Subjective beliefs and the choices of community college students

John Green[†]

Updated September 8, 2025

Abstract

Economists are increasingly modeling decision making with limited information and subjective beliefs. Education choice has been a fruitful area for this research, but despite their important role as engines of economic mobility, the decisions of community college students have been largely neglected. I present novel survey data from a large and diverse Maryland community college, with data on beliefs about the labor market returns to 2- and 4-year degrees and offering the first systematic look at the subjective beliefs of community college students about dropout risk and tuition cost after transfer to 4-year school. I find evidence that student beliefs about graduation and tuition costs in 4-year programs deviate significantly from rational expectations; respondents appear *pessimistic* about the cost of 4-year school, but *optimistic* about the probability of graduation after transfer, with ambiguous net effect. I use multinomial logit, motivated by a simple static random utility model, to show how intentions to pursue 2- and 4-year degrees depend on expected outcomes from each degree, and show that expected *non*-pecuniary returns are more important than the expected wage premium. I introduce a dynamic model of education choice, consumption, and labor supply; in future work I will estimate the model and use it to explore the costs of biases and to evaluate financial aid policies. The second wave of the survey, planned for fall 2025, will allow me to model belief formation and evaluate the accuracy of first round expectations.

JEL Classification: I21, I23, I26, D83, D84

[†] Ph.D. student at Johns Hopkins Department of Economics. This is a work-in-progress draft of my job market paper.

1 Introduction

One of the most active areas of research in education economics is the "college puzzle."[1] Given substantial returns to a college degree, why don't more individuals earn a degree? Going further, why has the growth in college attendance flattened in recent years?

Figure 1 shows attendance and attainment trends from the CPS. The portion of 25-29 year-olds with a 2-year community college degree increases significantly from 1962 to 1980, before leveling out at about 10%.[2] It decreases in the early 1990s, but recovers to previous levels by 2010 before beginning to decrease in 2020. 4-year college attainment among 18-24 year-olds also increases sharply from 1962 to the late 1970s, flattens out, and then begins to climb again around 2005. Enrollment data is not available as far back as attainment and in the CPS is not separated between community college and 4-year schools, but the percentage of 18-24 year-olds (excluding those enrolled in secondary school) enrolled in college at any level can be seen to grow significantly from 1986 to 2012. From its peak it has declined significantly, from 44.1% in 2012 to 38.3% in 2023. Importantly, we can see that despite substantial growth in degree attainment over the long run, it still the case that only a minority of young people (about 40%) have earned a 4-year degree, and 10% a 2-year degree. What's more, 2-year degree attainment is essentially flat over 45 years.

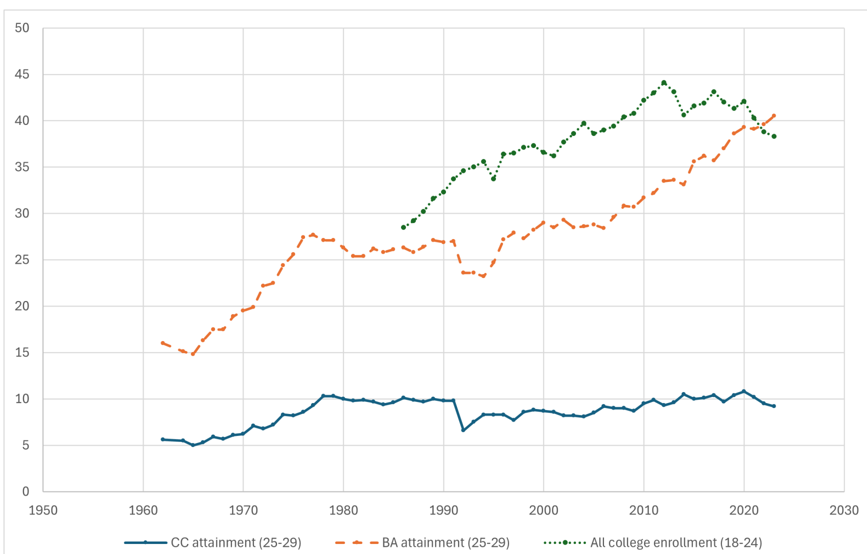


Figure 1: College attendance and attainment

Although the discussion of the college puzzle is usually focused on 4-year programs, the trends in figure 1 show a similar puzzle in the case of community colleges. Community colleges play a critical role in the US

[1] Dynarski et al. (2022) provides a good discussion of this issue.

[2] Tracking this age group lets us track trends in college attendance and attainment at the cohort level, which are then more gradually reflected in the population of prime-age workers.

higher education system, providing a local and financially viable pathway to a college degree which serve nearly 10 million students each year. However, community colleges are marked by two phenomenon: a high rate of dropout (per the National Center for Education Statistics (NCES), only 34% of the incoming cohort in 2017 earned a degree within 150% of the expected time), and a low rate of transfer to four year programs (although 80% of community college students intend to transfer per Horn & Skomsvold (2011), only about 33% will make the transition to a 4-year program, and only about half of these transfers will earn a BA (Tatiana Velasco, John Fink, Mariel Bedoya-Guevara, Davis Jenkins & Tania LaViolet (2024))). These twin phenomena may be framed as a *community college* puzzle: given the earnings premium from a 2-year degree, why do so many students leave community college with no degree? And given the further benefit to a BA, and in light of the stated mission of community colleges to facilitate access to 4-year schools, why do so few students transfer and earn a BA?

In this paper I study the role of subjective (and potentially biased) beliefs in the education choices of community college students. I collect novel survey data on a variety of beliefs and probabilities from students at a large and diverse Maryland community college system, providing the first evidence on the beliefs of community college students about outcomes from transfer. Although beliefs and choices obey sensible patterns, I document deviations from rational expectations in beliefs around tuition costs and graduation rates at 4-year programs. I use a simple static choice model in a random utility framework to motivate a multinomial logit which explains reported probabilities using demographic characteristiscs and expected outcomes. The parameter estimates follow the expected signs, with students particularly sensitive to increases in expected tuition at 4-year school. The model suggests that non-pecuniary returns to a 4-year degree are far more important than the expected wage premium. Finally, I move beyond the static model and present a dynamic model of the education choices of community college students which can be identified and estimated with my beliefs data.

This paper is a preliminary draft of my job market paper, and as such is a work-in-progress. In further work, I will estimate the dynamic model, and then use it to examine the efficacy of financial aid policies as well as to explore the welfare costs of biased beliefs. A second round of the survey, to be administered in the fall of 2025, will allow me to model learning, better control for endogeneity, and examine the accuracy of beliefs at the *individual* level by comparing expectations with outcomes.

The paper proceeds as follows. In section 2 I briefly review the literatures on education choice and subjective beliefs. In section 3 I detail the survey methodology and the context in which the work was carried out. 4 presents descriptive analysis of the survey data: summary statistics, a reduced-form regression of wage expectations, and the results of two tests for rational expectations. Following this, in section 5 I develop and estimate a simple multinomial logit to explain the reported choice probabilities using subjective beliefs. Section 6 moves beyond the static setup to show a dynamic model; future versions of this paper will include more details on the estimation procedure, results, and counterfactual analysis. Finally, section 7 concludes with next steps.

2 Literature Review

This project builds on several different literatures in education choice, labor economics, and subjective expectations.

2.1 Education choice

Workhorse models of human capital accumulation include on-the-job training (Becker (1964)), supplementary training (Ben-Porath (1967)), and learning-by-doing (Heckman et al. (2002)). Higher education is the most straightforward example of human capital investment and one of the most studied. In the standard framework, individuals face a tradeoff between the costs of additional education, which include tuition, opportunity costs from lost wages, as well as psychic or effort costs, and the benefits to earnings that education brings during the rest of the lifecycle. Agents choose an optimal level of education based on this tradeoff and in the face of uncertainty.

