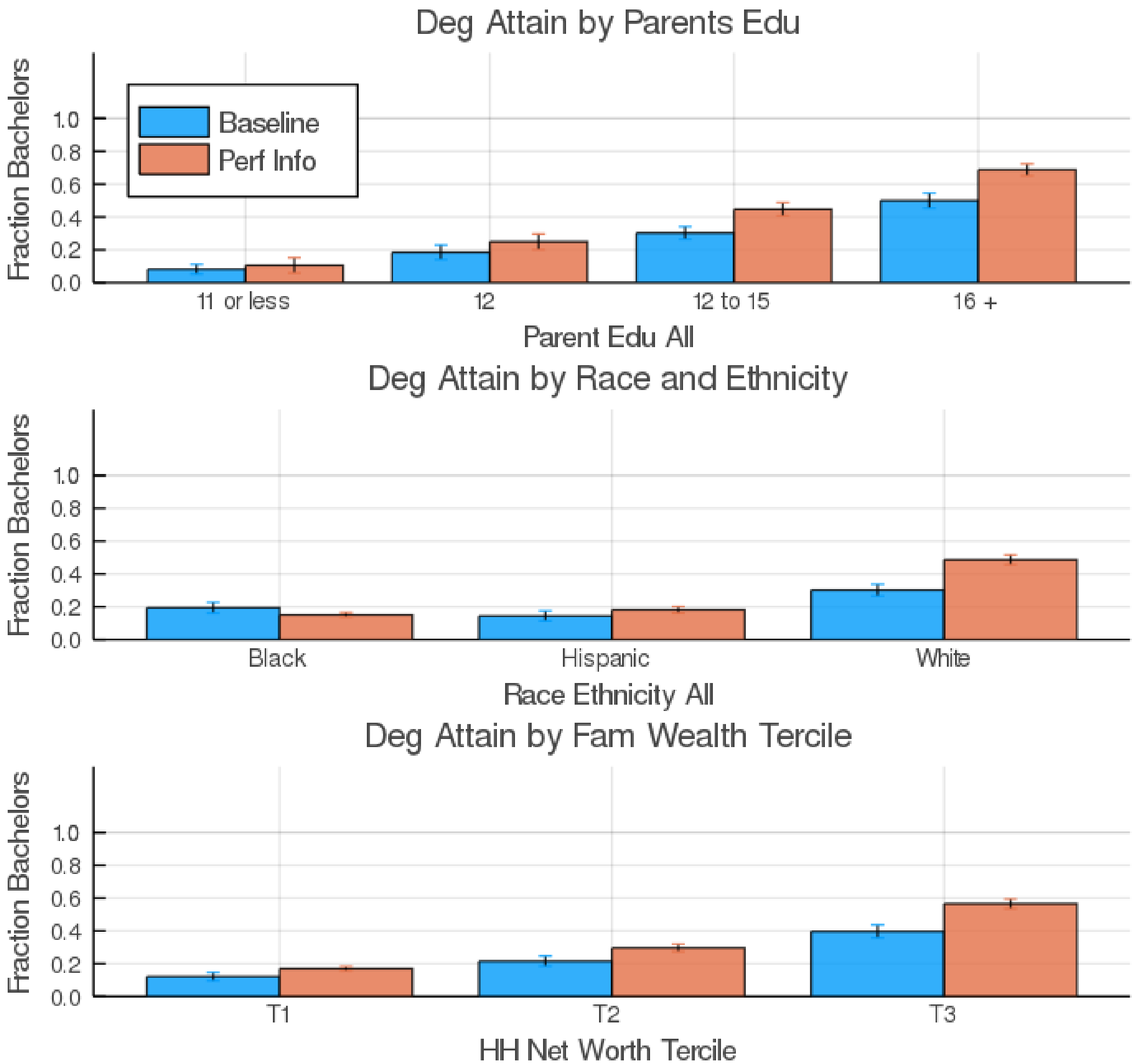

Figure 13: Shows difference in BA attainment under baseline model and under scenario where youth know their true probability of being a High Scorer. This graph looks at all youth regardless of scoring type by demographic group.

