Is College Worth It For Me?

Beliefs, Access to Funding, and Inequality in Higher Education

Sergio Ernesto Barrera

University of Minnesota

November 2021

Abstract

In the US, the bachelor's attainment rate of high socioeconomic status White youth is much higher than that 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, is targeting 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 the distribution of subjective prior beliefs about own ability using self reported beliefs about college outcomes from the NLSY97 and a dynamic discrete choice model with heterogeneous financial support and beliefs about one's ability. I find that for Hispanic and low socioeconomic status youth, differences in beliefs explain 38-49% of the gap relative to high-socioeconomic status White high-scorers. In contrast, for Black high-scorers beliefs play almost no statistically significant role in explaining gaps. In the policy analysis I show that the targeted policy is the most efficient at closing gaps by 25% to 42%, depending on the comparison group. Although targeting information and funding to low socioeconomic status high scorers narrows gaps, inequality will persist due to differences in early human capital.

1 Introduction

In the United States there are still large gaps in bachelor's attainment by race, ethnicity, and socioeconomic status (SES). 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). Empirical evidence suggest that an explanation for this is that information frictions lead to underinvestment in education for high ability youth in families which adults have less college experience (Hoxby and Turner 2013). Specifically, because the whole college experience might be less familiar to these youth, they may have less information about their own college ability and expected returns than their more affluent peers with similar measures of academic ability.

This paper focuses on differences in bachelor's attainment rates by demographic group. First, I focus on youth with high measures of academic ability and good behavior, which will be referred to as "High-Scorers," and I ask, how do differences in beliefs about own college ability affect inequality in bachelor's attainment rates for high-scorers. The measure of inequality that I use is the difference in bachelor's attainment rates between a given demographic group of interest and high-SES White youth. I focus on three demographic groups low-SES youth regardless of race, and Black and Hispanic youth, regardless of SES. This question is important to answer because if information frictions lead to underinvestment for high scorers from disadvantaged backgrounds than this could imply serious economic costs, such as foregone earnings or growth (Hsieh, Hurst, Jones, and Klenow 2020).

I find that differences in beliefs explain 38 percent of the low-SES high-scorer gap, and 49 percent of the Hispanic high-scorer gap. In contrast, differences in beliefs explains 33 percent of the Black high-scorer gap. However, I am unable reject a null hypothesis of a zero effect of beliefs on the Black high-scorer gap. Additionally, I find that for all three comparison groups differences in financial assistance play big statistically significant roles in explaining gaps, where explained contributions range between 45-50 percent depending on demographic group.

I then turn to studying policy interventions designed to narrow overall inequality regardless of ability. In particular, are policies that target low-SES high-scorers with information and funding more efficient at decreasing overall inequality than policies that are universally applied to all, like free college for all or a tracking policy? I use the same measure of inequality as in the decomposition exercise, but this time not restricting the analysis to high scorers. The efficiency measure that I use to evaluate the policies is college mismatch, which is the percentage of youth who would make different bachelor's obtainment decisions should they have complete knowledge about ability. This takes the form of over-investment of low-scoring youth, and under-investment of high-scoring youth in education. The more targeted policy can be likened to college recruiting efforts like those studied in Dynarski, Michelmore, Libassi, and Owen 2019 or information campaigns as in Hoxby and Turner 2013.

I find that the targeted policy (providing information and funding only to low-SES high scorers), is the most efficient policy at reducing inequality, because it most effectively narrows bachelor's attainment gaps. The reductions range between 25-42% depending on demographic group[1]. The targeted policy also decreases mismatch by encouraging more education investment from high-scorers and thus overcoming underinvestment in higher education. Free college for all decreases inequality at the cost of increased mismatch, because bachelor's attainment rates increase primarily among low-scorers. This happens because low scorers are too optimistic relative to the their actual ability. Although the tracking system reduces mismatch, it increases inequality, because increases in bachelor's attainment among high-scorers are offset by decreases in bachelor's attainment from low-scorers for the groups of interest.

[1]It reduces the Black gap by 25%, Hispanic gap by 28.3%, and the low-SES gap by 41.8%

My analysis is based on a dynamic discrete choice model with credit constraints, heterogeneous financial support, and heterogeneous beliefs about ability. The model includes two latent ability types, low and high-scorers, as well as learning about type through GPA. To estimate the model, I use the National Longitudinal Study of Youth 1997, which contains information on earnings, education outcomes, self-reported beliefs, financial assistance, and demographic information for youth born in the early 1980s. The data set is particularly useful for this exercise because it over samples Black and Hispanic youth.

The crucial objects of interest are the proportion of high-scorers by demographic group, and the distribution of beliefs about one's ability type. I estimate the proportion of high scorers outside of the model by using a finite mixture model with two latent types governing earnings, grades, and human capital measures. I estimate the distribution of beliefs together with the model parameters via indirect inference. The identification comes from two important targets: the coefficient from self reported beliefs about college outcomes regressed on enrollment, and the coefficients from grade categories regressed on college exit, both holding financial assistance and demographics constant[2]. Estimating beliefs this way allows beliefs to not be restricted to be rational priors that match the ability distributions estimated from the data. It also allows the distribution of beliefs to differ by demographic groups.

Overall my results suggest that targeting information and funding to low-SES high scorers can efficiently increase representation in higher education among Black, Hispanic, and low-SES youth. However,substantial inequality remains even after this intervention. This is because differences in early childhood human capital investment are important in explaining differences in bachelor's attainment rates by demographic group.

[2]For an example in the context of occupation choice for how belief paramaters are identified by exit/switching behavior see Papageorgiou and Lopes De-Melo 2016.

1.1 Contribution to the Literature

This paper contributes to the structural modeling literature that focuses on the role of information frictions in higher education decisions. One strand of previous work uses nationally representative panel data to study the role of information frictions in the decision to go to college (Navarro and Zhou 2017; Cunha, Heckman, and Navarro 2005[3]). The second strand of the literature uses panel data from a single university that include subjective beliefs and 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).

My paper is closest to Arcidiacono, Aucejo, Maurel, and Ransom 2016. As in this paper, I bridge the two strands of the literature by examining enrollment, non-continuation, bachelor's attainment, and efficiency in the presence of information frictions. My paper's main innovation is to combine data on self-reported beliefs about higher education outcomes and predicted model behavior to estimate agent's distribution of prior subjective beliefs about one's ability. This contrasts with other papers that set these priors to the estimated ability distribution observed in the data.

Because I allow these beliefs to be heterogeneous by demographic background, there is room for beliefs to be too optimistic or too pessimistic relative to rational expectations. This is what allows me to evaluate the role that differences in beliefs play in generating inequality in higher education outcomes. It also allows me to more accurately measure the effect of providing information on bachelor's attainment rates. This is because under a rational expectations prior this variation may be captured by unobserved preference shocks.

This paper also contributes to the empirical literature documenting that information campaigns can increase enrollment and completion for high achieving students from lower

[3]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.

income backgrounds. (Dynarski, Libassi, Michelmore, Owen 2020; Hoxby and Turner 2013; Bettinger Long, Oreopoulos, and 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 could increase representation in higher education and decrease mismatch across the United States. These policies can be more effective and generate less inefficiencies than free college for all.

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 education outcomes and belief measures, financial assistance, and grades, will be used as moments to identify model parameters.

2.1 Data

The dataset used in this analysis 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 and Hispanic youth in the US. This makes the NLSY97 useful for studying racial and ethnic inequality.

In the analysis that follows I control for cognitive human capital through the Armed Services Vocational Aptitude Battery (ASVAB) Armed Forces Qualifications Test (AFQT), which scores a youth's performance on mathematics and verbal test scores. The units of measurement for AFQT are percentiles. I also control for non-cognitive human capital using indicators for adverse behavior at young ages, sex before age 15, ever committed an act of violence before the start of the survey, and ever stole a value greater than $50 before the start of the survey.

2.2 Empirical Facts

In this section I review some empirical facts in the NLSY97. Summary statistics by parental education and by race are reported in Appendix A.1 under Table 11 and Table 12. The summary statistics in the appendix show that Black, Hispanic, and lower education background youth have low 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 the NLSY97, a good portion of the disparities in schooling outcomes and beliefs for Black and Hispanic youth can be explained by observables. Despite less enrollment, continuation, and optimism on average, in the analysis that follows, being Black or Hispanic is strongly positively correlated with enrollment and optimism towards enrollment. College completion and beliefs about having a degree by age 30 are also strongly positively correlated with being Black, holding all else constant. For Black and Hispanic youth, a large portion of the unconditional gap in schooling and beliefs can be explained by parental education, house hold net worth and human capital measures, with some unexplained portion remaining significant for Hispanics regarding beliefs (see Oaxaca Blinder decompositions in Table 14-16 in Appendix section A.2 and summary statistics Tables 11-12 in Appendix A.1.).

Table 1 studies whether self-reported beliefs are positively correlated with schooling outcomes. It shows that parental education and self reported beliefs about degree attainment are highly correlated with college enrollment, college continuation, and hence bachelor's degree attainment. Additionally, there is a strong role for parental education as well. Table

Table 1: College Outcomes

VARIABLES	(1) Ever Enrolled	(2) Bachelors Attained	(3) Complete College
Parent Education	0.0292***	0.0375***	0.0427***
	(0.0048)	(0.0056)	(0.0070)
Household 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)
Probability 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 the results of three linear probability models. Column (1) and (2) use the whole sample to regress Ever Enrolled and Bachelor's Attained on covariates in the table. Column (3) conditions on those we've observed ever enrolling in college to regress college completion on covariates. Of interest is that the subjective belief variable Prob Degr remains significant and positive after controlling for measures of human capital, access to funding, and parental background

1 shows that holding all else constant if my parent's average years of schooling increases by one year, the probability I attain a bachelor's degree increases by 3.75 percentage points. Additionally if my self reported belief of having a degree by age 30 increases by 0.2, that probability of bachelor's attainment increases by nearly 4 percentage points. This increase in one year of parents education or 0.2 of my self reported beliefs, holding all else constant, is equivalent the equivalent to a nearly $200,000 in household net worth, holding beliefs and parental education constant. The strong relationship between beliefs, parental education, and schooling outcomes continues to hold in Column 3 even with the inclusion of college GPA, and financial resources available in college. This is in addition to controlling for human capital measures, birth year, and geography variables.

Given the strong correlation between beliefs and schooling outcomes, Table 2 examines which covariates are strongly associated with beliefs. [4]. Table 2 shows that parental education and household net worth holding all else constant are associated with more optimism. An additional year in parents education is associated with a nearly 3 percentage point increase in the self reported probability of enrollment or degree by age 30. Number of peers with college plans is also positively associated with more optimism. If youth went to a high school where more than 90% of high school peers are planning on going to college as opposed to one where less than 10% were, than self reported probability of degree attainment and enrollment would increase by close to 20 percentage points. This is consistent with the findings of Hoxby and Avery 2013 and Hoxby and Turner 2013. Their findings suggests that low-SES youth may know less about suitability for college resulting from less adults in their social networks that have higher education experience and less peers that go to college.

[4] Sample size differences in Table 2, is due to the fact that the probability of degree question was only asked to the older cohort, and probability of enrollment was asked for the younger cohorts while in high school. 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.

Table 2: Measured Beliefs

VARIABLES	(1) Probability Degree	(2) Probability Enroll
Parent Education	0.0267***	0.0282***
	(0.0046)	(0.0058)
Household 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 results from an OLS specification regressing two measures of beliefs on covariates. Of importance is that subjective beliefs reported before age 18 are highly associated with parental education, race and ethnicity, holding net worth and human capital constant.

Although Table 1 establishes a relationship between beliefs and continuation, next in Figure 1 and Table 3, I evaluate the difference of college non-continuation behavior by observables, within grade categories. College non-continuation is measured by students who enroll and do not complete a 4 year degree. This includes those that enroll in community college and exit upon receiving an associates degree[5].

[5] 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.

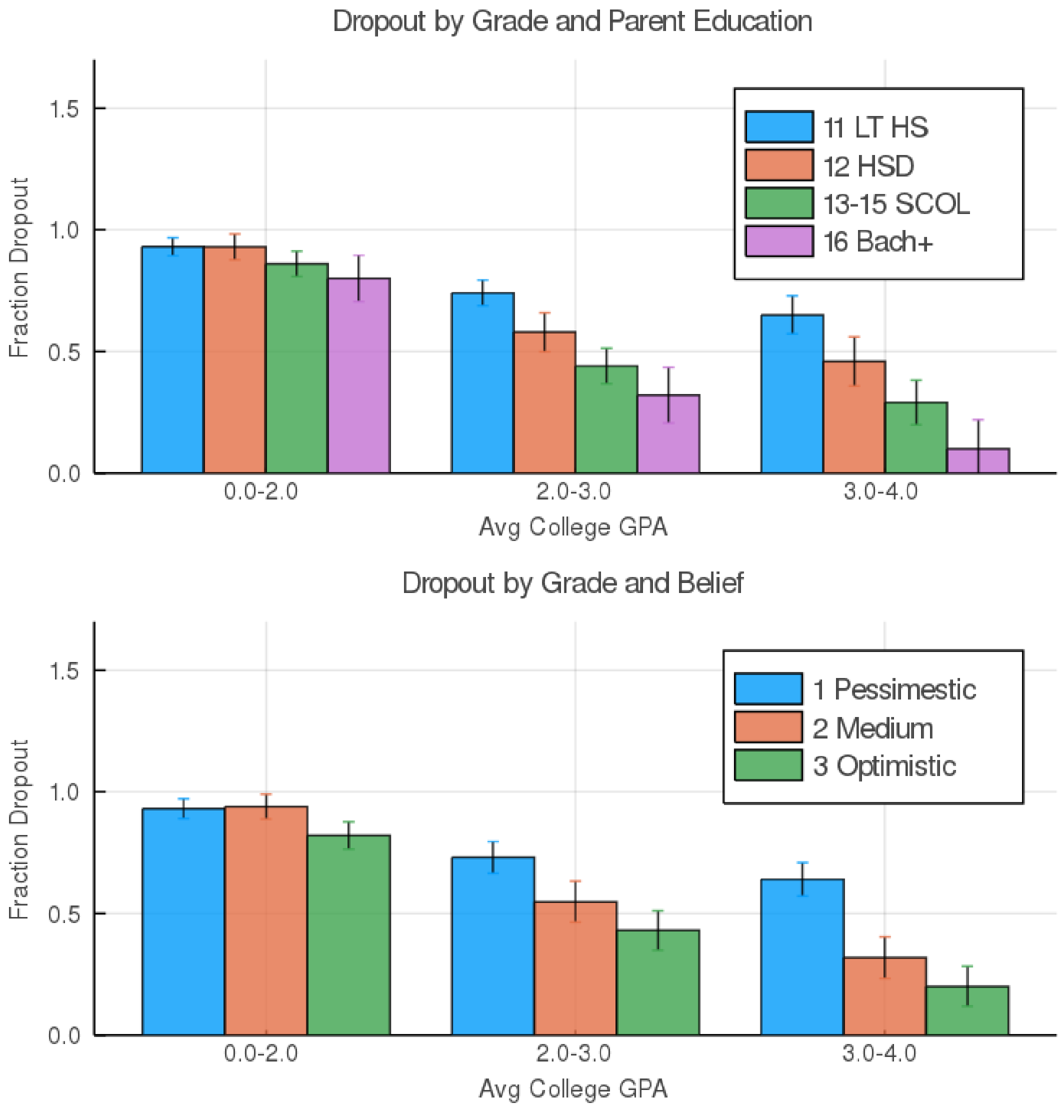

Figure 1: College non-continuation rates conditioned on grades/demographics.