2.1.1 College premia

Substantial evidence documents the college wage premium (CWP), taken as the average amount by which the wages of BA holders exceed those of high school graduates, typically on the order of 40-60% (Goldin and Katz (2008), Valletta (2016), Ashworth and Ransom (2019)). While some of this literature simply documents differences in earnings, research which seeks to control for the obvious selection problems still finds meaningful increases to earnings from attaining a college degree (Westphal et al. (2022)). Although the CWP literature has focused on 4-year degrees, there is significant evidence of a similar (though smaller) premium for community college (Marcotte et al. (2005), Belfield and Bailey (2011), Bahr et al. (2015), Jacobson et al. (2005)). This is usually on the order of 10-20% (Jacobson et al. (2005) estimate 9-12% for displaced workers), though Stevens et al. (2015) find for some degrees a premium of up to 45%. This is in line with the significant heterogeneity often found in the benefits of community college, in particular varying by gender (women benefit more than men) and by major (career-oriented programs, in particular healthcare and nursing, are far more beneficial than others).[3]

2.1.2 College puzzle

Although estimates of the premia depend on the type of degree and context, there is broad consensus that higher education provides a boost to earnings. As documented by Goldin and Katz (2008) and others, this premium was growing significantly from the 1980s through the 2000s, and remains substantial, even if it has recently flattened. Given this, the statistics presented in section 1, namely that only about 40% of young people have a BA, are puzzling and have spawned a literature examining the "college puzzle": given the wage premium for higher education, why do so many young people fail to earn a degree?

Undoubtedly there are many factors at play, and to date no one has done a precise decomposition of how much non-attainment is attributable to each. Dropout risk has emerged as a promising explanation; Hussey and Swinton (2011) argue that the substantial dropout rate (about 50% of students who start a 4-year degree fail to earn one within 6 years) means that the expected returns to *attending* college may be much lower than the expected returns to earning a college degree. This uncertainty creates a wedge: the expected value of having a degree exceeds the expected value of attending school, which lowers the income threshold at which individuals invest in education. Allowing for a heterogeneous dropout risk correlated with wages (as might be expected if both were influenced by the same unobservable, such as college preparedness at age 18) would further lower that threshold. This mechanism is further explored by Athreya and Eberly (2021), who show that policies aimed at increasing attendance, such as tuition subsidies, are unlikely to be successful at increasing attainment because the marginal students who are pushed into attendance are exactly those with a high risk of dropout.

Of course, dropout is not necessarily an exogenous event, and may instead be driven by endogenous choices students make as they progress through college. It is unfortunately difficult to find estimates for the percentage of voluntary versus forced exits from higher education, and in any case the line may be blurry: a student may realize they are performing very poorly as they receive grades and choose to leave rather than continue and risk being forced to exit. One of the better data sources available is from the Berea Panel Study, conducted by Todd and Ralph Stinebrickner. Stinebrickner and Stinebrickner (2008) focuses on the role of credit constraints in dropouts and Stinebrickner and Stinebrickner (2012) focuses on learning about ability; neither find a significant role for students being asked to leave because of poor performance or for a binding liquidity shock. Learning about ability then emerges as a primary channel for voluntary exits; besides Stinebrickner and Stinebrickner (2012), Arcidiacono et al. (2023) model student learning about ability, which

[3] See Jacobson et al. (2005), Belfield and Bailey (2011), Bahr et al. (2015), Stevens et al. (2015), Grosz (2020), Jepsen et al. (2014)

creates uncertainty around performance in college and future earnings and may lead to a rational choice to leave without a degree.

This draws out an important distinction between an *ex-ante* and *ex-post* analysis of dropout. If students face uncertainty about their ability and preferences, enter college, learn that it is not the right path for them, and then drop out, entering college in the first place was not a mistake *ex-ante*, even though *ex-post* it is clear they would have been better off not going. From the social planner's perspective, this form of dropout is not a problem and is in fact a net positive to the extent that individuals are updating their beliefs and making an optimal decision to leave. Inefficiency instead arises from information constraints, whereby everyone would be better off if young people were able to learn about their abilities and preferences without making a costly investment in college in the first place. The process of information revelation directly causes dropout, but the dropout decision itself is not problematic; rather, it is the existence of information deficits which is suboptimal.

Another channel for dropout might be broadly thought of as "adverse shocks," which include family- and own-emergencies, such as those related to a health crisis, job loss, or incarceration.[4] As DeLuca et al. (2021) discuss, this may be particularly important in the community college context, through a powerful selection effect: young people who face a high likelihood of adverse shocks may choose to go to lower-cost community college programs which have a lower labor market return but have limited downside in the event that the student has to leave unexpectedly. Here, as in the above case of learning about ability, the line between a voluntary exit and a forced dropout is not strictly clear, since many of the adverse shocks described in DeLuca et al. (2021) could, in principle, be solved with sufficient liquidity.

An especially large literature explores the role of financial constraints. Some of the best evidence on the role of credit constraints in dropout decisions comes from Stinebrickner and Stinebrickner (2008). They follow students from Berea College who have their tuition costs (and some living costs) entirely covered by the college. Still, dropout rates at Berea are high, and the authors interview students before and after dropout to determine the role of credit constraints. While they do find that many students would like to borrow money to smooth consumption, they find that the majority of dropout would remain even if credit constraints were entirely removed. Evidence on community college students in particular comes from Ortagus et al. (2021), who follow a large sample of academically successful community college students who leave with no degree. They find that financial considerations, for example an unexpected reduction in financial aid, contribute to the majority of dropout decisions. However this study is unable to verify the role these factors play, and it is possible that reporting bias plays a role.

[4]Mental health in particular may be an important factor in these kinds of shocks (Daniel et al. (2009)).

Shifting from dropout to the primary education decision, the literature which structurally models education choices usually finds that credit constraints do not bind, or that they only play a small role in the decision of young people to not attend college. Early models of this type include Keane and Wolpin (2001) and Cameron and Heckman (2001), neither of which find a large role for credit constraints; the latter argues that it is factors which correlate with credit constraints, rather than the constraints themselves, which drive the gaps in attendance between high- and low-income families. Johnson (2013) finds a similar result, and highlights a "buffer stock savings" style motive as in Carroll (1996): young people are averse to taking out large loans, perhaps out of caution for the possibility of unemployment and large future shocks. There are exceptions in the structural literature, such as Attanasio and Kaufmann (2009) and Kaufmann (2014), but this may have to do with their population (Mexican students rather than those in the US). On the other hand, there are several reduced-form studies which find positive impacts of increases in loosened financial constraints. Rau et al. (2013) and Solis (2017) look at the expansion of a loan program in Chile and find significant positive impacts on enrollment. Dynarski (2000) looked at the HOPE tuition program in the state of Georgia, and estimated that an additional $1000 of aid in 1998 dollars increased college attendance by 3.7-4.2 percentage points. Black et al. (2020) look at the impact of expanded loan access on currently enrolled students, and find (among other outcomes) a positive impact on completion, suggestive of a role for credit constraints on the dropout decisions of enrolled students.

The extent to which credit constraints bite undoubtedly depends on specific national and local policies and contexts. However, another explanation for this discrepancy arises from limited information and biased beliefs. Is it possible that based on policy, credit constraints *should not* bind, but that families believe they in fact do? In fact, there is significant evidence that young people (and their families) are uncertain about the costs of college, as well as about the future returns to a degree. Giustinelli (2022) provides an excellent review of expectations surveys in higher education; in general, it appears that beliefs about the costs of college the returns to a degree are heterogeneous and frequently systematically biased (Dominitz and Manski (1996), Betts (1996), Abbiati and Barone (2017); Boneva et al. (2022) looks at expectations for post-graduate degrees). Gong et al. (2019) find that students at Berea College have significant uncertainty about future earnings. There is also a large body of evidence looking specifically at the returns to *major choice* within college, which similarly finds heterogeneous and biased beliefs (this literature will be discussed below). One important explanation for the mismatch between individuals' information set and what is, in theory, an observable statistic is the large difference between *sticker* prices and actual *out-of-pocket* (OOP) prices. This is a discrepancy that has grown over time, as illustrated by figure 2.