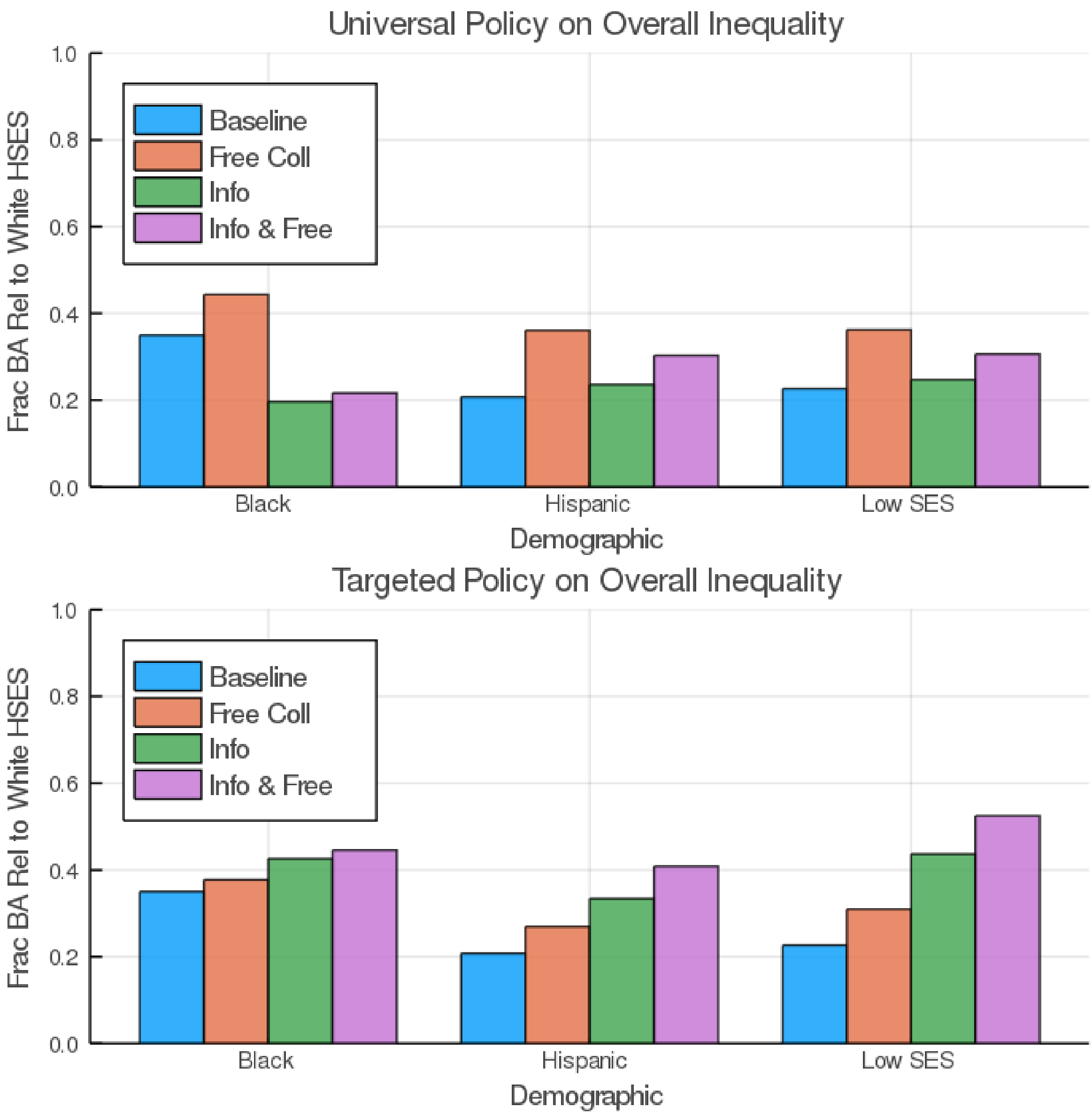

Figure 14: Shows relative BA attainment of Black, Hispanic, and Low SES High Scorers relative to White High SES High Scorers under different policy settings. Low SES refers to youth whose parents average years of schooling are high school or less or are in the bottom tercile of the household net worth distribution. Hi SES refers to youth in the top tercile of the household net worth distribution and whose parents have an average education of a Bachelors degree or more.