Table 3: Non Continuation Interacted with GPA

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 Education	-0.0003	-0.0077	-0.0317*
	(0.0096)	(0.0161)	(0.0163)
Probability of Degree	0.0677	-0.2748***	-0.2007*
	(0.0544)	(0.1032)	(0.1122)
Household 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

Table 3: Displays the results of an OLS regression of college non continuation on covariates, where GPA categories are interacted with the self reported belief variable Prob Degr, as well as Parent Edu. The second and third column show the results of the interacted coefficients, while the first column displays the stand alone coefficients.

Although non-continuation decreases with grades, Figure 1 shows that this exit behavior within grade categories differs by parental education and belief categories concerning bachelor's attainment[6]. In Table 3 we control for measures of human capital and financial resources and still find statistically significant coefficients for the belief variable interacted with GPA category. Similarly,the two interaction terms between parental education and high grades, and self reported probability of degree and high grades is marginally significant.

The decrease in non-continuation with higher grades, as well as the different effects of grades by belief levels is consistent with the hypothesis that agents don't know their individual returns to college and learn through grades. According to Figure 1 and Table 3 low grades are a strong signal for low returns, and high grades are a strong signal for high returns[7]. As suggested by the belief medium grade coefficient, more optimism might matter more for the medium grades, since the signal from medium grades is more ambiguous.

This means that at least for a prior belief not equal to zero or one, low grades may lead to downward adjustments in college returns and more exit, while high grades may lead to high adjustments in college returns leading to more persistence and continuation. Since medium grades provide a less salient signal, beliefs about returns depend more on the prior and hence the marginal effect of beliefs on estimated returns and drop out is stronger. This is consistent with the belief coefficient estimates in Table 3. Another implication of all of the coefficients on grades and the belief variable in Table 3 is that as an agent becomes more optimistic the difference in exit behavior between high and medium grades narrows. For example at a self reported belief of zero, getting high grades as opposed to medium grades leads to a 9 percentage points increase in probability of completion while at a belief of one to only a 1.6 percentage point increase in probability of completion. This is consistent with Bayesian learnings models were optimistic youth near certain of their type not sufficiently

[6] Beliefs were broken into three categories by setting quantiles for beliefs. 4 quantiles were used but since there was a lot of bunching 1 group contains two quantiles

[7] Given by positive stand alone belief coefficient and the marginally significant stand alone coefficient for high grades

adjusting their estimates downwards with bad signals, and hence exhibiting more persistence. Altogether this analysis suggests a connection between subjective beliefs, parental education 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, which will be explored in the quantitative analysis of the paper.[8]

3 Economic Model

In this 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 and inequality in education outcomes. 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.

The economic environment consists of agents who live $T = 24$ periods, where each period lasts 2 years and represents an age span from 18-66. In each period agents can save or borrow up to a specified borrowing limit. Once an agent begins work, they do not return to school, so the education problem acts as a three stage problem. In the first stage agents decide to enroll or work until the end of the life cycle. If agent's enrolled in the first stage, then in the second stage agents choose to continue with school or work until the end of the life-cycle. Finally if agents chose to continue school, then in the third stage agents work for the remainder of the life cycle and realize type dependent earnings and non pecuniary utility.

In the model agent i has an unknown type $\tau_i \in \{\tau_h, \tau_l\}$ for being a high-scorer or

[8] 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.

low-scorer, respectively.The latent ability type τ_i determines the distribution of grades while in college $\pi(g_k, \tau_i)$ for $k \in \{l, m, h\}$, for low, medium and high grades respectively. The latent ability type τ_i also determines earnings $w_c(\tau_i)$, and non-pecuniary utility, $\mu_c(\tau_i)$ in the final stage of the model. The two variables for τ_i are hence a parsimonious one dimensional representation of the important role that cognitive and non cognitive skills play in generating education outcomes. Because of this τ_h represents the latent ability value of high-scorers since if an agent has τ_h they have a higher probability of achieving higher grades, higher earnings, and non pecuniary utility from college.

The realization of τ_i depends on true probability $P_{\text{true},i}$ of being type τ_h. This true probability depends on parental education, household net worth, race, ethnicity, sex, and measures of human capital. As such, it also captures the effect of early childhood human capital investment on educational attainment and earnings.

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. This subjective belief captures a broad belief about success at college for the individual, since and individual with $\tau_i = \tau_l$ is more likely to have lower grades, lower earnings, and less utility from graduating college.

3.1 Timeline of the problem

A decision tree representation of the problem is shown in Figure 2. In the decision tree subscripts are suppressed for ease of illustration. In the first stage 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 working and earning non college earnings w_n for the rest of the life cycle, or enrolling in college where they pay net tuition, $f_{1,i} = tuit_1 - Aid_{GC,i} - Aid_{Fam,i}$, which is the sticker price net of financial aid from government/colleges and families for the first period of life.

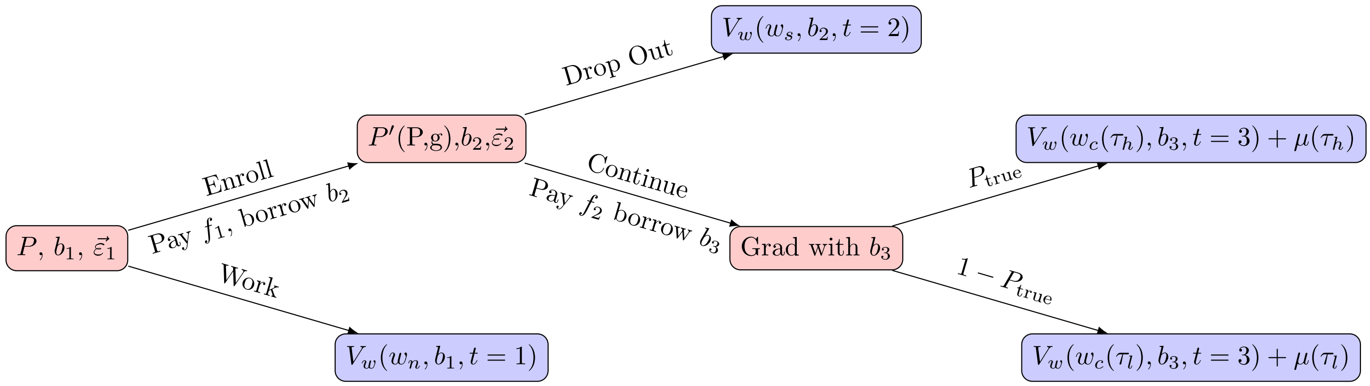

Figure 2: Decision tree representation of the quantitative model. Red nodes represent key stages of the model where decisions are made or different outcomes are realized.

In the second stage agents realize a signal for their latent type given by the GPA g_i during 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 period 2 net tuition $f_{2,i} = tuit_2 - Aid_{GC,i} - Aid_{Fam,i}$ for another period, or exit to work for the rest of their lives and earn w_s each period.

If agents complete school then after they work and earn earnings that depends on their type, $w_c(\tau_i)$ each year and receive non monetary utility $\mu(\tau_i)$ from work. Agents make borrowing and 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 $-B_s(t) \geq -B_n(w)$. Hence credit constraints are tighter while enrolled in school (Lochner and Monge-Naranjo 2012).

Heterogeneity by parental background, race, and ethnicity enters the problem through four channels. The first one is the distribution of initial subjective beliefs P_i of one's own ability type.The second is through transfers from parents, as well as from the government, and colleges that often provide need based financial assistance. This second channel leads to differences in net tuition $f_{t,i}$ for $t = 1, 2$ while in school. The third channel is the true probability of being type τ_h, $P_{\text{true},i}$ which determines the distribution of grade realizations and future earnings. Finally, the fourth channel is the distribution of non-pecuniary utility shocks $\vec{\varepsilon}_{t,i}$.

The model has no explicit role for discrimination. However discrimination could enter the problem through past policies that created differences in parental education and net worth. It can enter through the amount of aid the government or colleges distribute. It can enter through wages and grades, and hence $P_{true,i}$. It can also enter through early childhood human capital that determines the realization of human capital measures.

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 one's 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 are unable to consumption smooth and face lower consumption for the first two periods of their life. If youth receive less financial assistance during college, or are more pessimistic about being a high scorer, than this will make college less appealing for students from demographic groups where this is more likely the case.

3.2 Enrollment Stage

At age 18, agents either enroll in school or work. If they choose to work at this stage they will do so until the end of the life cycle. Agents begin with a belief P_i that they are of type τ_h, a net tuition realization $f_{1,i}$, initial assets $b_{1,i}$, and unobserved tastes for college and work $\vec{\varepsilon}_{1,i} = (\varepsilon_{c,1,i}, \varepsilon_{w,1,i})$. The agent's stage 1 problem is thus given by (3) below, where $V_w(w_n, b_{1,i}, 1) + \varepsilon_{w,1,i}$ is the utility from working and $V_{c,1}(P_i, f_{1,i}, b_{1,i}) + \varepsilon_{c,1,i}$ is the utility from enrolling in college.

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

$$s.t.$$

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

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 $\pi(g_k, \tau_j) = Prob(g_k|\tau = \tau_j)$.

$$(4) \quad P'(g_k, P_i) = \frac{P_i \times \pi(g_k, \tau_h)}{P_i \times \pi(g_k, \tau_h) + (1 - P_i) \times \pi(g_k, \tau_l)}$$

3.3 Continuation Stage

At age 20, agents make the decision to continue and complete college or exit and work for the remainder of the life cycle. Agents observe GPA g from the first stage and then update belief P_i to $P'(g, P_i) = P'_i$. They realize period 2 net tuition $f_{2,i}$ and begin the second stage with debt/savings from the first stage $b_{2,i}$. They also 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

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

$$s.t.$$

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

During college grades reveal information about τ_i because the grade distribution depends on τ_i. But because τ_i also determines one's non-pecuniary utility, the information revealed in school can also include psychosocial elements of higher education that are often discussed in other contexts. In this model, the assumption is that this is closely tied to grade performance, and a bad signal in grade 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 are captured by a age-constant location parameter of non-pecuniary shocks $\vec{\varepsilon}_{t,i}$.

3.4 Workers Problem

Finally, the workers problem is given by (1) below. The state variables are earnings w, assets/debt b_i, and age t.

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

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

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

For every period the borrowing constraint is given below.

$$\tilde{B}_{T-n}(w) = \sum_{m=1}^{n} w(1 + r)^{-m} \quad \text{for} n \geq 1$$

$$\tilde{B}_T = 0$$

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

3.5 Optimal Choice

Since in the first two stages the agent faces a discrete choice problem, the optimal decision for each agent is described by a cutoff rule with respect to one's belief about one's type. If P_i is higher than a certain threshold the agent will enroll. For example in the first stage the optimal decision is characterized by equation 6 below, where $\sigma_{d,2}, \mu_{d,2,i}$ are the normalized scale parameter and location parameters[9] 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 $\{\pi(g_k, \tau_j)\}_{k,j}$ 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-pecuniary utility shocks and distribution constant, allow us to predict the effects of financial assistance, subjective beliefs, and human capital embedded in $P_{true,i}$.For instance, if the distribution of non-pecuniary utility is such that utility from the decision to work is less than the utility from college for being τ_h, and greater than the utility from college for being τ_l, then the decision rules are strictly increasing in $f_{1,i}, f_{2,i}$, depending on the period. $P'(g_i, P_i)$ also increases in P_i and for $P_i \in (0, 1)$, if higher grades

[9] Normalized with respect to the difference in Type I extreme values. This because the difference in shocks is what is identified

provide a strong signal of being τ_h, $P'(g_i, P_i)$ will increase in g_i. Therefore probability of enrollment and probability of continuation would increase with financial assistance, one's belief about being τ_h, and with higher $P_{true,i}$.

3.6 Example for Model Prediction

We now turn to illustrating how financial assistance, subjective beliefs of being τ_h, and grades (whose realizations reflect earlier human capital investments) affect one's 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, the probability of enrollment displays a flat portion and then increases with subjective beliefs. This is the case for all net tuition levels displayed in the graph. Thus if two youth have the same beliefs but different access to resources, their probability of enrollment will still be different. Likewise, if their access to resources is the same but beliefs differ, their probability of enrollment also differs.

Figure 4 shows that conditional on enrolling, probability of continuation differs by grade revelation. This illustrates how learning affects one's continuation decision. After a certain belief threshold, higher grades thus lead to an increase in one's probability of continuation. The large difference in continuation probability between high grades and medium grades diminishes as a youth become more optimistic.

Figure 5 takes probability of obtaining a degree 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 one's probability of obtaining a degree.

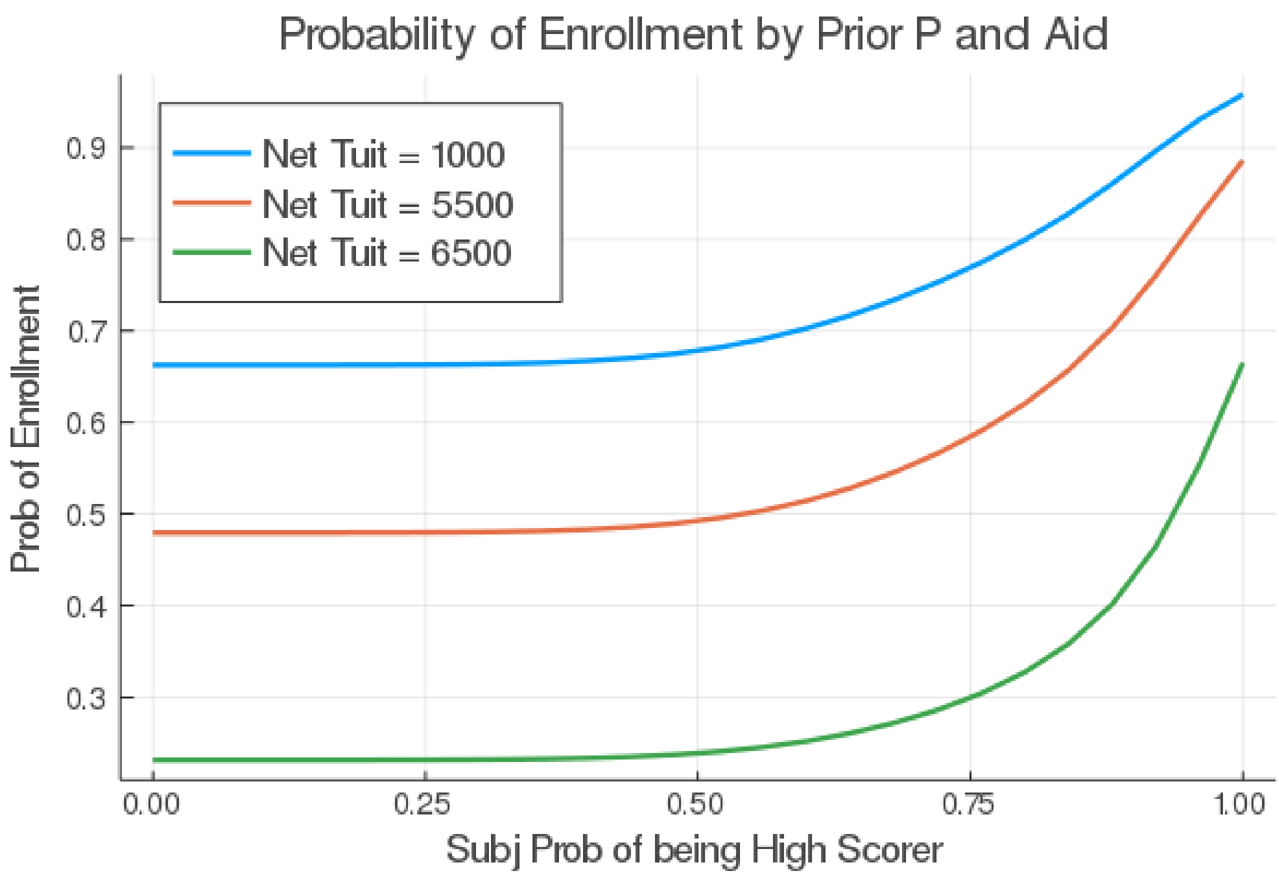

Figure 3: Model predicted probability of college enrollment by net tuition and prior subjective belief of being a high-scorer. Net tuition is the ”sticker price” of tuition minus any financial aid received from family, government or colleges.