As explained in Dynarski et al. (2022), there is a robust system of need- and merit-based financial aid

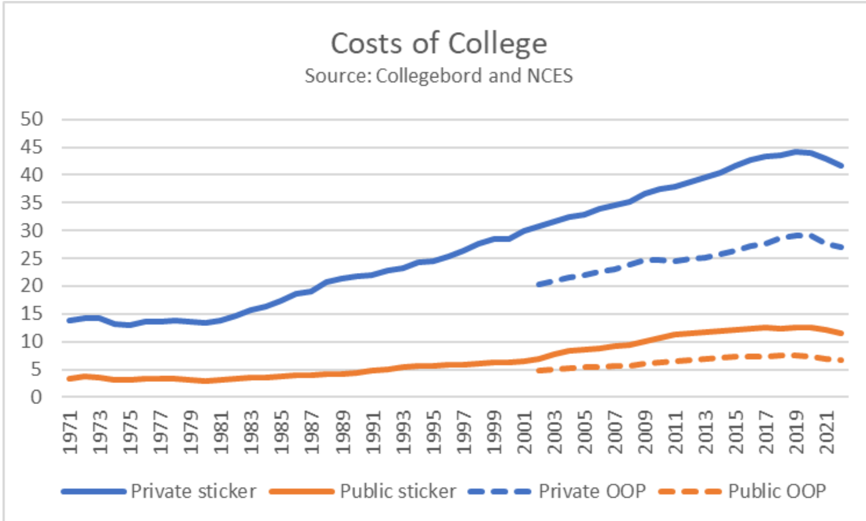


Figure 2: Sticker and out-of-pocket tuition costs

in the US which creates this divergence in price. Essentially there is significant cross-subsidization wherein some students paying the full sticker price offset the costs of lower-income (and other) students who receive significant aid. One analysis of NCES data from Kantrowitz (2017) estimates that only 25% of incoming college freshmen and 38% of all undergraduates pay the full sticker price of attendance. To the extent that the true OOP cost of attendance is unknown, young people may make their attendance decision based on uncertain and potentially biased beliefs which are anchored to the sticker price. Indeed, there is evidence that students respond to sticker price changes even when OOP costs are unaffected (Levine et al. (2022)).

A lack of information about costs is different from a difficulty in *processing* what is a highly complex system. Dynarski et al. (2021) show that guaranteeing aid up front, even when the amount of aid is unchanged, leads to an increase in application and attendance. Marx and Turner (2019) shows a similar result with respect to student loan offers. Help in filling out the famously-complicated Free Application for Federal Student Aid (FAFSA) form had significant impacts on attendance in Bettinger et al. (2012). A paper from Mulhern (2021) provided personalized information on the college process (focused on admission probabilities) via the Naviance software, and found significant (though heterogeneous) impacts which are particularly large for low-income students. This is an important distinction but one which is hard to make and has not been fully explored.

Several strands of the literature discussed above come together in research done on the choices and outcomes of high-achieving low-income (HALI) high school students. These students tend to apply to and matriculate at selective 4-year schools at a significantly lower rate than their higher-income peers, even though low-income students receive a larger premium from selectivity (Stacy Berg Dale and Krueger (2002)). In

Hoxby and Turner (2015) the authors provide targeted information about the probability of acceptance and costs of attendance at more-selective schools to HALI students, and find significant impacts on application and attendance behavior. Hoxby and Avery (2013) decomposes HALI students into "achievement-typical" students who behave like their high-income peers and "income-typical" students who behave like their low-income peers. They find that the former group is usually found in school districts with competitive magnet public schools and in districts with a critical mass of HALI students; from this they conclude that peer-effects and the impact of guidance counselors and college programs which help HALI students navigate the higher-education process are important. These findings fit in well with a narrative of limited information and biased beliefs driving the choices of young people.

Finally, there is substantial evidence that non-pecuniary enjoyment of higher education also plays a significant role in education decisions. In particular, it emerges in several expectations-based surveys, such as Wiswall and Zafar (2015) and Delavande and Zafar (2019) for 4-year students and Baker et al. (2018) for community college students, that the wage premium which economists tend to be focused on is sometimes dwarfed by non-pecuniary tastes for majors or courses of study. A deeper understanding of how young people value the elusively-measured experience of education, as opposed to the more easy-to-observe labor market returns, will be helpful in understanding education decisions.

Considering the role of information frictions, non-pecuniary returns (notoriously difficult to infer from data), and broader behavioral puzzles naturally brings us to the growing literature on subjective expectations, which provides a ready-made framework in which to understand several of the above-discussed phenomena.

2.2 Subjective expectations

Largely beginning with Manski (1993), economists began directly asking survey respondents their beliefs about uncertain outcomes. Manski (2004) provides an overview of the problem motivating this research: observed choice data may be compatible with many different combinations of preferences and beliefs, which has spurred economists to place strong assumptions on beliefs (typically assuming rational expectations) in order to estimate our models. If those assumptions fail, then estimates will be biased and any inference or policy analysis done with the model may not hold in the real world. The fundamental problem of separately identifying preferences from beliefs has given rise to a literature which uses solicited expectations and subjective beliefs to estimate structural models without the assumption of rational expectations. Koşar and O'Dea (2022) provides a thorough review of this literature, which has now been applied to a wide variety of economic problems ranging from labor market decisions (van der Klaauw (2012); Jäger et al. (2021); Blesch et al. (2023)), marriage and fertility (Delavande (2008); Miller et al. (2025)), consumer behavior (Blass et al.

(2008)), health (Wang (2014)), and, most relevant for our purposes here, education choices.

Dominitz and Manski (1996), one of the earliest surveys of subjective expectations, began a rich literature on the role of subjective beliefs in education choice. Kaufmann (2014), Abbiati and Barone (2017), and Delavande and Zafar (2019) look at higher education choice. Delavande et al. (2022) goes a step deeper and looks at the time use of students, who can divide between academic and non-academic activities, each of which have different expected returns. A particularly fruitful vein has been on the major choice of college students. Noting that different college majors have significantly different labor market returns, these papers tend to solicit students' beliefs about labor market outcomes conditional on majoring in different areas. This data can then be used to estimate preferences for different courses of study. In general, students have heterogeneous and inaccurate beliefs about earnings, revise in logical ways when presented with new information, and place a high weight on non-pecuniary enjoyment. Examples of this work include Arcidiacono et al. (2010) with students at Duke, Zafar (2011) with students at Northwestern, and Wiswall and Zafar (2015) with students at NYU. Wiswall and Zafar (2021) focus on gender differences and find that male and female students have different preferences for labor market returns, with female students prioritizing flexibility and stability and male students prioritizing earnings growth.

One strain of this literature combines subjective expectations data with information treatments and other experiments, with a nice overview presented in Fuster and Zafar (2023). The purposes of such interventions are usually twofold. The treatment may be interesting in its own right, for example to test methods for correcting biases; and it also may be used in identifying the causal effect of a belief, either in reduced form analysis or in a structural model. A primary example of the latter is Wiswall and Zafar (2015), which provides information about major-specific returns and labor market outcomes. In that paper, the information treatment with repeated survey questions before and after provides identification of the preference parameters for each major.[5] These treatments are not always effective; Dinkelman and A. (2014) find no impact on the enrollment behavior of Chilean 12$^{\text{th}}$ graders, and Bleemer and Zafar (2018) find that while information on returns is effective at encouraging enrollment, correcting misperceptions around costs is not. In Barone et al. (2017) a counseling program for Italian high schoolers is effective on the intensive margin of behavior among enrolled college students, but does not change the extensive margin of enrollment. Cunha et al. (2018) examine the rollout of Texas GO centers which provide peer-based counseling on the college process, and find a significant effect on application, a small effect on attendance, and no effect on completion.

This literature has been almost entirely silent on the beliefs about community college students.[6] The

[5] Other examples include Oreopoulos and Dunn (2013), Jensen (2010), Ehlert et al. (2017), and Peter et al. (2021), Berkes et al. (2022), among others.

[6] It is worth mentioning that the evidence on the beliefs of 4-year college students also skews heavily towards students from

exception is Baker et al. (2018), who surveyed community college students about the labor market returns to different majors. They find that beliefs are highly inaccurate, and that students in general overestimate earnings and underestimate employment probabilities. They provide an information treatment with specific information about the earnings of graduates from the subjects' own institutions, which causes the sample to generally revise beliefs in the expected direction. However, despite the majority of community college students reporting intention to earn a 4-year degree, this study did not look at beliefs around transfer and outcomes with a 4-year degree.