Table 8: Mismatch: Percentage of Population Switch with Type Knowledge

Policy	% Pop Mismatched Overall	% Pop Mismatched High Scorer	% Pop Mismatched Lower Scorer
Baseline	28 %	22 %	6 %
Free College For All	31%	21 %	10 %
Targeted College	29%	22 %	7%
10 % improvement in Beliefs	19%	17 %	1%
Targeted 10 % improvement in Beliefs	25%	19 %	6%

Table 8: Shows the percentage of the population in the simulations that are mismatched, that is 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 Low 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 Econornics, 106, 979-1014.

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.

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.

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 9: 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 10: 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 11: 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 12: 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

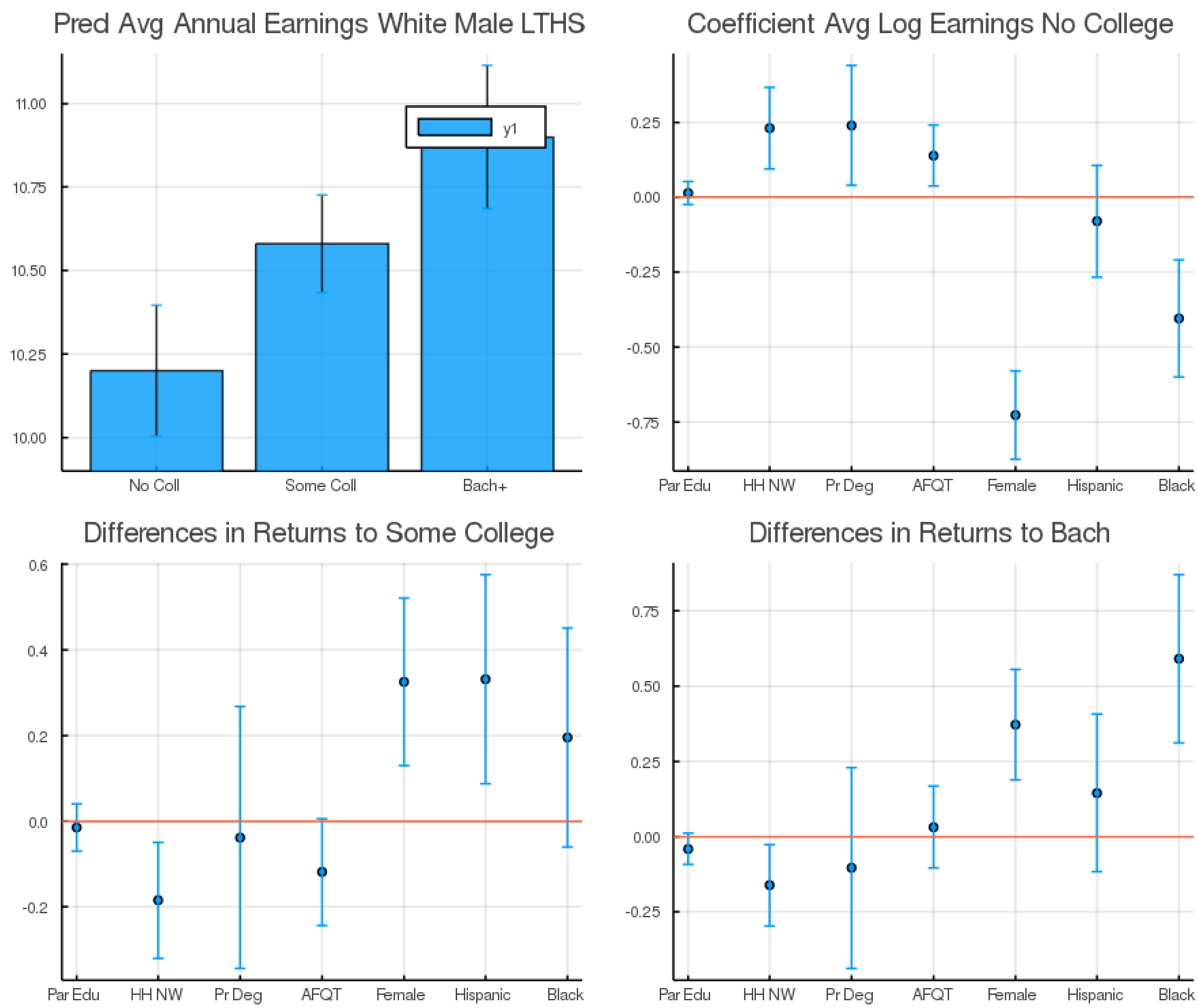

Figure 15: Simulated Update Rule

A.2 Likelihood Function

Specifically the likelihood function is given by the product of the likelihood of observing human capital measures given in RHS of the first two lines of (a.1) below. ϕ and Φ denote the pdf and CDF of the standard normal distribution after subtracting the mean and dividing by the standard deviation of the error term. The third line through fifth line are the conditional likelihoods by schooling level for wages, and grades dependent on enrolling.