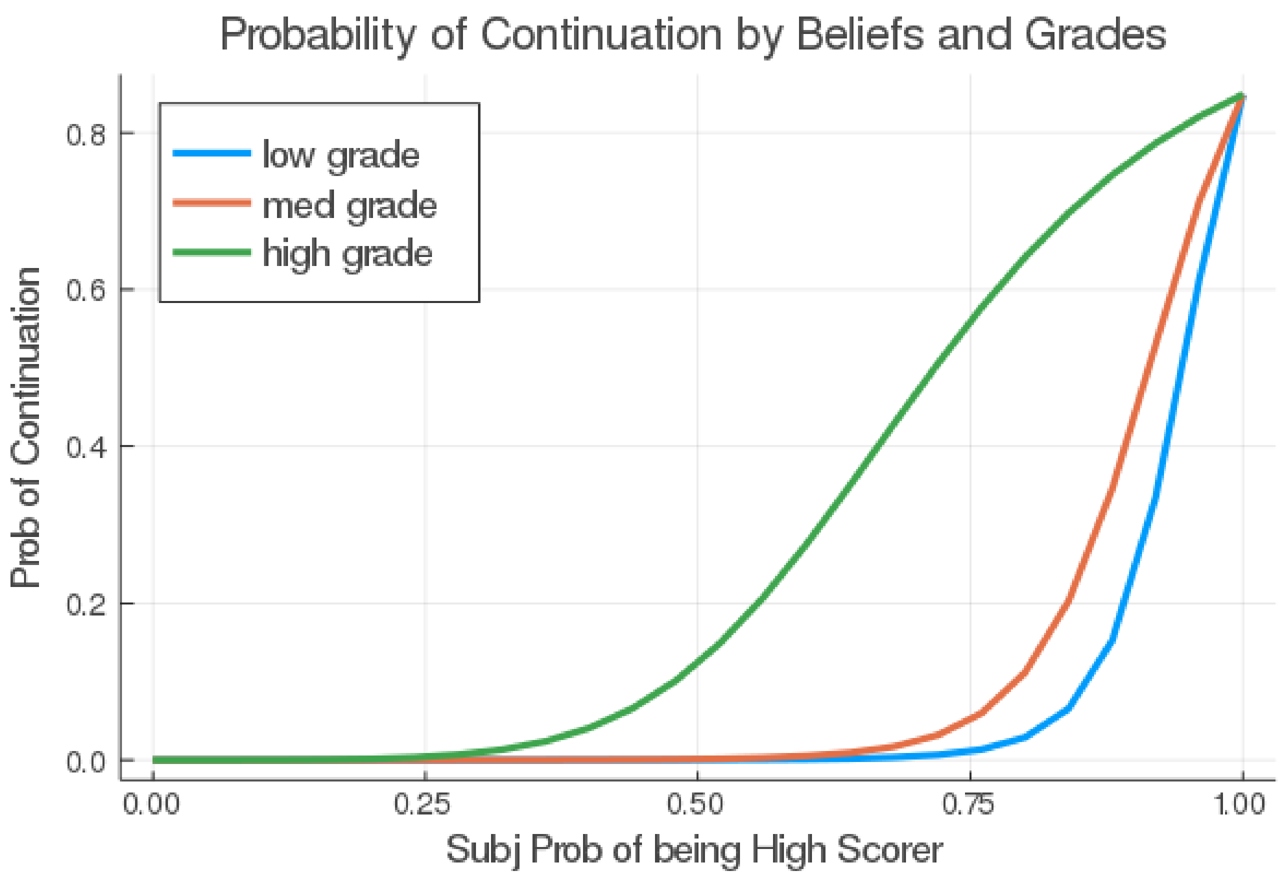

Figure 4: Model predicted probability of college continuation by average GPA realized before the second stage after the first stage.

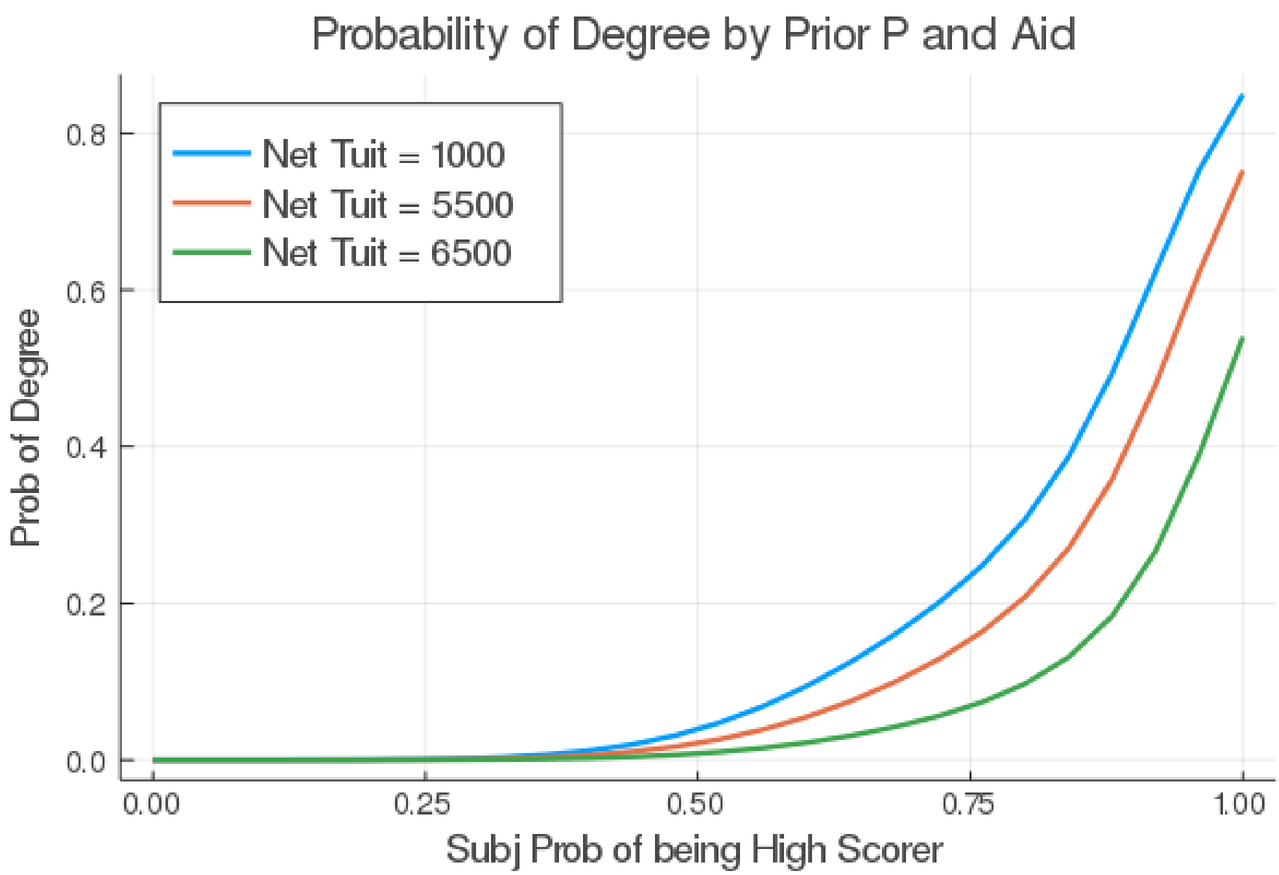

Figure 5: Model predicted probability of bachelor's attainment, which is enrollment and continuation, by net tuition and prior subjective belief of being a high-scorer. Net tuition is the "sticker price" of tuition minus any financial aid received from family, government or colleges.

4 Estimation of Quantitative Model

In this section I discuss how I identify and estimate the parameters of the structural model described in section 3. Specifically, I describe some of the assumptions governing the distribution of earnings, financial assistance, and probability of being a high type, as well as parameters whose values will be set outside of the estimation routine. I discuss what moments are used to identify parameters related to the distribution of beliefs about one's type.

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 4: 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.

4.1 External Parameters

The parameters I 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 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. The distribution of financial assistance is assumed to follow a log normal distribution, of the form below, where parameters are estimated by OLS.

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

Where X_i includes demographic variables like race, ethnicity, gender, household net worth, parental education, and a constant term. The subscript k indicates that Equation 9 above is estimated separately for family assistance $k = 1$ and government/college 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 effect of being τ_l is normalized to 0. The effect of being τ_h will be determined 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 the specification for $Z_{i,j}$ follows Hai and Heckman 2017.

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

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 degree $s \geq 16$.

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.

The distribution of type will also be important for the effect of policies targeting by ability as 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 finite mixture model can be estimated using the individual likelihood function given by $f(\vec{Z}_i, w_i, g_i; \tau_k, X_i, s)$. These parameters in (9)-(12) are estimated by solving for the maximum likelihood given below in equation (13). 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.3-A.4.

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

After estimating financial assistance and the finite mixture model, I use the sum of the predicted financial assistance variables for total financial assistance, and predicted earnings from $\ln w_{i,s}$ for $w_n, w_s, w_c(\tau_l), w_c(\tau_h)$[10]. I use the finite mixture model $\pi(g, \tau)$ for the conditional grade probabilities that are used as signals and grade realizations. The individual probability of being a high-scorer that I use, $P_{\text{true},i}$, 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, given education outcomes, earnings, grades, and human capital measures estimated from the finite mixture model. $P_{\text{true},i}$ will be used to simulate high types and low types in the quantitative model. Given the simulated type the appropriate $\pi(g, \tau)$ will be used to generate grades for those missing grades.

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

4.2 Internally Estimated Moments

The remaining parameters that are estimated are the sticker price of tuition for stage 1 and stage 2, $tuit_1$, $tuit_2$, the distribution of subjective beliefs of being type τ_h, the non-pecuniary utility that depends on type, $\mu_c(\tau_i)$, as well the distribution of non-pecuniary utility shocks. The distribution of non-pecuniary utility shocks is given by the type I extreme value draws whose location parameters differ for White, Black, and potential first generation students. The scale parameter for the type 1 extreme value shocks are allowed to differ by stage.

The distribution of subjective beliefs of being high type is given below by equation (15) where values are truncated at zero and one given 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 Probability of Degree, $ProbDegr$, from the NLSY97, is a noisy measurement of the subjective belief of being type τ_h. The 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 one's and zero's since these imply certainty of type. Thus these beliefs are not amenable to change 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 boot strapping. The moments that will be targeted in the indirect inference specification are the coefficients for the following two regressions in equation (16) and (17).

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

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

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

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

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 specific problem that will be solved is given below, in equation (17). The parameter vector Γ are those parameters that minimize the difference between the simulated regression coefficients and data regression coefficients. The vector $\tilde{\beta}(\Gamma)$ is the vector of simulation coefficients given Γ, while the vector $\vec{\beta}$ is the vector of regression coefficients from the data. The weighting matrix is given by W which is the inverse of the diagonal matrix of the standard errors from the data regression coefficients.

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

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 will also be determined by $P_{true,i}$. 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

This section briefly discusses the identification strategy used to choose the targeted moments to estimate the parameters. For a quick reference see Table 5.

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.

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. The importance of beliefs also depend on relative utility of non college and some college. All of these depend on $w_c(\tau), w_s, w_n$, which is externally estimated. The difference in utility from college between low and high types is also determined by $\mu_c(\tau)$. The difference in expected utility between school and work is also determined by the location difference of type I extreme value shocks for school and work.

If the difference in utility between school and work,[11] is such that work is always preferred, or college is always preferred, then we would have trouble matching some of the patterns between beliefs, grades, and outcomes observed in the data. If type dependent non pecuniary utility $\mu_c(\tau)$, is such that there is no difference in utility between high and low scorers than we would have the same problems matching patterns in the data. This restricts these parameters to be such that there is an effect of measured beliefs and grades.

[11] College minus work type I extreme value shocks

Table 5: Key Internal Parameter Results

Parameter	Parameter Description	Target	Target Description
$\gamma_{p,0}$	Belief Constant	$\beta_{C,0}, \beta_{C,G_m}, \beta_{C,G_h}$	Constant, Coefficient med, high GPA on continuation
$\mu_c(\tau)$	Type dependent non pecuniary utility	$\beta_{C,0}, \beta_{C,G_m}, \beta_{C,G_h}$	Constant, Coefficient med, high GPA on continuation
$\gamma_{p,b}$	Belief: Meas Belief	$\beta_{E,B}$	Coefficient Meas Belief on enrollment
$\gamma_{p,h}$	Belief: Parent Education HSD	$\beta_{C,PH}$	Coefficient $Pedu_{hsg}$ on continuation
$\gamma_{p,s}$	Belief: Parent Education SCOL	$\beta_{C,PS}$	Coefficient $Pedu_{scol}$ on continuation
$\gamma_{p,c}$	Belief: Parent Education Bach	$\beta_{C,PB}$	Coefficient $Pedu_{bach}$ on continuation
$\mu_{d,0}$	Non-Pec Util: Black 1st Gen Col Stud	$\beta_{E,0} + \beta_{E,1G}$	Constant and $FirstGen$ Coefficient on enrollment
$\mu_{d,C}$	Non-Pec Util: Col Educated Parents	$\beta_{E,0}$	Constant Coefficient on enrollment
$\mu_{d,W}$	Non Pecun Util: White	$\beta_{E,W}, \beta_{C,W}$	$White$ Coefficient on enrollment,continuation
$\mu_{d,H}$	Non Pecun Util: Hispanic	$\beta_{E,H}, \beta_{C,H}$	$Hisp$ Coefficient on enrollment,continuation
$tuit_1$	Tuition Pd 1	$\beta_{E,F_2}, \beta_{E,F_3}$	$T2(Finaid), T3(Finaid)$ Coefficient on enrollment
$tuit_2$	Tuiton Pd 2	$\beta_{C,F_2}, \beta_{C,F_3}$	$T2(Finaid), T3(Finaid)$ Coefficient on continuation

Table 5: Description of Internally Estimated Parameters and moment targets identifying parameters. See equations (16) and (17) to see full regression specification.