Most of the above cited work takes beliefs as given and does not explicitly model belief formation or the process of evolution over time, though there are examples of this in the literature (Erdem et al. (2005) model consumer learning about computers, Stinebrickner and Stinebrickner (2012), Gong et al. (2019) and Stinebrickner and Stinebrickner (2013) examine learning and resolution of uncertainty during college, and the literature on ability such as Arcidiacono et al. (2023) is explicitly focused on this process). In general, the macro literature seems to be more focused on modeling learning in dynamic models than the micro literature, and Koşar and O'Dea (2022) explicitly calls this out as a promising direction for future research in microeconomics. The macro literature has largely focused on learning about macroeconomic variables such as inflation and proposed several alternate expectations models such as sticky expectations (Carroll et al. (2020)), noisy information (Wang (2024)), diagnostic expectations (Bordalo et al. (2018)), and epidemiological expectations (Carroll and Wang (2023)).

3 Data

3.1 Context

Survey data were collected at Anne Arundel Community College (AACC), in Maryland's Anne Arundel County. Anne Arundel County is home to the state capital of Annapolis, is relatively close to Washington D.C., and as a result is somewhat more prosperous than the state as a whole. Per-capita personal income is \$79,000, compared to \$71,000 for Maryland, making it the 6$^{\text{th}}$ most prosperous county (out of 25). It also has the 4$^{\text{th}}$ lowest poverty rate, at 5.8%, compared to 9.6% for the state at large. It is just slightly more educated than Maryland, with 45% of residents holding a BA or more compared to 43% statewide.[7] Anne Arundel county is one of the more diverse counties in Maryland, but regional heterogeneity means it is still slightly less diverse than the state as a whole: 65% identify as white compared to 50% in Maryland, 18%

elite institutions.

[7]This masks significant regional heterogeneity: while 64% of Howard County residents hold a BA or better, only 35% do in urban Baltimore City and just 17% in rural Somerset County.

identify as black compared to 30%, and 10% as Hispanic compared to 12%. Among Maryland's 24 counties, Anne Arundel ranks 4th in Hispanic population, 10th in Black population, and 5th in Asian population.

Anne Arundel Community College (AACC) is the third largest community college in Maryland, with around 12,000 students enrolled. 73% are part-time, which is close to the statewide figure of 70%. It is the 5^{th} most Hispanic, 8^{th} most black, and 10^{th} most white community college. As is typically the case in higher education, Maryland community colleges skew heavily female, with only 38% male students statewide and at AACC.

AACC itself presents as a fairly successful community college among its MD peers. It has the 5^{th} lowest rate of remediation needed, at 36%, compared to 48% statewide. It has one of the lowest rates of Pell grant recipients, which are awarded to 23% of its students, compared to 33% across all MD community colleges. It has the 3^{rd} highest 4-year transfer rate: 32% of students transfer to a 4-year program within 4 years of enrolling, compared to 29% overall. Considering 4-year transfer and 2-year graduation together as student "success", AACC has the 6^{th} highest 4-year success rate.

On the whole, Anne Arundel Community College is a good context in which to conduct this research. While slightly more educated and slightly wealthier than the state, Anne Arundel county is neither particularly poor (like Baltimore City) nor especially wealthy (like Montgomery County). It is racially and ethnically diverse and has a mixture of working class and more highly educated residents, as well as urban, suburban, and rural areas. The community college is fairly successful, but does not dramatically surpass its peer institutions. With this context in mind, I designed a survey to capture students' subjective beliefs about education pathways.

3.2 Survey

The survey was administered online via Qualtrics software from November 21^{st} to December 7^{th}, 2024. The survey was open to all AACC students who met the following criteria:

1. Between 18 and 32 years old
2. Pursuing a 2-year degree
3. Not in a dual enrollment (high school) program

The survey was distributed via the school's Canvas system. Canvas is a learning management system where students access their course materials, such as lecture notes and quizzes, receive messages from professors, and submit assignments. The survey was posted as an announcement to every student's Canvas page so that they would see it when they logged on to their account. Respondents were offered a $10 Amazon gift card in exchange for what was described as an approximately 30 minute survey.

The survey design suggests potential selection bias along a few relevant dimensions. First, since the survey is distributed via Canvas, it is reasonable to think that students who spend more time on their Canvas accounts would be more likely to take the survey; this would include more diligent students (more likely to have a high GPA) as well as students taking a heavier courseload (more likely to be full-time). Second, the survey is relatively long, which may skew the sample towards non-workers and non-caregivers who are more likely to have the free time to participate. Finally, the $10 incentive may skew the sample towards younger and lower income students who value the gift card more than their peers. As I will show, I do find evidence of selection along several of these dimensions; however, such bias (or worse) would be likely to exist if the survey were distributed via school email addresses (which would suffer from many of the same issues) or school survey recruitment lists.[8]

The survey consisted of two main components. The first focused on collecting demographic and background information on respondents. This included questions about race, ethnicity, gender, age, and other fixed characteristics. It also included questions about background, such as family income and education during adolescence. Finally, it asked about their previous educational experiences, such as high school grades, SAT scores, and grades during previous semesters of community college enrollment.

The second component, and the heart of the survey, consisted of a series of questions on the potential paths students might take through higher education and their expected outcomes from each choice. Subjective beliefs are centered around three primary factors: the cost of attending a 4-year college, graduation probabilities from community college and 4-year programs, and the labor market outcomes (employment and earnings) conditional on three levels of education. The basic format for the questions consists of placing students in one of three mutually exclusive hypothetical situations:

1. Leaving community college with no degree
2. Earning a community college degree and entering the workforce
3. Transferring to a 4-year school, with or without finishing a 2-year degree

While separate probabilities are elicited for transferring with and without first earning a 2-year degree, expected outcomes are elicited without conditioning on 2-year receipt. The assumption here is that conditional on transfer to 4-year school, students do not view holding a 2-year degree as influencing outcomes. This is admittedly a strong assumption, and one which I will further explore in the second round of survey

[8]Nothing like this is available at AACC, but such lists have been used in previous work such as Wiswall and Zafar (2015).

work.

In each case, I elicited beliefs about expected work hours and earnings at three different points in the life cycle: 5 years, 10 years, and 20 years after completing higher education. Each scenario also has beliefs questions which are specific to that education path. For example, in the case of transferring to a 4-year school, I ask about the likelihood of then graduating with a 4-year degree. In the case of leaving community college with no degree, I ask students to consider the various reasons why they might do so and how likely each is.

An important feature of the survey is that for several questions I elicited both student beliefs about their own outcomes as well as beliefs about outcomes for the broader population of students from AACC. Students' own-beliefs are the relevant beliefs to understand their decisions and to estimate the education choice model, but beliefs about the broader population can be more easily used to measure biases as these can be benchmarked against data.

I received 1,492 attempted survey responses, with succesful submissions from 446. After filtering using the results of two built-in attention checks and a few other filters for reasonable responses, I end up with a final sample of 319 high quality responses. More details about the process for cleaning the sample may be found in appendix A.

4 Reduced-form results

4.1 Descriptive statistics

I check the representativeness of my sample against registrar data on the full population of AACC students who meet my filter criteria. There are 6,888 such students, meaning that my sample represents 4.6% of the total population. Table 1 compares the full sample (including those who failed attention checks), cleaned sub-sample, and the full student body along several critical dimensions.[9] There are no significant differences between the full sample and the clean subsample, however there are a few significant differences between the population and sample which deserve mention.

First, I oversample female students. While 60% of the population is female, my sample is 67% female. Excluding those who identify as nonbinary (not a classification kept by AACC), my sample is 70.5% female and 29.5% male.[10] I also undersample part-time students, who make up 62.4% of the population but just 42.6% of my sample. Differences in PT/FT status may in part be due to measurement error; I define PT

[9] A full table of demographic characteristics can be found in appendix A.

[10] 4.4% of my sample identify as nonbinary, a statistic in line with other recent survey data.

Variable	Population	Sample	CI (sample)	Full sample	CI (full)
Female (norm.)	59.9	70.5	(65.2, 75.8)	71.2	(66.3, 76.1)
Male (norm.)	40.1	29.5	(24.5, 34.6)	28.8	(24, 33.6)
Age	21.49	21.92	(21.5, 22.4)	22.02	(21.6, 22.4)
Mean CC GPA	2.81	3.35	(3.15, 3.41)	3.3	(3.25, 3.36)
Part-time-school	62.4	42.6	(37.2, 48.1)	45.4	(40.7, 50)
White	49.5	53.8	(48.3, 59.2)	50	(45.3, 54.6)
Black	20.3	19.1	(14.8, 23.4)	21.7	(17.9, 25.6)
Hispanic/Latino	17	15	(11.1, 19)	16.3	(12.8, 19.7)
N	6888	319	-	443	-

Table 1: Sample demographics

status as taking 12 or fewer credits, following a common definition. However, some students may not be sure how many credits they are taking. Students may take up to 4 separate 3-credit classes, which would make them technically PT students for my purposes (and those of the registrar who provided the population data) but might reasonably consider themselves FT. That said, as discussed in section 3, the distribution method makes it reasonable to expect some selection from full-time students.