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

$$\Pi_{j_n}\Phi(Z^n_{i,j_n};\tau_k,X_i)^{1(Z^*_{i,j_n})}\times(1-\Phi(Z^n_{i,j_n};\tau_k,X_i))^{1-1(Z^*_{i,j_n})}$$
$$\times[\text{Prob}(s<12|X_i))\phi(\ln w_{i,s};X_i)]^{1(s<12)}$$
$$\times[\text{Prob}(s\in(12,16)|X_i)\Pi_{g_k}\pi(g_k|\tau)\phi(\ln w_{i,s};X_i)]^{1(s\in(12,16))}$$
$$\times[\text{Prob}(s\geq 16)|X_i)\Pi_{g_k}\pi(g_k|\tau)\phi(\ln w_{i,s};X_i)]^{1(s\geq 16)}$$

A.3 Indirect Inference: Targeted vs Simulated Moments

Table 13: Indirect Inference OLS Targets

VARIABLES	(1) Enrolled Data	(2) Enrolled Sim	(3) Continue Data	(4) Continue Sim
Intercept	0.303	0.261	-0.068	-0.012
	(0.0307)	(0.069)	(0.0502)	(0.036)
High Belief	0.223	0.200		
	(0.019)	(0.027)		
Fin Assist T2	0.150	0.141	0.072	0.056
	(0.024)	(0.024)	(0.034)	(0.008)
Fin Assist T3	0.286	0.247	0.095	0.083
	(0.030)	(0.036)	(0.0403)	(0.0134)
Parent HSD			0.0767	0.064
			(0.0390)	(0.027)
Parent SCOL			0.128	0.134
			(0.0379)	(0.025)
Parent Bach	0.070	0.023	0.216	0.235
	(0.031)	(0.015)	(0.0478)	(0.027)
White	0.150	0.112	0.015	0.025
	(0.0257)	(0.037)	(0.036)	(0.022)
Hispanic	0.104	0.052	-0.016	-0.001
	(0.032)	(0.048)	(0.044)	(0.024)
GPA Med			0.214	0.154
			(0.0348)	(0.015)
GPA High			0.3724	0.412
			(0.0371)	(0.026)

Table 13: 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.