Given these restrictions on non pecuniary utility, 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}, as well as the constant term in continuation $\beta_{C,0}$. We can also see in Panel 1 of Figure 6 that high and low grades do provide strong signals of type. For medium grades this is less so.

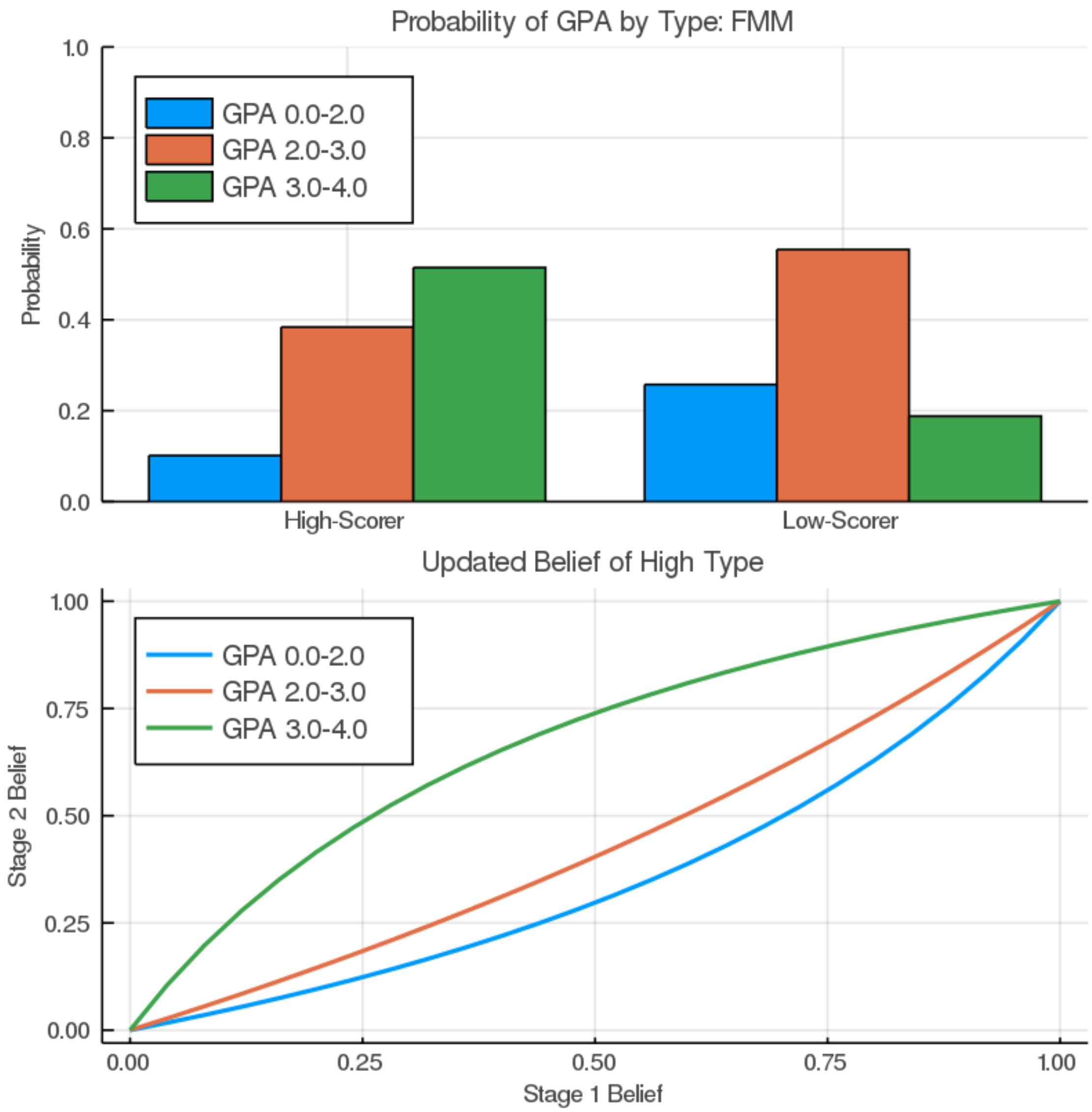

Figure 6: The top panel shows grade probability condition on being a high-scorer or low-scorer. While the bottom panel shows how prior subjective beliefs on the x-axis are updated to the posterior belief on the y-axis after grade realization.

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 in enrollment which enters the model through subjective beliefs about type.

Panel 2 in Figure 6 shows the updated belief given the type-dependent grade probabilities in Panel 1. Figure 6 shows that grades do provide a signal of type showing that there should be a differential response in non-continuation of grades.

The level of beliefs given through $\gamma_{p,0}$ is identified through difference in response to GPA. 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 as suggested by Figure 6 Panel 2. Therefore $\gamma_{p,0}$ will be set to where this best matches equation (17) from the data.

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 (15) and (16), given by $\beta_{C,W}, \beta_{C,H}, \beta E, W, \beta E, H$. 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).

The variance on the unobserved portion of belief in equation (15), the period specific scale parameters for the type I extreme values, and the period constant location parameter for the type I extreme value shocks, would help match the levels of enrollment and continuation, as well as create extra variation needed to fit the data. The type dependent non-pecuniary utility, $\mu_c(\tau)$ helps to adjust response to grades if the response implied by the finite mixture model through earnings and $\pi(g_k, \tau_j)$ is too restrictive. In total there are 16 parameters that are estimated by 17 moments.

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 6: Expected value of earnings from finite mixture model by education choice and type.

4.4 External and Internal Estimation results

Table 6 shows the estimated model earnings from the results of the log earnings equations of the finite mixture model. We see that regardless of type, 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, then all youth would choose to enroll and complete college. However in the presence of binding credit constraints the lower utility from low consumption for the first two periods may deter some youth from pursuing education. This is especially the case if they believe they will incur some non-pecuniary utility costs from being a low-scorer as well.

Table 7: Key Internal Parameter Results

Parameter	Description	Estimate
$\gamma_{p,0}$	Belief Constant	0.0057
		(0.0133)
$\gamma_{p,b}$	Belief: Meas Belief	0.88***
		(0.0103)
$\gamma_{p,h}$	Belief: P-Edu HSD	0.026**
		(0.0116)
$\gamma_{p,s}$	Belief: P-Edu SCOL	0.028***
		(0.0103)
$\gamma_{p,c}$	Belief: P-Edu Bach	0.055***
		(0.0102)
$\mu_{d,0}$	Non Pecun Util: Black 1st Gen Col Stud	-0.000056
		(0.000044)
$\mu_{d,C}$	Non Pecun Util: Col Edu Parents	0.00004
		(0.000037)
$\mu_{d,W}$	Non Pecun Util: White	0.000017
		(0.000028)
$\mu_{d,H}$	Non Pecun Util: Hispanic	0.000023
		(0.000034)
$\mu_c(\tau_h)$	Non Pecun Util high	0.00052***
		(0.000065)
$\mu_c(\tau_l)$	Non Pecun Util high	-0.0028***
		(0.00031)
$tuit_1$	Tuition Pd 1	$7583.61***
		(120.5)
$tuit_2$	Tuiton Pd 2	$6972.45***
		(16.05)

Boot strapped standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 7: Description of key internally estimated parameter results from indirect inference estimation. Standard errors are bootstrapped standard errors.

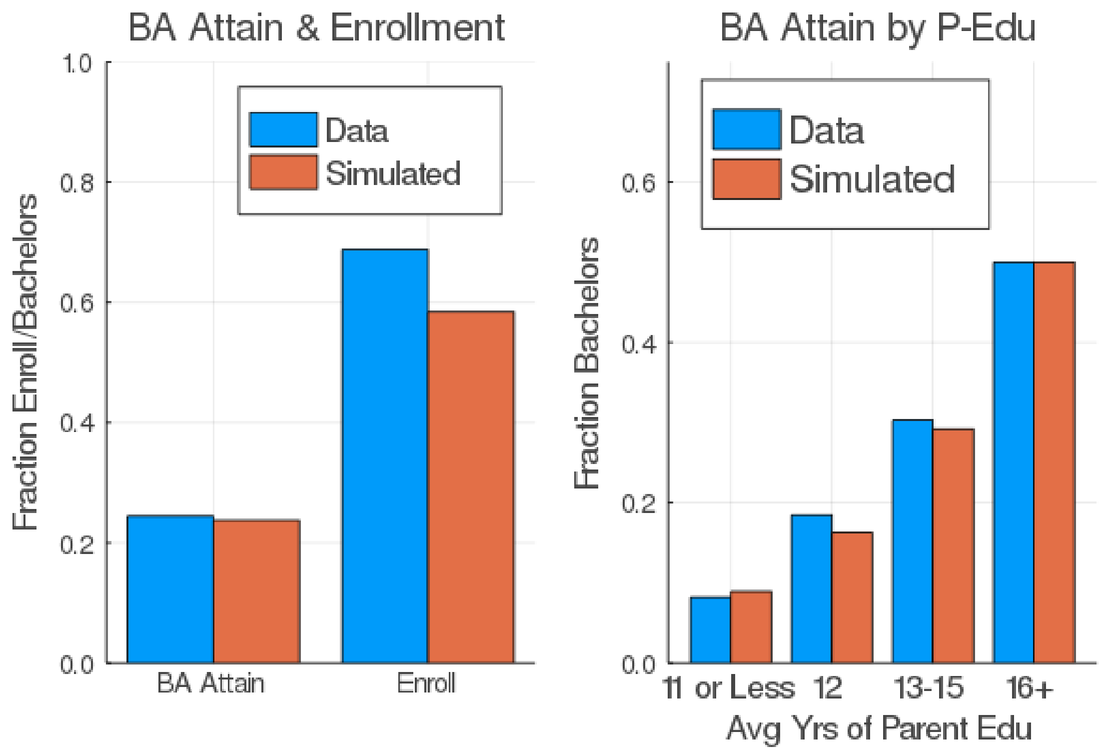

Figure 7: Fit of the Estimated Model: Enrollment, bachelors (BA) attainment. The blue bars are from the NLSY97 and the orange bars are simulated from the estimated model.

Table 7 shows several of the key parameters that were estimated in the internal calibration exercise. The coefficient on self-reported 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 college ability higher and perhaps closer to the truth if they are high-scorers.

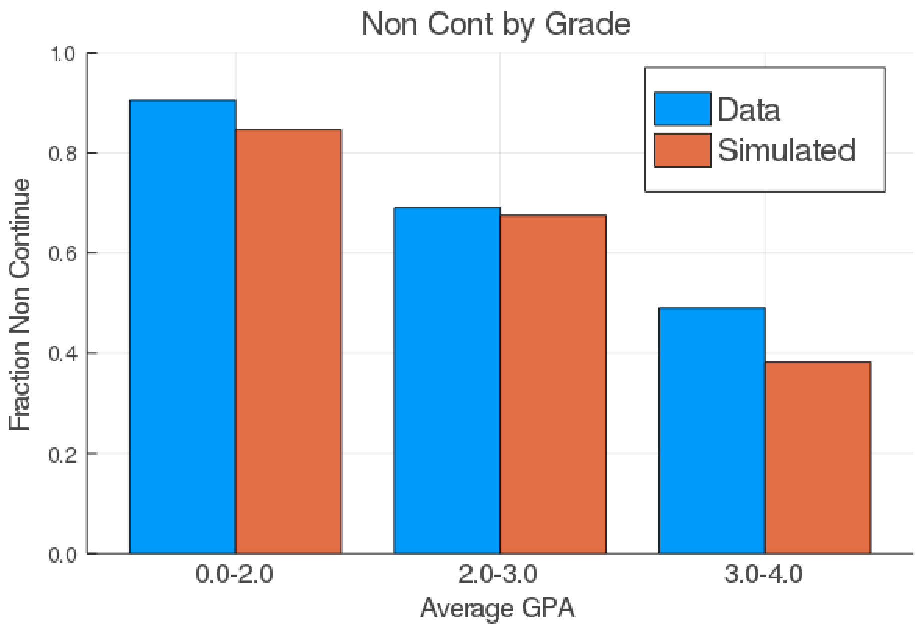

Figure 8: Fit of the Estimated Model: non-continuation by GPA level. The blue bars are from the NLSY97 and the orange bars are simulated from the estimated model.

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.5.For the results of the finite mixture model and financial assistance estimation see Appendix A.4.

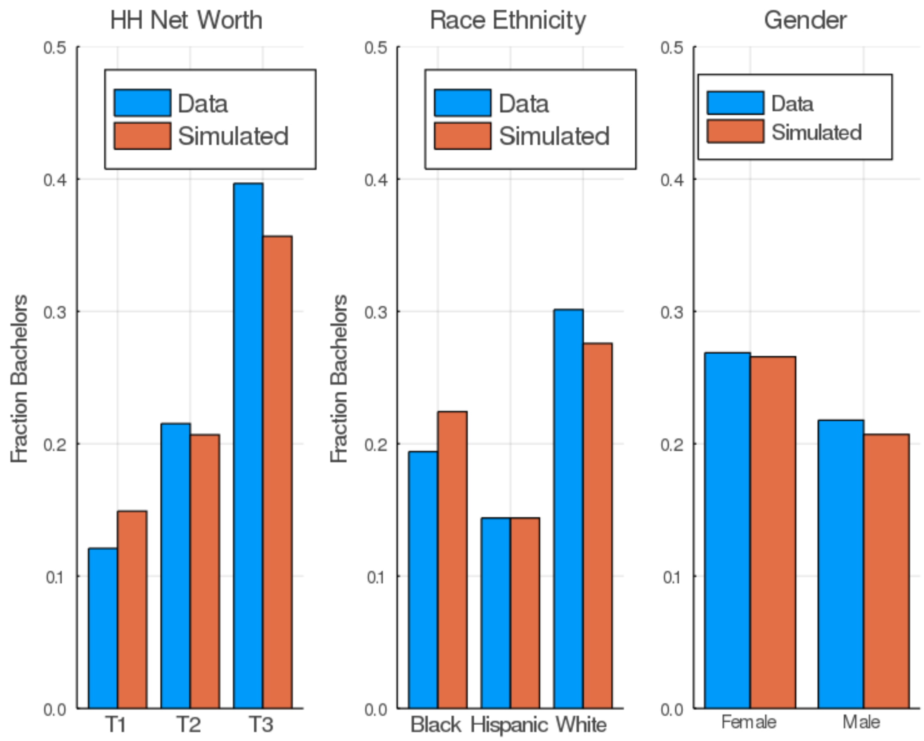

Figure 9: Fit of the Estimated Model: Bachelor's attainment by demographics.The blue bars are from the NLSY97 and the orange bars are simulated from the estimated model.

5 Quantitative Results

In this section I use the estimated quantitative model to answer two questions. The first is, for high scorers, how much of the gap in bachelor's attainment rates relative to high-SES White youth is explained by differences in beliefs and financial assistance? The second question is, will a policy that targets information and funding to high scorers from low-SES backgrounds be more efficient at closing overall bachelor's attainment gaps[12] than universal policies such as free college for all or instituting a tracking system in the United States.