This mechanism may also be at play when looking at GPA. It should be noted that this statistic is only relevant for continuing students, since first time students in the fall of 2024 could not report a cumulative GPA. We see that the sample has a higher GPA by about 0.5, representing half a letter grade on average. Interestingly, the difference is larger for male students than female students, suggesting more significant selection on ability for male students than female students. In fact, while in the administrative data female students have a higher GPA than male students, the opposite is true in my sample (though not statistically significant). While this selection on ability is in line with expectations, there are likely several other factors at play. I expect measurement error, where students who are unsure of their GPA give a best guess (evidenced by the majority of responses being precise to the first decimal place or integer). While this on its own may not systematically bias the mean, I also expect some positive response bias, where students may inflate their GPA when answering the survey. In the future, I hope to match my survey records to the school's administrative data to validate self-reported GPA.

On several other metrics the sample appears close to the population. The mean age in my sample is 21.5, close to the mean age of 21.9 in the population (although just outside the 95% CI). The racial composition also appears similar, although some caution is necessary here. In my survey I adopted the methodology of the 2020 census, which allowed respondents to choose all racial categories with which they identify. I did not ask respondents which racial identity they consider most *salient*, which could have been used to do an apples-to-apples comparison to the AACC metrics (which include just one racial group per respondent). As a result, I normalize the elicited racial categories in order to compare directly; details can be found in appendix A. The result is that my sample has a similar proportion of Black, White, and Hispanic/Latino respondents as the population at large.

My sample also appears similar to the population in terms of major choice. Table 2 shows the proportion of the population and sample in 6 major categories. Here again some caution is needed, as I manually classify majors in the administrative data into the six categories I included in my survey.[11] For example, although I classify "teacher education" as a major in arts and humanities, it is not inconceivable that a respondent might select social science, as this degree does not fit neatly into any one category. In addition, students who the college considers to be majoring in "general exploration" and "custom transfer" (categories of general study for students who do not yet know their major or do not plan to declare one prior to transferring to 4-year school) may consider themselves in a specific major that they have not registered with the school administration. Similarly, students may technically major in "transfer studies" (which has some advantages when transitioning from community college to a 4-year program) but firmly consider themselves as majoring in STEM or health. Still, the proportions suggest that the sample is fairly representative in terms of major choice.

Major	Population	Sample	Full.sample
Arts and humanities	14.1	12.2	12.9
Business and economics	12.5	13.5	13.1
Medicine and health	22.9	26.3	27.5
STEM	22.3	27	25.5
Social science	7.8	6.3	5.9
Transfer studies	20.4	14.7	15.1
N	6888	319	443

Table 2: Majors

A joint F-test between the student body and my sample using gender, age, and race fail to find significant differences ($p = 0.227$). Adding in part-time status leads the test to find significant differences ($p < 0.01$) but, as mentioned above, there may be more measurement error with attendance status than other variables. Joint tests between my final sample and the larger, pre-cleaning sample on the same vector of characteristics (including PT enrollment status) fail to find any significant differences ($p > 0.99$).

Ultimately, the most that can be said is that the estimates presented in the rest of this paper will be internally valid for my survey sample, and there is surely some evidence of selection along gender, full-time status, and GPA. However, the sample is representative in terms of race and area of study, leading me to believe that it is fairly representative of the student body at AACC. Future work will allow me to better control for selection by evaluating realized outcomes.

[11] Categories are taken from the College Board.

4.2 Expectations

I now turn to examining the heart of the survey: data on expected outcomes.

4.2.1 Earnings

Respondents were asked to report expected full-time earnings 5, 10, and 20 years in the future when working with a high school degree only, a 2-year degree only, and a 4-year degree. They are specified to ignore inflation and respond in real terms (that is, in 2024 dollars). Specifically, they are asked a question of the type:

> *Consider the case where you are working full-time with a 2-year degree as your highest degree. What annual salary do you expect to earn 5 years from now? That is, what is the most likely salary?*

Answers are given as a positive integer. Respondents are then asked a follow-up question, based on their response:

> *You listed your expected salary while working full-time with a 2-year degree 5 years from now as X, but it is possible you may earn more or less than this amount. What is the percent chance that your salary falls into each of the following ranges?*

They are then prompted to drag sliders for five different ranges of possible salary:

1. Less than \$[0.7×response]
2. Greater than or equal to \$[0.7×response] and less than \$[0.9×response]
3. Greater than or equal to \$[0.9×response] and less than \$[1.1×response]
4. Greater than or equal to \$[1.1×response] and less than \$[1.3×response]
5. Greater than \$[1.3×response]

The total across all sliders is forced to sum to 100.[12]

Eliciting the PDF is useful in two ways. First, it can be used directly to measure the uncertainty individuals face about any continuous variable. If we want to use the mean inside any non-linear function, being able to approximate the distribution is critical. Second, eliciting the density allows me to construct a second measure for the expected mean, as well as the median, mode, and other moments. This is important, because it is not always clear which moment respondents have in mind when they answer a point-expectation

[12] Alternatively, I could have allowed the distribution to be unrestricted and then normalized to 100 afterwards.

question of the sort described above. As Engelberg et al. (2009) detail and Armantier et al. (2016) helpfully review, some respondents may answer with a mean, some with a median, and still others with a mode. Instead of using the point-expectations directly, I follow these authors in fitting parametric distributions to the elicited densities, and then use the density-implied moments rather than the directly elicited point expectations.[13] Details are given in appendix B. Figure 3 shows the relationship between elicited and implied mean earnings, pooled across degree type and time horizon. These have a correlation coefficient of 0.837, and display little systematic bias upward or downard (across 2,871 measurements, in 49.5% of cases the implied mean is less than the elicited mean, in 5.6% of cases it is equal, and in 44.9% of cases it is greater).

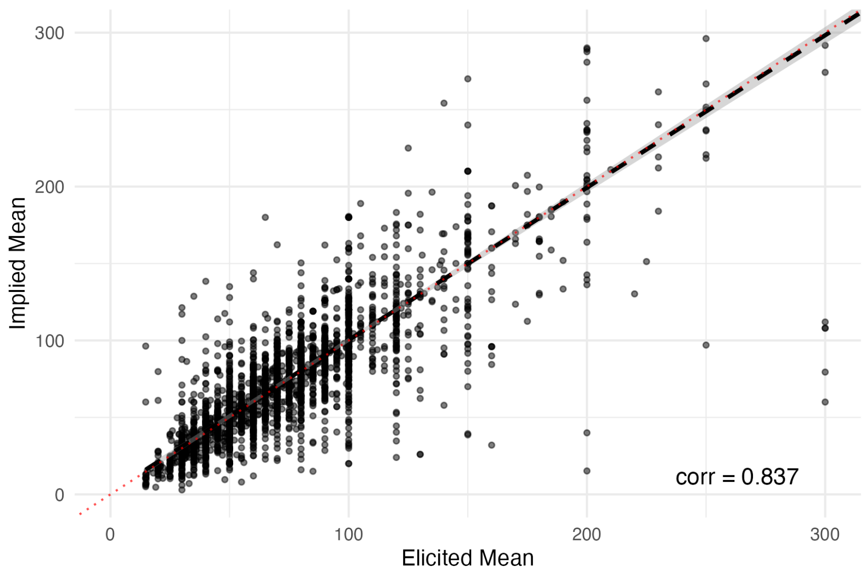


Figure 3: Elicited vs. Implied means

Respondents expect significant premia to both 2- and 4-year degrees over a high school degree. They also expect earnings to rise over time, with the largest increases coming in the first 5 years after graduation, resulting in a familiar concave earnings path. Figure 4 shows average expected earnings with each highest degree 5- 10-, and 20-years into the future.