A.4 More Counterfactuals

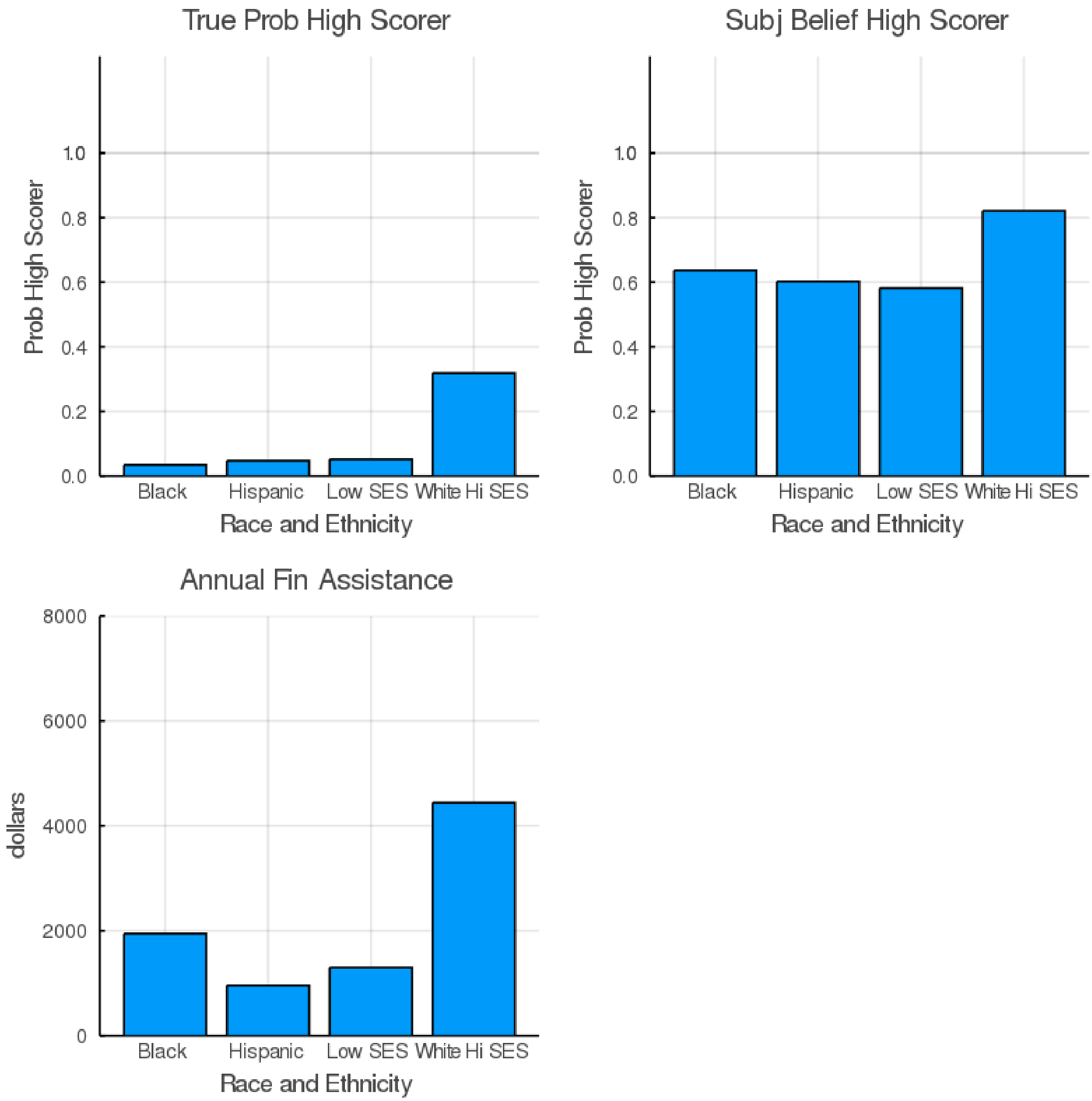

Figure 16: Financial Assistance by Demographic