For both questions the main outcome of interest, bachelor's attainment gaps, is de-

[12] gaps independent of scorer type relative to high-SES White youth.

fined as the difference in bachelor's attainment rates between high-SES White youth versus the three comparison groups, Black, Hispanic, and low-SES youth[13]. For the first question inequality is measured within high-scorers only. Where high scorers are those simulated by the model, whose realizations depend on human capital measures, earnings, and grades. For the second question inequality is measured independent of scorer type.

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 whose household is from the top tercile of the wealth distribution and whose parents have at least a bachelor's degree.

Before discussing the main results of this paper I discuss the estimated information frictions and mismatch present in the baseline version of the model.

5.1 Information Frictions and Mismatch

In this section I discuss information frictions and mismatch by scorer type, with some discussion on how this differs by demographic group. I also explain how this mismatch can help us predict the effect of policy on inequality.

In the model subjective beliefs can differ with respect to $P_{\text{true},i}$, which is how much they differ from a rational expectations prior. Subjective beliefs can also differ from the truth, which is the actual type of the agent.

Differences relative to $P_{\text{true},i}$ will determine the effects of policy providing more accurate estimates of P_{true} to youth. While differences in beliefs relative to the youth's actual type will affect the measure of efficiency used in this model, mismatch. In this analysis, mismatch is defined as the percentage of youth who would change their decision to get a bachelor's or not, if they knew their type with complete knowledge.

[13]Black and Hispanic youth includes youth from all socioeconomic backgrounds. Low-SES youth includes youth from all racial and ethnic groups in the sample

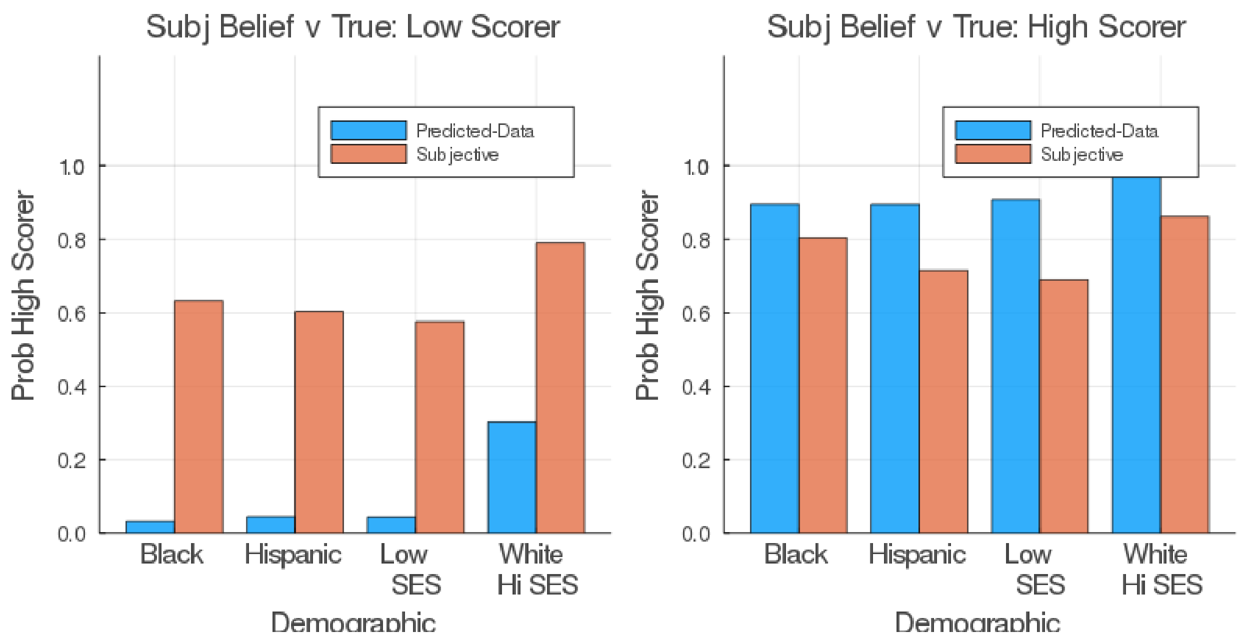

Figure 10: Compares the mean FMM estimate of prob high-scorer vs the mean subjective belief of being a high-scorer by scorer type. The left hand side is only predicted low-scorers, and the right hand side is only high-scorers.

Figure 10 shows the difference in mean subjective beliefs vs estimated $P_{\text{true},i}$ by demographic group and scorer type. We see that for Black, Hispanic, low-SES, and high-SES White youth subjective beliefs are on average inaccurate compared to those estimated from the data. 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 high-SES White youth beliefs are almost the same between low and high-scorers.

High-scorers are also closer to the truth than low-scorers for all demographic groups under consideration. Therefore we should expect that policies revealing estimates of $P_{\text{true},i}$, like tracking and the targeted policy, will have different effects by type. For instance tracking should lead to a bigger readjustment of low-scorers' beliefs than for high-scorers' beliefs. Providing information to everyone can perhaps lead to a bigger decline in bachelor's attainment from low scorers than the increase in bachelor's attainment from high-scorers. Additionally if there are more low scorers than high scorers among Black, Hispanic, or Low SES youth, then this can generate more inequality.

For the targeted policy that provides information only to low-SES high scorers, we should see that it should close gaps to low-SES youth more than for Hispanic and Black youth. This is because Black and Hispanic youth benefit only to the extent that they are also low-SES high scorers. Most importantly, the difference in beliefs with respect to $P_{\text{true},i}$ is larger for low-SES youth than for Black and Hispanic high-scorers.

Figure 10 can also help us predict the effect of for free college for all, a policy that does not address beliefs but reduces net tuition for everyone. Since in figure 10 low-scorers are overly optimistic, we should see that increasing funding to everyone will likely lead more low-scorers to enroll as well as some high scorers. This may increase over investment from low-scorers and perhaps mismatch as well.

Next Figures 11-12 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 11 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 all high-scorers. However, this is less the case for high-SES White youth.

The bottom panel in Figure 11 shows the bachelor's attainment rate of low-scorers in the baseline scenario and under complete information by demographic group. Conversely there is over investment occurring in the higher education market for low-scorers from all demographic groups. This over investment is highest for high-SES White low-scorers.

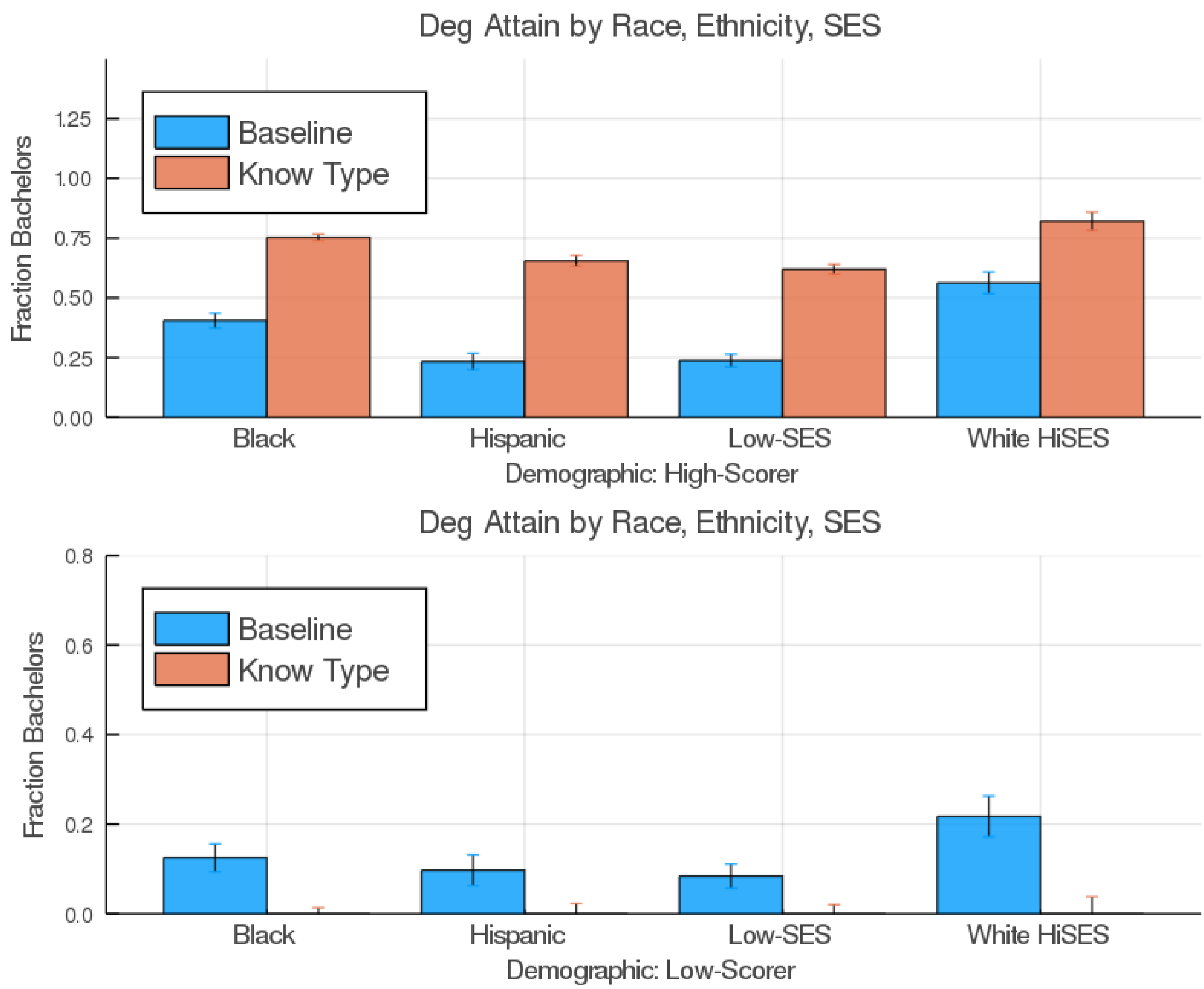

Figure 11: Shows difference in bachelor's attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at high-scorers by demographic group in the top panel, and low-scorers by group in the bottom panel.

Finally the top panel in Figure 12 shows the aggregate effect, independent of type, 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 12.

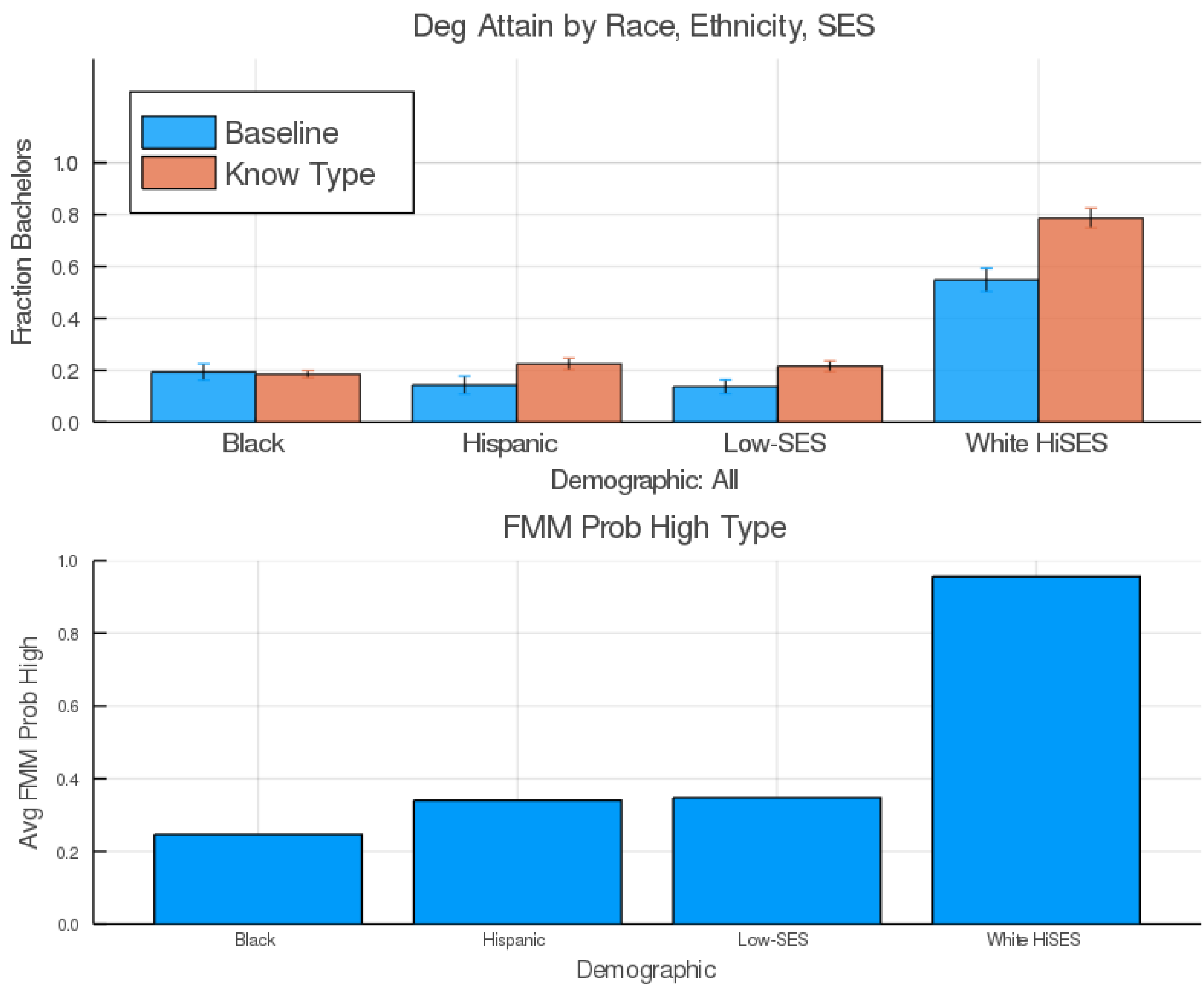

Figure 12: Top panel shows difference in bachelor's attainment regardless of type under baseline model and under scenario where youth know their true type with certainty. The bottom panel shows the proportion of high-scorers by demographic group in order to explain aggregate results.

Figure 12 suggests that independent of type, levels of mismatch are actually higher for high-SES White youth. We can also see that for Black youth there is little change in bachelor's attainment, and for the rest there are increases in bachelor's attainment. Since high-SES White attainment increases the most, complete knowledge of type might actually increase inequality, despite the gains in attainment for Hispanic and low-SES youth.

The difference of beliefs with respect to $P_{\text{true},i}$ and with respect to the actual individual's type will affect the results of the policy analysis. This will be explored in section 5.3. This is a separate question from the role that beliefs play in explaining gaps in the baseline scenario relative to high-SES White youth. What matters for this question is the differences of subjective beliefs and financial assistance between demographic groups. This will be answered in section 5.2 for high-scorers.

5.2 Decomposition

The first question I use the estimated model to answer is, for high-scorers, how much of the gap in bachelor's attainment rates between high-SES White youth versus Black, Hispanic, or low-SES youth, is explained by beliefs? A related question is how much is explained by differences in access to financial assistance?

To answer this question, I sequentially set beliefs of all high scorers to the mean value of high-SES White high scorers, then I set the financial assistance of all high scorers to the mean value of high-SES White high scorers. Since these are all high-scorers, human capital is assumed to be equal in the model. Therefore any remaining gaps are due to differences in non-pecuniary utility, entering the model through the distribution of the type I extreme value shocks.