Respondents were also asked questions of the following type about *population level* earnings:

> *We would now like you to consider a typical, 25-year old former AACC student who left with no degree and is working full time with a high school degree or equivalent. What annual salary do you expect them to earn?*

Answers were given as integers. Due to time and space constraints, I did not collect a density for

[13] An alternative would be to use the elicited density to set non-parametric bounds for the moments of each distribution, and perform any analysis with respect to those bounds.

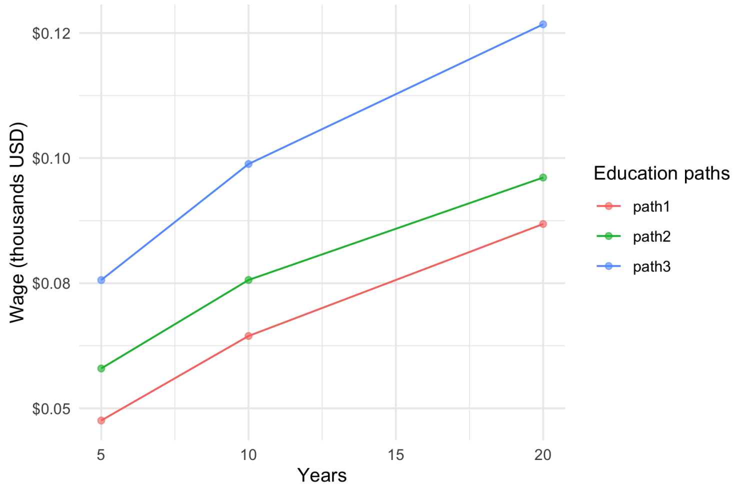


Figure 4: Expected wages at fixed distance in the future

the population level measures, and so will not transform them in the same way as the individual level expectations. These questions were also only collected for age 25 and 40, and so provide 6 measurements (rather than 9) per person, for fixed age ranges. Despite these caveats, they provide an interesting benchmark against which to compare the subjective expectations of own earnings.

Figure 5 shows the expected wage path by age, for each highest degree, in transparent bubbles representing the sample size at each hypothetical age. The crosses at ages 25 and 40 show the median population-level earnings beliefs, corresponding to questions of the above kind. The dashed lines show median full-time annual earnings for MD workers, taken from the CPS (details can be found in appendix C). Of course, there are many reasons why looking at the difference in population-level median earnings for workers with each type of degree clearly does not provide an unbiased measurement for the same statistic that survey respondents report. Most obviously, respondents describe their own earnings in different scenarios, and so in theory the well-known issues of selection are controlled for and the difference in expected earnings represents the expected treatment effect of the degree. Furthermore, respondents are making a projection into the future, and may be imagining changes in macroeconomic conditions. However, it still may be helpful to benchmark survey expectations against some objective data, if only because we might imagine survey respondents forming expectations to be influenced by exactly these numbers. Median population-level beliefs are closer to the CPS data the own-beliefs, in particular at age 40, but are still significantly higher. In general, the shape and ordering of subjective earnings trajectories mimics that of the population-level medians from the CPS, but with a sizable upward level shift.

The dip at age 35 in figure 5 reflects the cohort effects mentioned above: a respondent who is currently

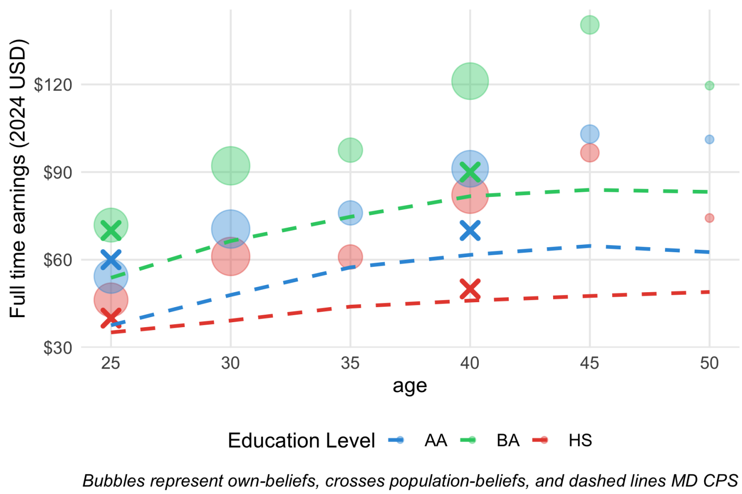


Figure 5: Expected wages by age

age 30 may have a lower expected age 35 wage than a respondent who is 20. This suggests that blending cohorts may obscure underlying trends, and that considering younger and older respondents separately may be helpful. I do so in table 3, which breaks down expected earnings at 5, 10, and 20 years into the future for the "young" (who are 18-22) and the "old" (who are 23+). Older respondents expect higher earnings with an AA or BA, but lower earnings with a high school degree only. Interestingly, young students expect that earnings for HS and AA degrees converge after 20 years, a trend not seen in the older cohort.

Years out	**All**			**Young**			**Old**		
	HS	AA	BA	HS	AA	BA	HS	AA	BA
5	40.9	56.0	70.8	40.0	51.8	68.1	45.7	62.9	80.0
10	54.6	69.7	85.6	53.1	64.2	82.2	55.4	75.0	92.0
20	64.3	80.8	99.5	64.1	78.8	98.0	65.3	92.2	106.4

Table 3: Median expected earnings

Rather than focusing on levels, we can also look at the median expected wage premium, measured as the percentage by which earnings with a 2-year degree exceed earnings with a HS degree, and earnings with a 4-year degree exceed those with a 2-year degree. Table 4 shows the median wage premium for the full sample and broken down into young and old cohorts. The older cohorts expect higher premia than the younger cohorts, in particular for the returns to a 2-year degree. In table 5 I benchmark the survey expectations against the population-level figures from the CPS, once again making the caveat that if beliefs were rational we would *not* expect these two numbers to align. Still, it is nonetheless interesting to note a rough alignment in these figures, which arises because although the *level* of expected earnings is very different from the CPS, the ordering and shape are not.

While these averages provide evidence that responses are broadly reasonable, they obscure widespread

Years out	All		Young		Old	
	AA	BA	AA	BA	AA	BA
5	0.30	0.70	0.23	0.69	0.44	0.73
10	0.25	0.61	0.18	0.57	0.36	0.72
20	0.22	0.62	0.20	0.57	0.30	0.77

Table 4: Median expected wage premia

Age	AA premium		BA premium	
	Survey	CPS	Survey	CPS
23-27	0.23	0.17	0.30	0.29
28-32	0.21	0.19	0.27	0.32
38-42	0.23	0.24	0.28	0.35

Table 5: Expected wage premia

heterogeneity in expected returns. Figure 6 shows the distribution of wage premia for AA and BA degrees over HS only, averaged across the 5-, 10-, and 20-year time horizons. The expected BA premium is centered to the right of the AA premia, indicating a larger expected premium, but is much wider, indicating higher variance in the expected returns to a BA than to an AA.

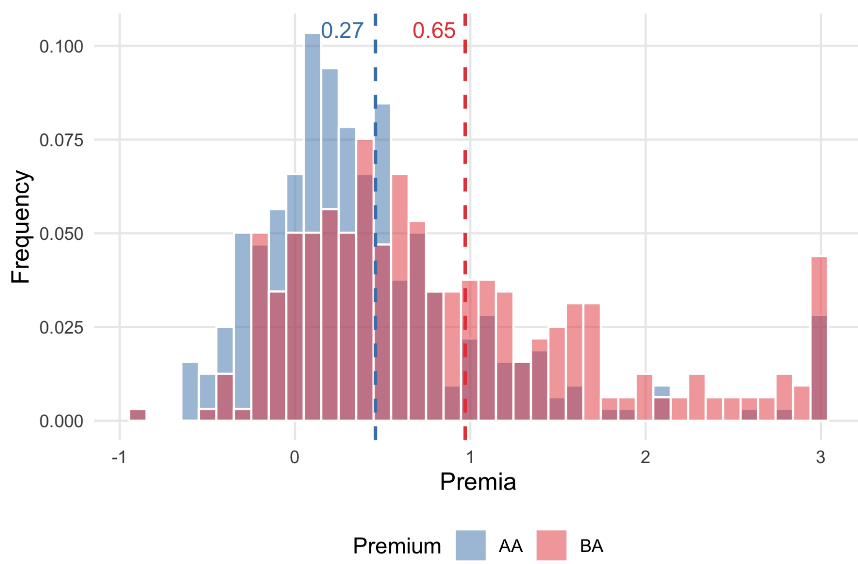


Figure 6: Wage premia for AA and BA over HS

The expected wage premium may reflect many different factors not strictly related to the value of a degree. In order to better understand the population level variation in degree premia beliefs, and to check that the correlations in my data appear sensible, I estimate variations on the following model.

Let $\mathscr{E}_i\left[Y_{i,s,d}\right]$ be the subjective belief of person i about their annual, full-time earnings at age s if they were to hold highest degree d. I estimate the following model:

$$\log\left(\mathscr{E}_i\left[Y_{i,s,d}\right]\right) = \beta' X_i + \beta_s s + \beta_s s^2 + \alpha_d + \epsilon_{i,s,d} \tag{1}$$

where α_d will be the degree-specific effect on the log of expected earnings, i.e. the expected premium that individual i holds about their age-s earnings with degree type d. X_i is a vector of personal characteristics which do not vary. Table 6 presents the OLS results of eq. (1) with a full set of controls.

Table 6: Log expected earnings

	Dependent variable:
	log(earnings)
Age	0.086*** (0.012)
Age^2	−0.001*** (0.0002)
Associate Degree	0.239*** (0.025)
Bachelor Degree	0.508*** (0.025)
White	−0.047** (0.022)
Male	0.113*** (0.023)
Parent had BA	−0.032 (0.022)
FARM status	0.035 (0.023)
STEM+ major	0.119*** (0.023)
log(HS GPA)	0.144*** (0.046)
Returning student	−0.177*** (0.053)
Currently working	0.113 (0.069)
log(college GPA) × returning	0.020 (0.042)
log(hourly wage) × working	−0.014 (0.023)
Constant	1.853*** (0.220)
Observations	2,871
R^2	0.247
Adjusted R^2	0.243
Residual Std. Error	0.540 (df = 2856)
F Statistic	66.807*** (df = 14; 2856)
Note:	*p<0.1; **p<0.05; ***p<0.01

The expected AA premium is 0.24 and the expected BA premium is .51, both of which are precisely estimated. The data correlations in other characteristics, for the most part, align with reasonable priors; men, students in STEM+ majors (defined as STEM, economics and business, and health), and those with higher grades have higher expected wages. Although statistically insignificant, the two variables related to adolescent SES may be surprising; having a parent with a BA or higher is associated with *lower* expected earnings, and having received free or reduced-price lunch as a child is associated with *higher* expected wages. In a similar vein, being a returning student (as opposed to being in one's first semester) is associated with lower expected earnings, as is being a white student (despite population-level data showing a persistent racial earnings gap).

These estimates are good reminders of the unusual nature of this data, which *beliefs*, and thus subject to behavioral and informational biases. For example, it is quite possible that first-generation students have more limited access to information about the college wage premium, and thus tend to overestimate the returns to a degree. At the same time, these estimates also reflect the population of community college students, and thus also reflect the process of selection into community college that occured prior to taking the survey. For some students, such as those from low-income backgrounds or those who struggled in high school, attending community college may be evidence of motivation and persistence which give good reason to expect high wages in the future; on the other hand, for some students (such as those from high income backgrounds), attending community college may be viewed as a failure and indicative of unobservables correlated with lower future earnings. Overall, caution is called for when interpreting these results, but I take them as further evidence that patterns in the data follow sensible trends, indicating meaningful responses to my survey questions.

4.2.2 Employment

Parallel to the questions about earnings, I elicited expectations of discrete labor supply and unemployment. Respondents were asked questions of the following form:

> *Assuming your highest degree is a 2-year degree, what is the percent chance that you will work the following hours 5 years from now?*

Respondents then dragged sliders to allocate 100 percentage points between:

1. No work (voluntary)
2. Unemployed
3. Part-time
4. Full-time

In separate instructions, the distinction was made between being out of the labor force (option 1) and being unemployed (option 2). Part-time is defined as working about 20 hours per week, and full-time as working about 40 hours per week.

Table 7 reports the mean probability of each labor force status, 5-, 10-, and 20-years in the future. Perhaps unsurprisingly, given that this is a sample actively investing in their education, respondents report an extremely high probability of working a positive number of hours, assigning low probabilities to both

Years out	**High school**				**2-year**				**4-year**			
	0	Un.	PT	FT	0	Un.	PT	FT	0	Un.	PT	FT
5	0.04	0.05	0.20	0.72	0.02·	0.03*	0.19	0.76	0.01***	0.03**	0.17	0.8**
10	0.03	0.03	0.12	0.82	0.02	0.03	0.14	0.82	0.02	0.01**	0.14	0.83
20	0.04	0.03	0.11	0.82	0.04	0.02·	0.14·	0.8	0.05	0.01***	0.15*	0.79

Table 7: Mean LS probabilities by highest degree

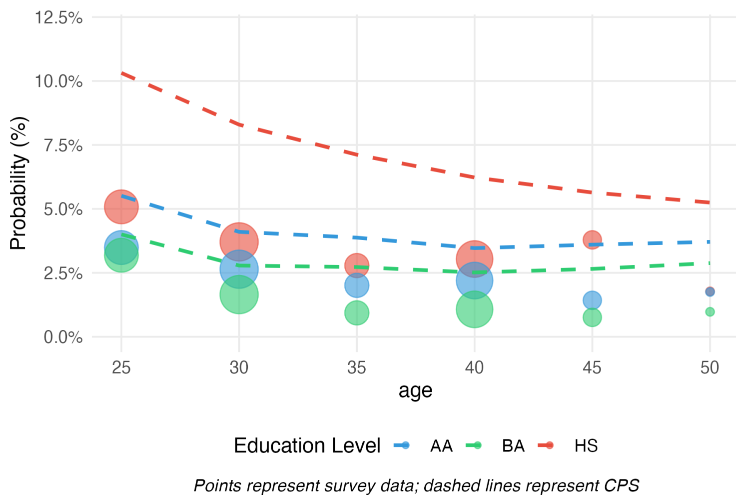


Figure 7: Probability of unemployment

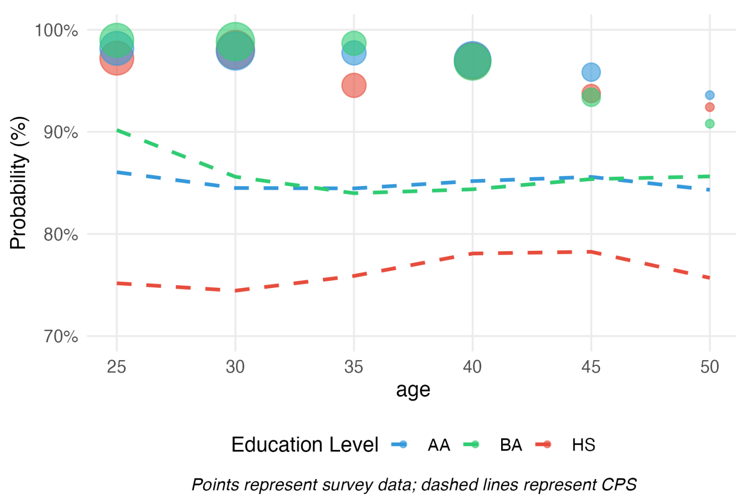


Figure 8: Probability of labor force participation

unemployment and to non-participation. Significance levels are given for 2-year and 4-year probabilities with respect to the corresponding HS probability.

As with earnings, we can also look at expectations by age, and once again I will here benchmark the subjective beliefs against population-level data from the CPS. Figures 7 and 8 show the mean probability of unemployment and labor force participation respectively. The expectations data respects the ordering of unemployment probabilities by degree, as well as the slightly declining level over time, but are significantly lower than the unemployment rate in the population at large. Expected labor force participation is much higher than that seen in the CPS.

To some degree, the sample means obscure the degree of certainty in labor supply expectations; the median probability of both unemployment and non-participation is in fact 0, across degrees, time horizon, and sample. However, statistically significant differences in the distribution of beliefs (in particular between unemployment and non-participation when holding a 4-year degree versus a high school degree) suggest that this margin may still be relevant in motivating people to pursue a degree.