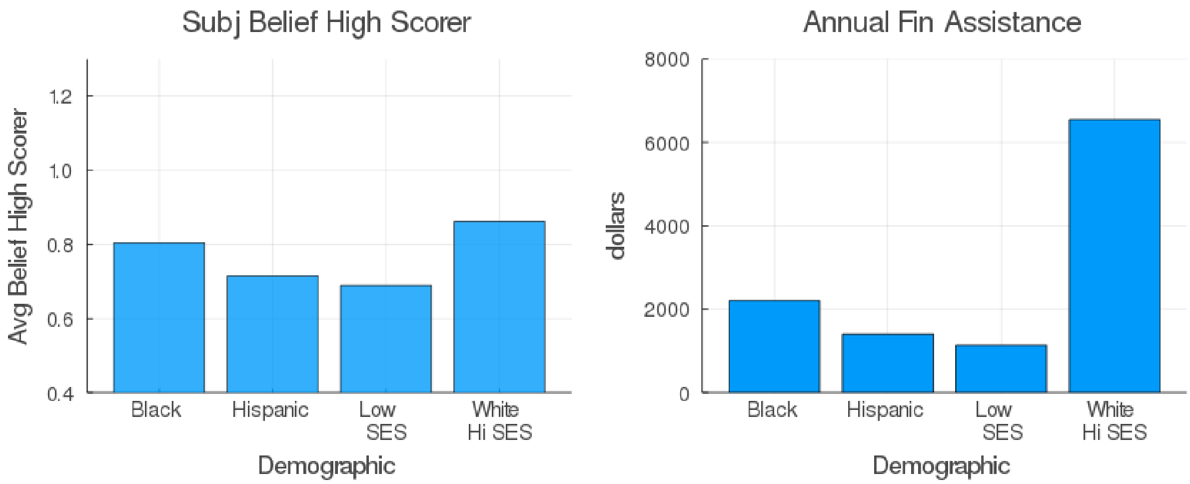

Figure 13: Shows the estimated variables relating to causal mechanism by demographic group. The left panel is the average subjective belief of being a high-scorer which is a function of measured beliefs in the NLSY97 and parental education. The right panel is predicted total financial assistance by demographics which is the sum of family assistance and govt/college aid.

Figure 13 shows mean subjective beliefs about being a high-scorer, and financial assistance by demographic group. Figure 13 shows that high-SES White youth not only receive higher levels of financial assistance on average but are also more optimistic on average. Gaps in average beliefs and financial assistance are also smaller between Black and high-SES White youth.

Figure 14 and Table 8 show the results of the decomposition exercise. In Figure 14, the y-axis shows high-SES White high-scorer bachelor's attainment rate minus the bachelor's attainment rate of the comparison group from the x-axis under the three scenarios in the legend. The rows titled difference in Table 8 provide the numerical values and standard errors for the information shown in the graph. The row titled "% Explained" shows the percentage decline in the gap after each step of the decomposition exercise for each of the demographic groups.

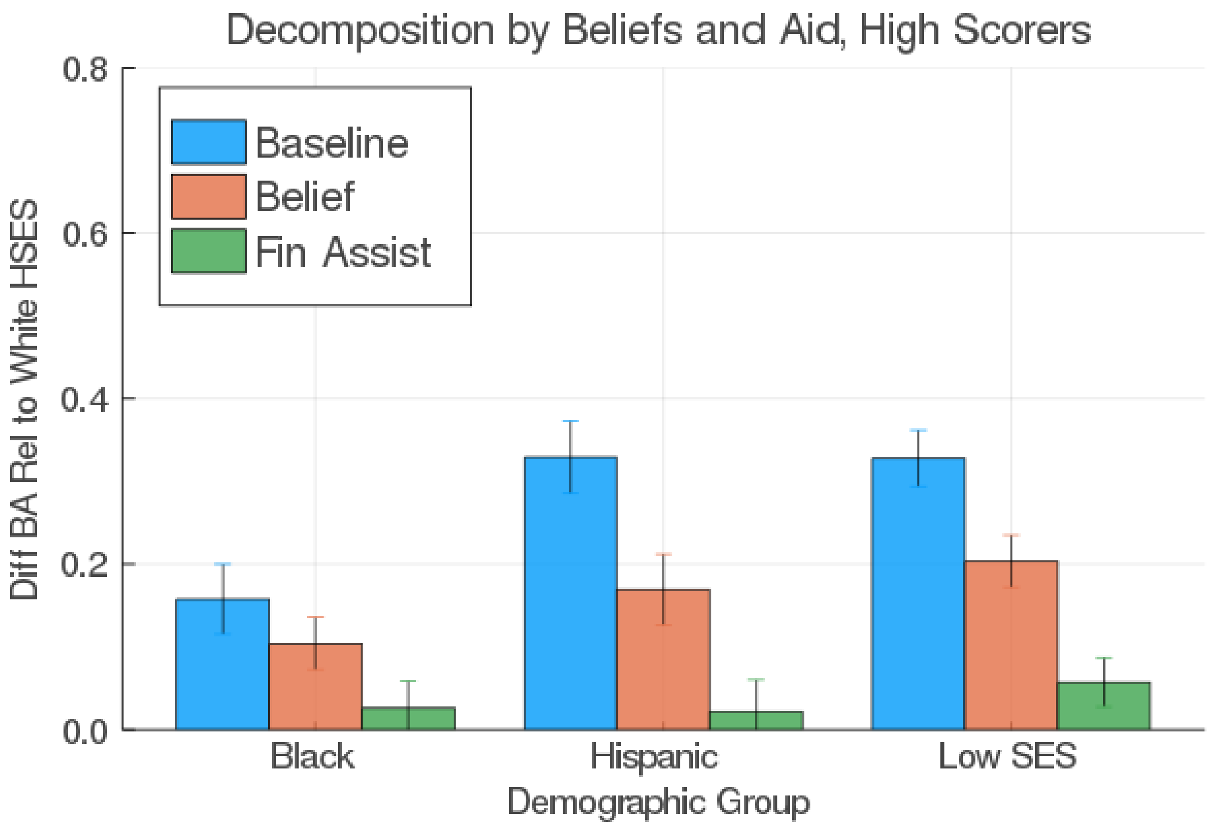

Figure 14: Difference in bachelors attainment relative to high-SES White high-scorers after sequentially equalizing variables. Standard errors are bootstrapped standard errors.

In Figure 14 and Table 8 we see that at each step of the decomposition exercise gaps are narrowed for each of the three comparison groups. However we would not be able to rule out a statistically zero effect of beliefs for Black high scorers, as shown by the large standard errors in Table 8. For Hispanic and low-SES youth however this is not the case. We can reject a null hypothesis that the effect of beliefs is zero. For these groups beliefs are estimated to explain an estimated 49% of the gap for Hispanic high-scorers, and 38% for low-SES high-scorers.

Table 8: Mechanism Decomposition: High Scorers

Demographic	(1) Baseline	(2) Beliefs Equal	(3) Fin Assist Equal
Black			
Difference	15.8*** (4.24)	10.4 (3.19)	2.6** (3.32)
% Explained		33 % (20.4)	50%*** (11.22)
Hispanic			
Difference	33*** (4.39)	16.9*** (4.29)	2.2*** (3.85)
% Explained		49 %*** (13.67)	45%*** (6.34)
Low SES			
Difference	32.8*** (3.39)	20.5*** (3.13)	5.7*** (2.96)
% Explained		38%*** (10.97)	45%*** (6.17)
White High SES Bachelor's attain	56		

Boot strapped standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 8: Shows how much of gaps in bachelor's attainment are explained by beliefs and financial assistance for high-scorers. The row "Difference" is the high-SES White bachelor's attainment rate minus the bachelor's attainment rate for the comparison group. The "% Explained" rows show how much of the difference in the baseline scenario is explained by each mechanism. For "Baseline" values on "Difference" rows and all values in "% Explained row", null hypothesis tested is whether the value is equal to zero. For the Beliefs Equal and Fin Assist Equal values on Difference rows, null hypothesis tested is whether the decrease relative to previous step in decomposition is equal to zero.

For all three comparison groups differences in financial assistance are also statistically significant. If the % Explained values in column 2 and 3 are added we can see the portion explained by differences in non-pecuniary utility. For Black and low-SES high-scorers, differences in non pecuniary utility explain nearly 17% of the gap in bachelor's attainment. For Hispanics the estimated portion explained by differences in non-pecuniary utility is only 6%.

The results in this exercise suggests that if Black, Hispanic, or low-SES high scorers are provided with information about their type as well financial assistance then gaps can be narrowed amongst high-scorers. The effect that an intervention like this has on overall inequality regardless of scorer type and how this compares to universal policies will be explored in the next section.

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, high-SES White youth, and the three comparison groups; Black, Hispanic, and low-SES youth. In this section gaps are measured independent of type. Mismatch takes the form of under investment in college for high-scorers and over investment for low-scorers. Where high and low scorers are simulated by the model. Realizations of being a high or low scorer depend on human capital measures, grades, and future earnings.

The first of the policies under consideration is a targeted policy that provides free college [14] and information about type to low-SES predicted high-scorers. Notice, for the predicted high scorers, we use only information that would be available to school administrators or policy makers before college, like standardized test scores. In this case, ASVAB measures and adverse behavior indicators are used to predict high scorers for the targeted policy.

The last two policies target everyone regardless of SES and predicted scorer type. These two policies are free college for all and the institution of a tracking policy which

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

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 $P_{true,i}$ that incorporates information that would be available before college completion.

Figure 15 and the rows titled "Difference" in Table 9 shows the difference in bachelor's attainment rates between high-SES White youth and each of the three comparison groups under each scenario. The row "% Change in Gap Relative to Baseline" shows by what percentage the gap changes after implementation of the policy compared to the Baseline. Negative percentage values indicate that the bachelor's attainment gap shrunk, while positive percentage values indicate that the gap increased.

We see that among universal policies free college for all decreases inequality for the three comparison groups,where decreases range between 14.7% to 16.9%. Tracking or better information for everyone actually leads to more inequality where the gap increases at a range of 41.5% to 70%. This is because tracking increases bachelor's attainment for high-SES White youth. Additionally for Black, Hispanic, and low-SES youth, gains in bachelor's attainment from high scorers are offset by decreases in bachelor's attainment rates for low scorers.

The targeted policy providing information and funding to low-SES high scorers decreases inequality the most, decreasing gaps between 25.2% for Black youth, 28.3% for Hispanic youth, and up to 41.8% for low-SES youth. 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 are also low-SES high-scorers. The policy may be strengthened if information and funding is targeted to Black and Hispanic high scorers regardless of SES.

Table 9: Policy Effect on Overall Inequality

Demographic	Baseline	Free College For All	Tracking: Info to All	Targeted: Info & Free
		for All	to All	Info & Free
Black				
Difference	35.4***	28.95**	60.22***	26.5***
	(3.11)	(3.16)	(3.10)	(3.18)
% Change in Gap		-18.3** %	70%***	-25.2 % ***
Relative to Baseline		(8.59)	(8.43)	(8.65)
Hispanic				
Difference	40.5***	33.6**	57.42***	29.02***
	(3.45)	(2.94)	(3.23)	(3.33)
% Change in Gap		-16.9 %**	42%***	-28.26%***
Relative to Baseline		(7.04)	(7.74)	(7.96)
Low SES				
Difference	41.1***	35.05**	58.2***	23.9***
	(2.69)	(2.71)	(2.95)	(3.08)
% Change in Gap		-14.7%**	41.5%***	-41.8%***
Relative to Baseline		(6.38)	(6.95)	(7.27)
White High SES	54.8			
Bachelor's attain				

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

Table 9: Shows the effect of the three policies on the difference in bachelor's attainment rates for Black, Hispanic, and low-SES youth relative to high-SES White youth, independent of scorer type. The row "Difference" is just the high-SES White bachelor's attainment rate minus the bachelor's attainment rate for the comparison group in the panel. "% Change in Gap Relative to Baseline" is calculated as the percentage change in bachelor's attainment differences after the policy. For "Baseline" column on "Difference" rows and values in "% Change in Gap" rows, null hypothesis tested is whether each individual value is equal to zero. For the corresponding policy values in "Difference" rows, hypothesis tested is whether the change in the gap relative to the corresponding gap in the Baseline column is equal to zero.

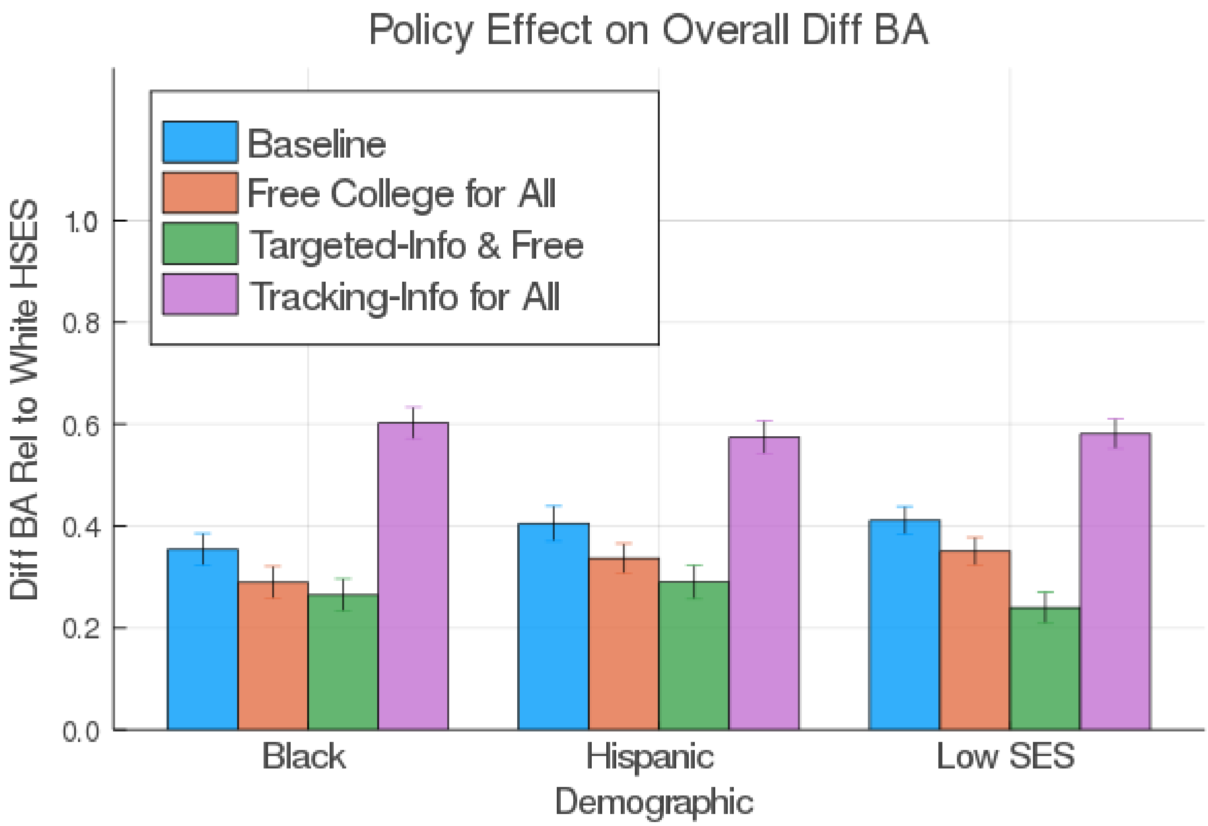

Figure 15: Shows differences in type independent bachelors attainment relative to high-SES White high-scorers after policy implementation. Standard errors are bootstrapped standard errors.

In Figure 15 gaps still exists after all three policies are implemented. This highlights the important role that disparities in early childhood human capital investment likely still plays in generating inequality. Specifically, even if the targeted policy lead to 100% bachelor's attainment for low-SES high scorers, the fact that discrepancies in early childhood human capital generate a lower proportion of high scorers among low-SES youth, would mean that there would still be inequality in overall attainment.

Table 10: 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%

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

Table 10: Shows the percentage of the population in the simulations that would change their decision to obtain a bachelor's degree or not if they knew their type. Values in the second and third columns or each row sum to the first column since they are percentage of population that are mismatched and high-scorer and percentage of population that are mismatched and low-scorers.

Finally Table 10 shows the amount of mismatch present in higher education and how it is distributed among high-scorers and low-scorers. We see that 27 % of youth would change their college decisions if they knew their type for certain. 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 30.5%, with no decrease in under investment of high-scorers, but a larger 3.3% increase in over investment of low-scorers. As expected the tracking system decreases mismatch the most by almost completely removing all mismatch for low-scorers. This is because it brings all youth closer to the truth by revealing $P_{true,i}$.

The targeted policy decreases overall mismatch but primarily only through high-scorers. The percentage of the population that are mismatched and high-scorers decreases by 8 percentage points under the tracking policy.

Together Table 9-10 shows that if we are interested in policy that decreases inequality with minimal effects on mismatch, then the targeted policy is to be preferred. This is because it not only decreases inequality the most but also decreases mismatch. Tracking increases inequality which would make it undesirable if decreasing inequality was our main policy objective. Free college for all decreases inequality less effectively than the targeted policy and generates more mismatch as well.

An additional benefit to the targeted policy is that 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 than the model would suggest (Hoxby and Avery 2012, and Dynarski, Libassi, Michelmore, and Owen 2017).

Even if the targeted policy is to be preferred, 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.

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 correlated 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 high-SES White 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 in terms of decreasing inequality while minimizing mismatch, targeted policies that provide information about ability type and funding to low-SES high-scorers are to be preferred to free college for all and instituting a tracking system in the US. This is because the targeted policy not only more effectively closes gaps, but also decreases mismatch. The other two policies exhibit equity efficiency trade offs, where free college for all decreases inequality and increases mismatch, while tracking increases inequality and decreases mismatch.

Therefore this paper shows that information frictions lead to less high scoring youth from all backgrounds under investing in education. These information frictions also contribute to inequality in higher education for Hispanic and low-SES youth. Because of that representation in higher education can be increased through more recruiting of academic high achievers from disadvantaged backgrounds. However because of differences in early childhood human capital development and perhaps college experience, gaps are likely to still persist. Therefore In order to fully close all gaps we must still study the effects of improving K-12 education and household environment.

7 Bibliography

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

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

Angrist and Krueger. 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, and Ransom. 2016. "College Attrition and the Dynamics of Information Revelation." Working Paper.

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

Bettinger, Long, Oreopoulos, and 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 and 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. 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. 1999."The Causal Effect of Education on Earnings," in Handbook of Labor Economics.

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

Chetty, Hendrin, Jones, and Porter. 2018. "Race and Economic Opportunity in the United States: An Intergenerational Perspective." NBER.

Chetty, Friedman, Saez, Turner, and Yagan. 2017 "Income Segregation and Intergenerational Mobility Across Colleges in the United States." NBER.

Bell, Chetty, Jaravel Petkova, and Van Reenan. 2019 "Who Becomes an Inventor in America? The Importance of Exposure to Innovation." Quarterly Journal of Economics. 134(2)

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

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

Cunha, Heckman, and 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, Libassi, Michelmore, and Owen. 2020. "Closing the Gap: The Effect of a Targeted, Tuition-Free Promise on College Choices of High Achieving, Low-Income Students." American Economic Review.

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

Keane and 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, and Lopes De-Melo. 2016. ”Occupational Choice, Human Capital and Learning: A Multi-Armed Bandit Approach.” Working Paper.

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

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

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

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

Wiswall, and 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 Data Construction and Summary Statistics

From the NLSY97 I use data on parental education, household net worth, self reported probabilities of school enrollment and obtaining a degree by age 30[15], labor market earnings, 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.

I use measures of cognitive human capital and non-cognitive human capital to in the empirical and quantitative analysis to control for early childhood human capital stock (Heckman and Kautz 2014). 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 (Hai and Heckman 2017).

For the empirical analysis and the structural model estimation 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. For grades I use transcript data in the NLSY97 for GPA, as opposed to the self reported data. I impute transcript GPA for individuals that are missing GPA information from their college transcripts. I do this by regressing transcript GPA on self reported college GPA, demographic characteristics, and human capital measures. I then use the predicted transcript values from the portion of the sample that only includes self-reported GPA.

For parental education I take the average of mother's and father's years of schooling

[15]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.

if both parents education level is in the data. If only one parent is in the data then I use that parent's years of schooling to measure parental education I bottom code at 8 years of schooling and top code at 16 years of schooling. For household net worth, I use the parent's reported household net worth at the start of the survey, before agents enroll in college. For individuals that do not have parental reports of net worth recorded, I impute household net worth using the individual youth's report.

I drop individuals that identify as Asian, Native American 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 created by the Bureau of Labor Statistics for the NLSY97[16].

Summary statistics by race an parental education are reported in Table 11-12.

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

Table 11: 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 12: 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

A.2 Supplementary Analysis

Table 13: Financial Assistance

VARIABLES	(1) Any Family Aid	(2) Total Fam Aid	(3) Any Govt/Coll Aid	(4) Total Govt/Coll 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: Shows the results from OLS specifications regressing financial assistance variable on covariates. Columns (1) and (2) are for family assistance while in college, while (3) and (4) are for government or college financial aid while in college, including grants, scholarships, and work study. Columns (1) and (3) are linear probability models since they are indicators for if any assistance was provided.

Table 14: 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 14: Shows Oaxaca Blinder decomposition results explaining unconditional differences in mean self reported probability of degree for Hispanic/Black youth compared to White youth. Explained is portion explained by differences in explanatory variables and unexplained column is portion explained by coefficient results in race/ethnicity separate regressions.

Table 15: 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 15: Shows Oaxaca Blinder decomposition results explaining unconditional differences in mean college enrollment for Hispanic/Black youth compared to White youth. Explained is portion explained by differences in explanatory variables and unexplained column is portion explained by coefficient results in race/ethnicity separate regressions.

Table 16: 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 16: Shows Oaxaca Blinder decomposition results explaining unconditional differences in mean college continuation for Hispanic/Black youth compared to White youth. Explained is portion explained by differences in explanatory variables and unexplained column is portion explained by coefficient results in race/ethnicity separate regressions.

Table 17: 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 17: Shows the OLS results from regressing log average earnings (where 1 is added to values to include zeros) on education seperately (first three columns) then interacted with all variables (Last two columns).

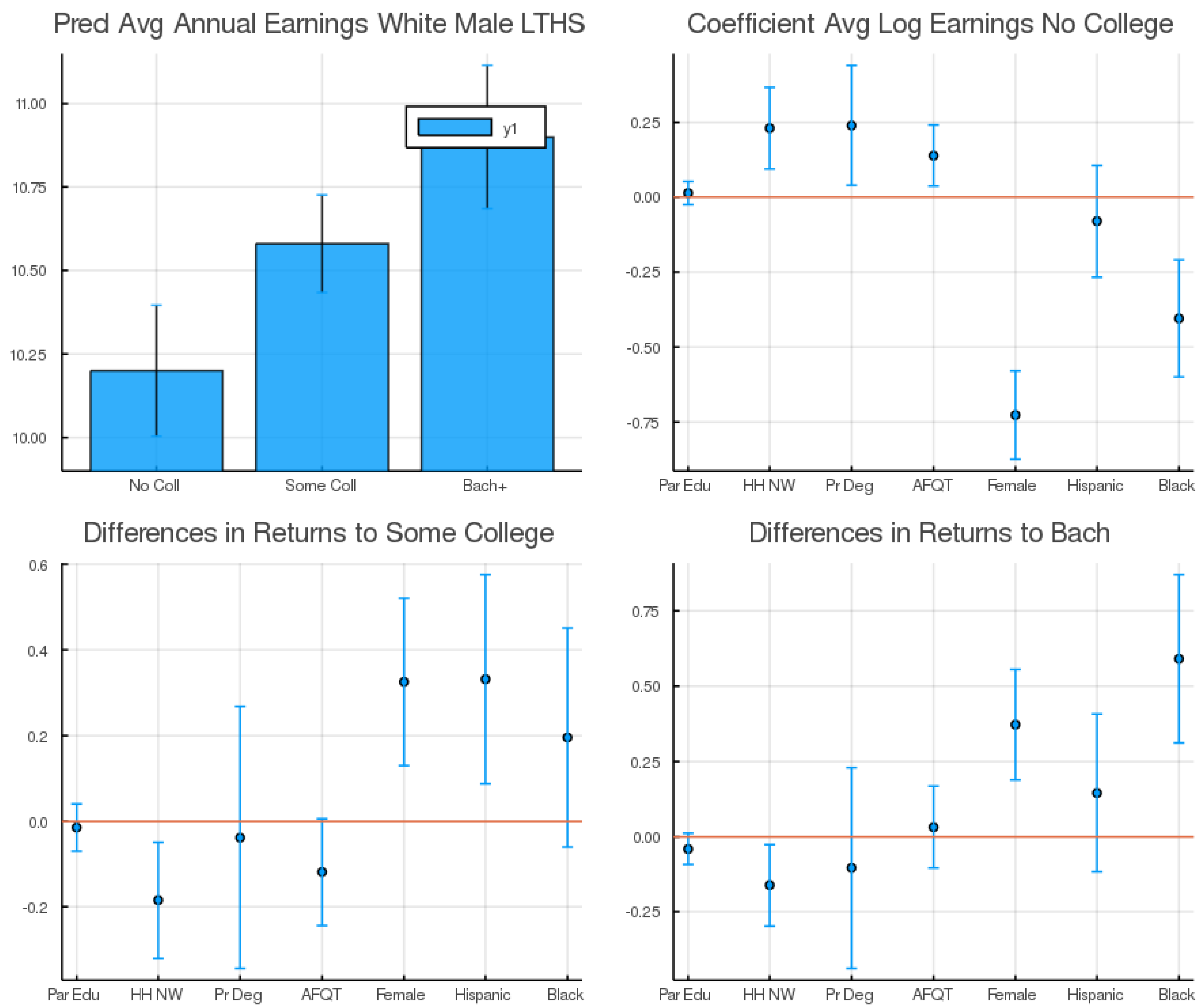

Figure 16: Shows predicted average earnings from Table 17 in top left panel, while other three panels plot coefficients to provide scale for the results.

A.3 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 the difference 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 the difference 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, multiplied 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 probability 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 $\text{Prob}(s \in (12,16))$ and probability of having a bachelor's degree $\text{Prob}(s \geq 16)$.

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

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

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

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

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

A.4 Finite Mixture Model Results

Table 18: Funding by Demographic: External Estimate

VARIABLES	OLS log Family Aid	OLS log Gov Coll Aid
Intercept	-0.963	3.67***
	(0.637)	(0.722)
Parent Edu	0.347***	0.0455
	(0.045)	(0.0513)
HH Net Worth ($1000s)	0.0032***	-0.0012***
	(0.0004)	(0.00046)
Black	-0.718***	1.093***
	(0.217)	(0.246)
Hispanic	-0.144	0.311
	(0.258)	(0.292)
Female	0.182	0.587
	(0.171)	(0.194)
Birth Yr 1981	0.329	0.0436
	(0.245)	(0.278)
Birth Yr 1983	0.114	-0.0238
	(0.247)	(0.280)
Birth Yr 1984	0.415*	0.161
	(0.245)	(0.277)
Observations	1,467	1,467
R-squared	0.1554	0.0345

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 18: Shows the results from the OLS estimates of log financial assistance from family and government/college. A value of one is added to account for zeros. Total financial assistance in the model is calculated by summing the predicted financial assistance from family with the predicted financial assistance from governments or colleges.

Table 19: 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)
Hispanic	-0.655***	0.307**	-0.040
	(0.201)	(0.145)	(0.189)
Black	-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)

Boot Strapped standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table 19: 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.

Table 20: 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)	

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

Table 20: Shows the results from the finite mixture model for human capital variables. High Type is a binary variable if agent is high type.

Table 21: 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 Type	0.256***	
	(0.075)	
Std Error Unobserved Shock	0.83***	
	(0.0223)	

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

Table 21: Shows the results from the finite mixture model. High Type 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 Intercept and Ever Enrolled coefficient if low type. If high type also add Bachelor*High coefficient.

A.5 Indirect Inference: Targeted vs Simulated Moments

Table 22: 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 22: 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 23: Key Internal Parameter Results

Parameter	Description	Estimate
$\gamma_{p,0}$	Belief Constant	0.0057
		(0.0133)
$\gamma_{p,b}$	Belief: Meas Belief	0.88***
		(0.0103)
$\gamma_{p,h}$	Belief: P-Edu HSD	0.026**
		(0.0116)
$\gamma_{p,s}$	Belief: P-Edu SCOL	0.028***
		(0.0103)
$\gamma_{p,c}$	Belief: P-Edu Bach	0.055***
		(0.0102)
σ_p	Belief: Var Error	0.00018***
		(0.000043)
$\mu_{d,0}$	Non Pecun Util: Black 1st Gen Col Stud	-0.000056
		(0.000044)
$\mu_{d,C}$	Non Pecun Util: Col Edu Parents	0.00004
		(0.000037)
$\mu_{d,W}$	Non Pecun Util: White	0.000017
		(0.000028)
$\mu_{d,H}$	Non Pecun Util: Hispanic	0.000023
		(0.000034)
$\sigma_{d,1}$	Non Pecun Util Scale pd 1	0.000043
		(0.000066)
$\sigma_{d,2}$	Non Pecun Util Scale pd 2	0.000027
		(0.000066)
$\mu_c(\tau_h)$	Non Pecun Util high	0.00052***
		(0.000065)
$\mu_c(\tau_l)$	Non Pecun Util high	-0.0028***
		(0.00031)
$tuit_1$	Tuition Pd 1	$7583.61***
		(120.5)
$tuit_2$	Tuiton Pd 2	$6972.45***
		(16.05)

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

Table 23: Shows the full list of 16 parameters estimated by indirect inference

A.6 Mismatch by Net Worth, Parental Edu, Race, and Ethnicity

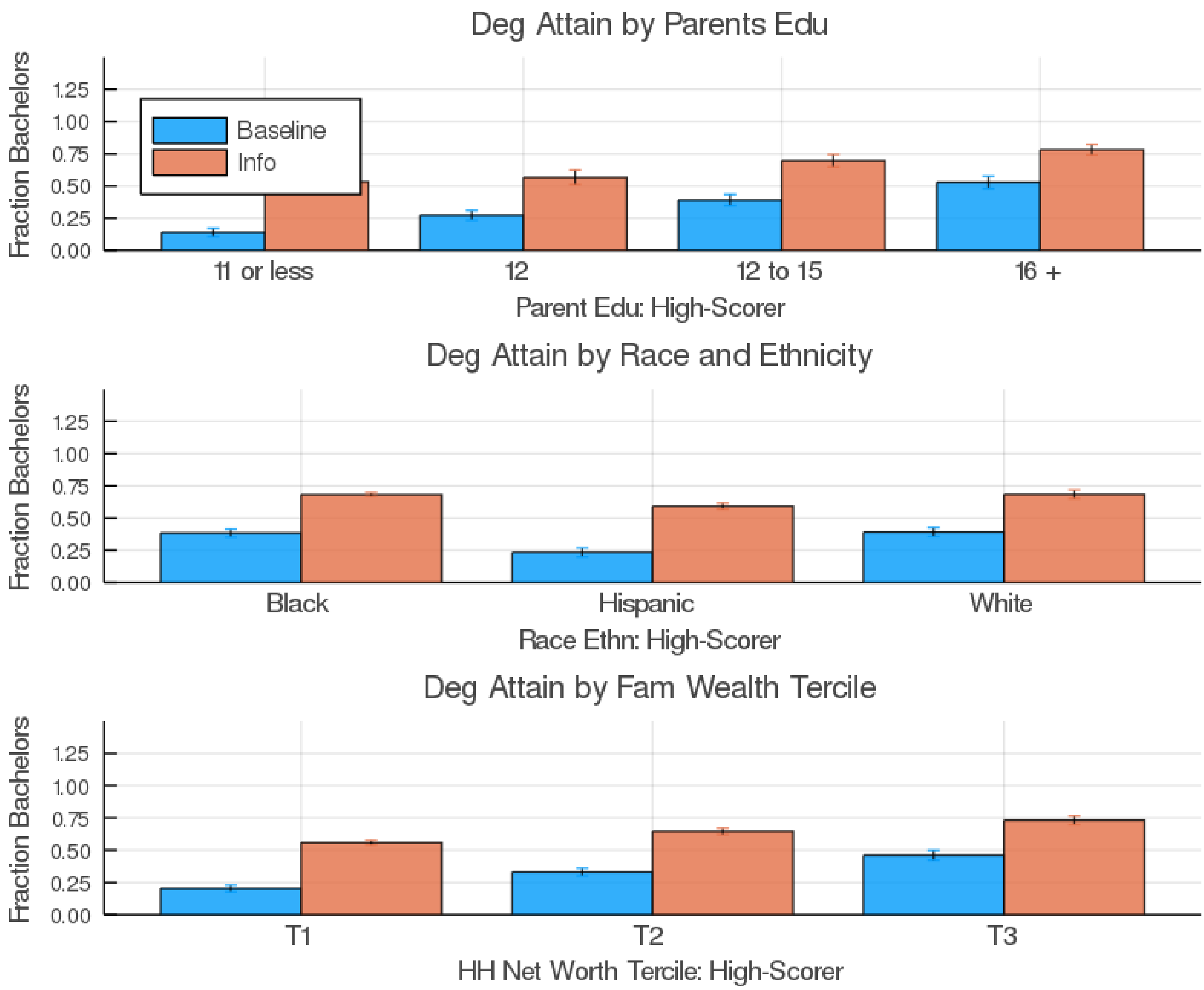

Figure 17: Shows difference in BA attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at predicted High Scorers by demographic group.

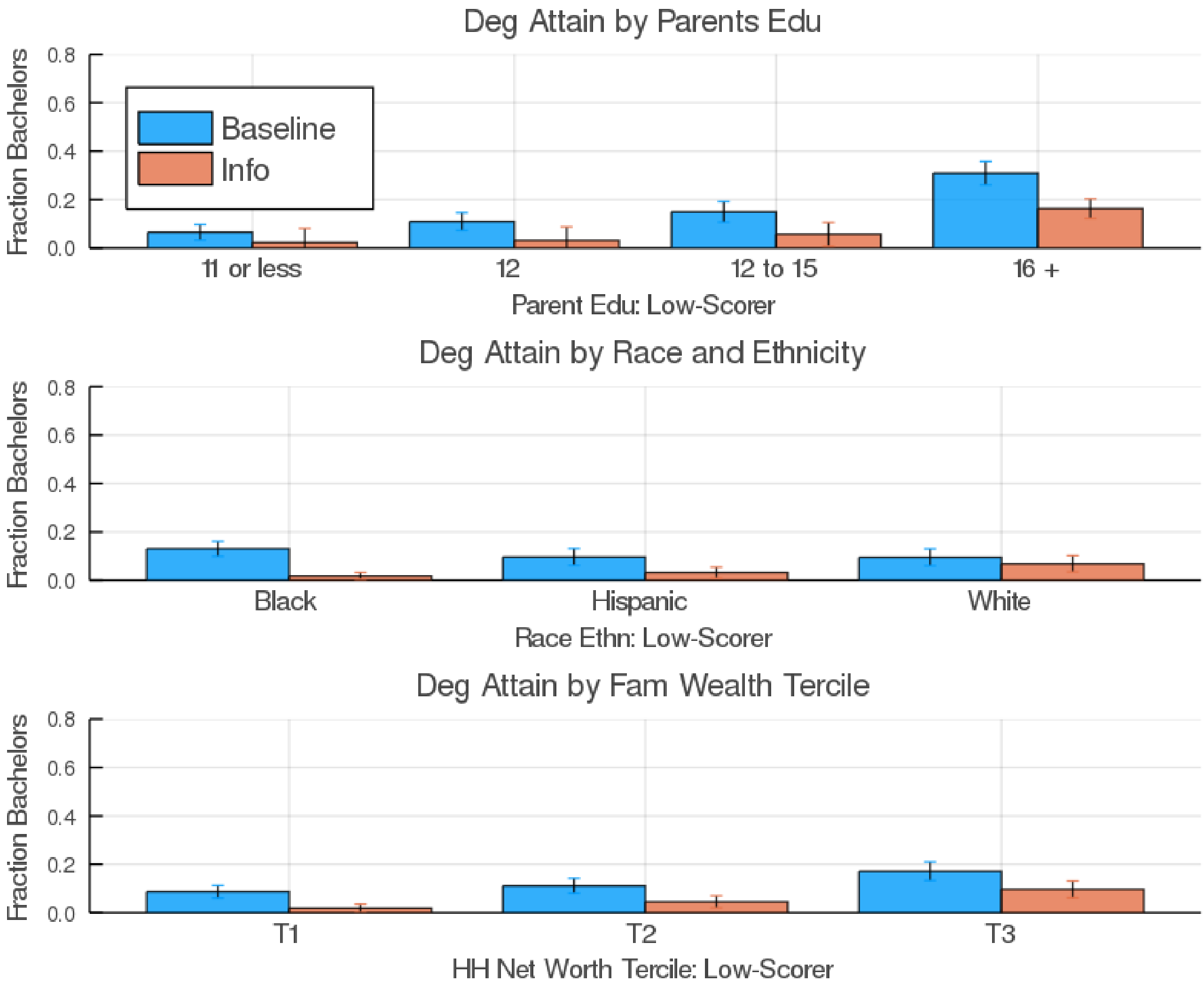

Figure 18: Shows difference in BA attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at predicted Lower Scorers by demographic group.

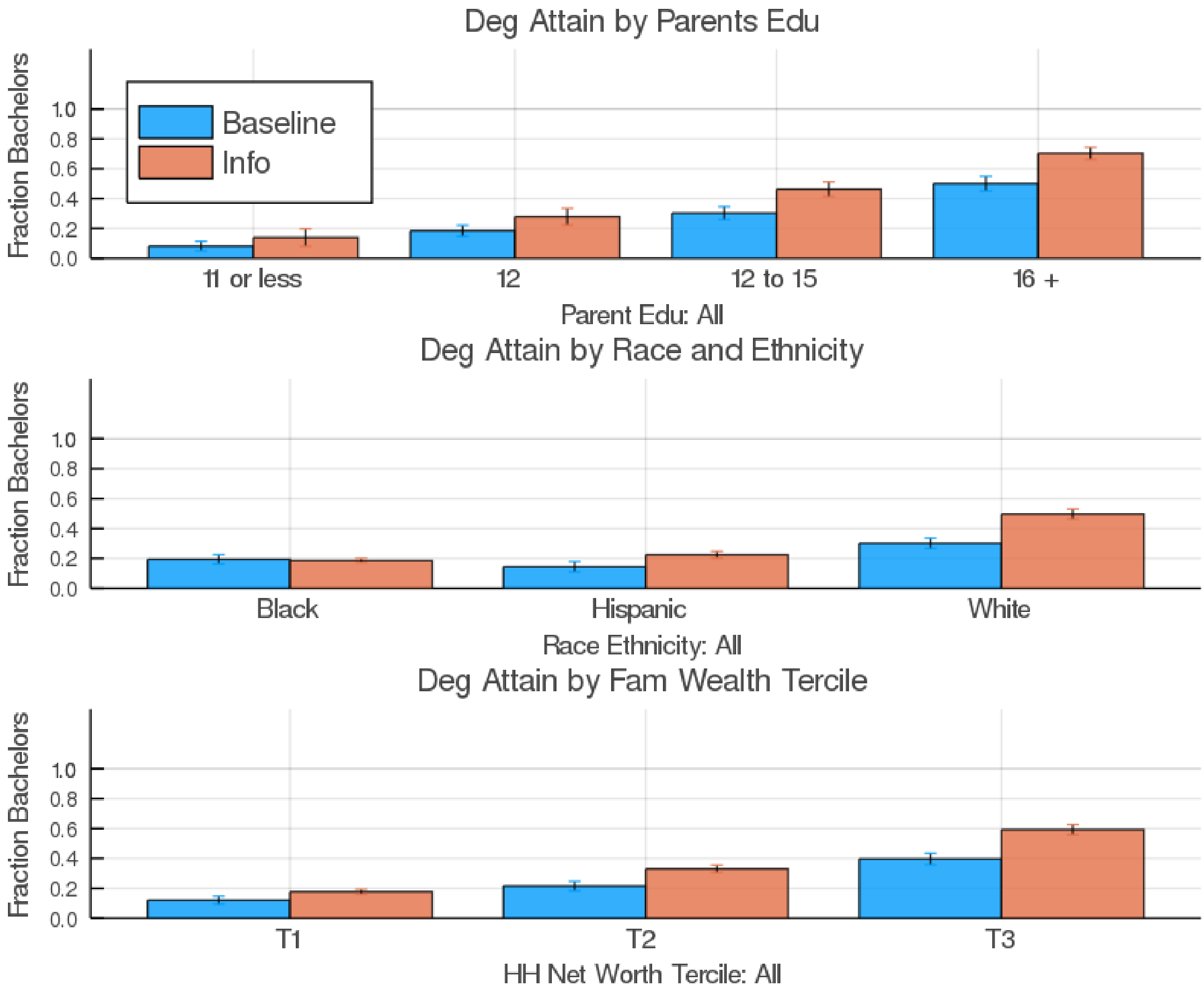

Figure 19: Shows difference in BA attainment under baseline model and under scenario where youth know their true type with certainty. This graph looks at all youth regardless of scoring type by demographic group.

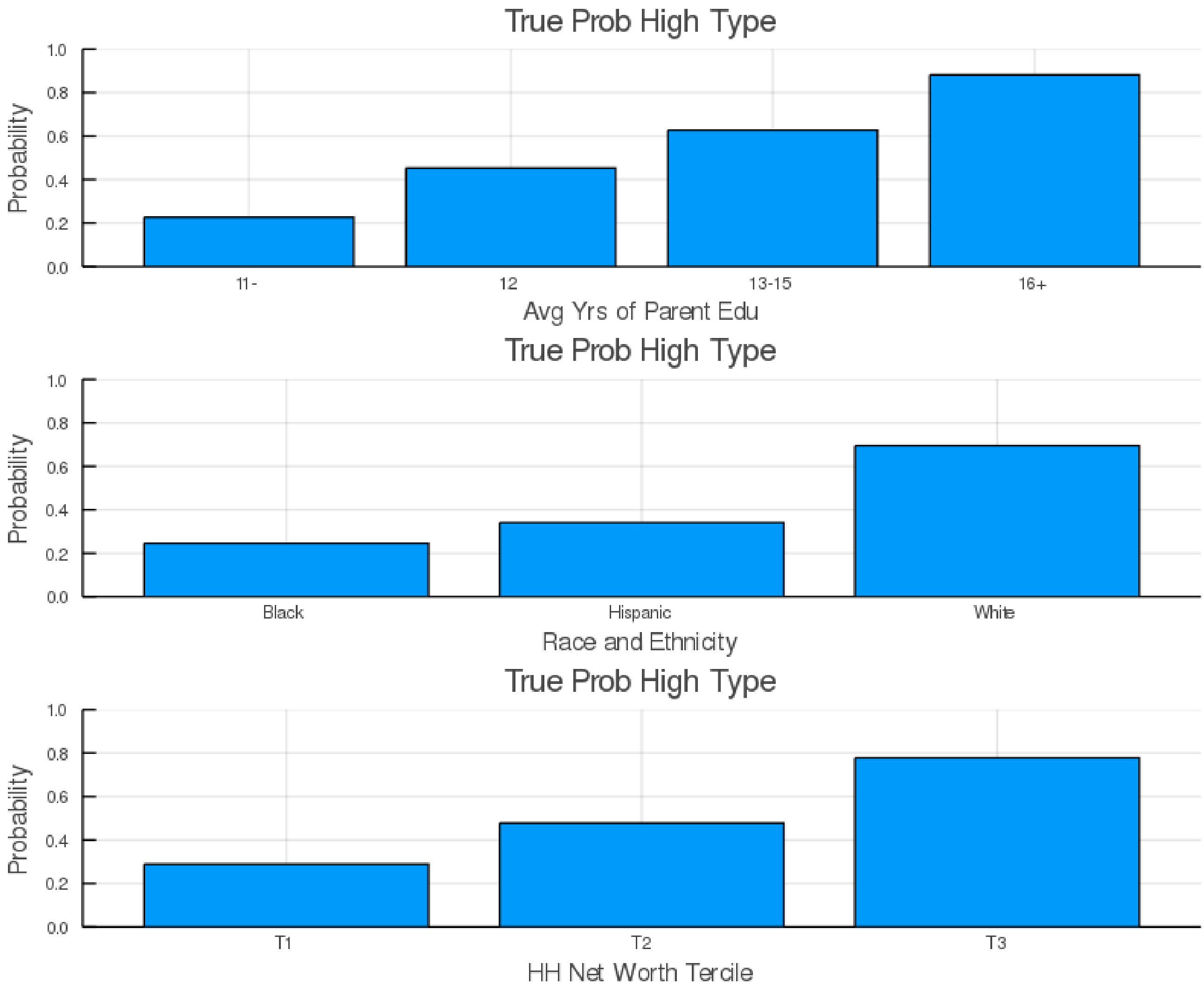

Figure 20: Shows the estimated fraction of high-scorers by demographic background from the finite mixture model.