4.2.3 Transfer variables

Finally, two variables were elicited about the transfer pathway.

Tuition costs

First, respondents were asked about their expected tuition costs. There are ways in which it may be surprising to consider tuition costs as a random variable, since many aspects of it are in theory easily observable; prospective transfer students can look up the sticker price of their desired program online, can research the parameters of financial aid programs, and can use calculators and other resources from the federal government. On the other hand, respondents face uncertainty about what (if any) *merit based* financial aid they will be eligible for, and they also may be uncertain about what type of school they will ultimately attend. The details of the financial aid system, which mean out-of-pocket costs are frequently much lower than the sticker price schools advertise, may be difficult to parse for students from low-SES backgrounds.

Respondents are asked for the probability, in the event that they transferred to a 4-year program, that they would attend an in-state public, out-of-state public, or private school. The overwhelming majority of respondents assign the highest probability to in-state public (86%), with smaller portions thinking it most likely they attend out-of-state public (8.8%) and private (5.2%). For whichever type of school respondents thought it most likely they would attend, they are asked for their expected annual tuition cost, with specific instructions to not consider room, board, and other costs of living; this is elicited both as a point expectation and as a density in a similar manner to earnings. As with earnings, I fit parametric distributions to these expectations in the manner described in appendix B and will primarily work with the implied mean of those distributions. Also as with earnings, I elicit expected tuition cost, at the same type of school, for the *typical* transfer student.

Figure 9 shows the distribution of own-expected tuition costs at public, in-state schools. The 2025-2026 tuition sticker price of $10,400 at the University of Maryland is shown as a benchmark against which to compare beliefs.[14] Although this should provide an *upper bound* for tuition costs, with financial aid and scholarships meaning the out-of-pocket cost should be weakly smaller, the majority of respondents report an expected tuition cost higher than this figure, with a median of $20,000 and a mean of $25,600.

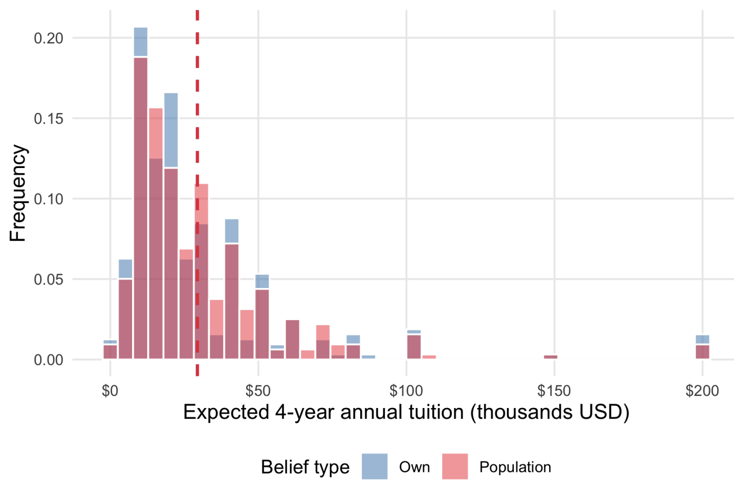


Figure 9: Expected annual tuition cost at 4-year school

Interestingly, the majority of respondents do not seem to consider heterogeneity in tuition costs, and those who do display little systematic relationship to SES or current grades. When examining the ratio of own- expected tuition to population- expected tuition, both the mean and the median are 1, with roughly equal numbers on either side. 26% of respondents expect their tuition to be more than 2% greater than the population at large, 31% less than 2% smaller, and the remaining 43% expect their tuition to be between 98% and 102% of the population level.

Graduation probabilities

The second transfer-related expectation elicited was the probability of successfully earning a degree, conditional on transferring to a 4-year program. This measures the overall probability of graduation, which includes both the risk of non-voluntary dropout and of choosing to leave voluntarily. This variable was

[14]This figure is comparable to other Maryland state schools. Including mandatory fees increases this figure to about $12,000, although respondents were not directed to consider fees.

collected both for oneself, and for the typical transfer student from AACC.

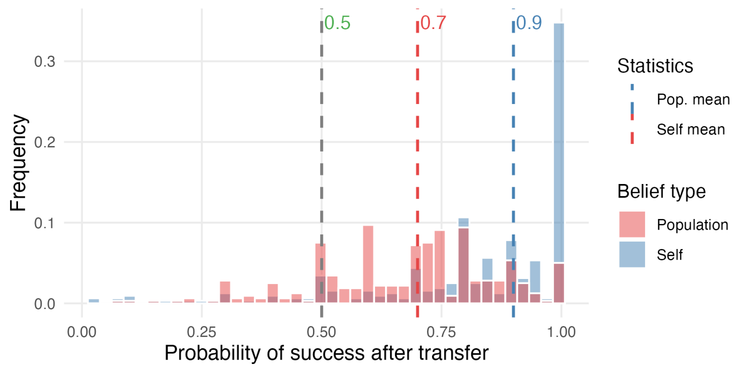


Figure 10: Successful attainment of BA after transfer

Results from these questions are shown in figure 10. The blue bars give own-beliefs, which skew clearly to the right of the population beliefs (in red); at the median, respondents believe that they are about 15 percentage-points (20%) more likely to earn a BA than the "typical" transfer student, and there is a sizable minority of respondents (35%) who report a greater than 98% chance that they will succeed after transfer. Still, even the population-level beliefs skew to the right of the data; analysis of National Student Clearinghouse data from Tatiana Velasco, John Fink, Mariel Bedoya-Guevara, Davis Jenkins & Tania LaViolet (2024) at the Community College Research Center shows that in the 2015 cohort, 50% of MD transfer students successfully completed their BA, just 2 percentage points below the US rate of 48%.[15]

The presence of a clear, objective benchmark for student beliefs in this particular instance makes it a tempting case for trying to say something rigorous about rationality and bias. This brings us naturally to the last section of descriptive evidence: tests for rational expectations.

4.3 Tests for rational expectations

Given data on a population's beliefs about a random variable, $\mathscr{E}_i\left[X_i\right]$, and realizations of that random variable, X_i, there are several tests available for whether or not expectations are rational. At their simplest, these are simply tests of whether the means of the distributions align, with the intuition that although it may be perfectly rational that any individual $\mathscr{E}_i\left[X_i\right] \neq \mathbb{E}\left[X_i\right]$, it is not possible for expectations to be rational and for $\mathbb{E}\left[\mathscr{E}_i\left[X_i\right]\right] \neq \mathbb{E}\left[X_i\right]$. Given the nature of my data, testing for rationality is an obvious temptation.

However, great caution is needed before proceeding with such a test and declaring whether a group's beliefs are "rational" or not. Any test for rational expectations (RE) is conditional on the objective distribution of

[15]Long lag times are generally necessary for such statistics, because a student who intends to go to school part-time for two semesters per year for the duration of their degree would take 8 years to complete.

X_i truly being the correct benchmark against which to compare the subjective beliefs $\mathscr{E}_i[X_i]$. For example, it may not be appropriate to test whether the subjective beliefs of my respondents with respect to age 50 earnings with a college degree are rational by comparing them to the empirical distribution of earnings of full-time college educated workers in 2025. Survey respondents may be making projections into the future about macroeconomic conditions, and may be accounting for unobserved heterogeneity which will not be properly controlled for in such a comparison. Concluding that their expectations are not rational because those means do not align would be highly dubious.

The two sets of expectations related to transfer outcomes present the clearest opportunity in my data to test for rationality in an environment in which we can be reasonably confident that student expectations should align with some measurable benchmark.

I first focus on the probability of successfully attaining a BA after transferring to 4-year school. In the case of a binary variable, rational expectations simply implies that the mean of population expectations matches the observed success rate, the testing of which collapses to a simple t-test. Defining $X_i \sim Bernoulli(p)$ as a random variable which is 1 if transfer student i successfully graduates, and 0 if not; in Maryland, the observed mean is $\mathbb{E}[X_i] = 0.5$.

If my sample were representative of the overall population of transfer students, and their beliefs were rational, then it would have to be the case that $\mathbb{E}[\mathscr{E}_i[X_i]] = 0.5 = \mathbb{E}[X_i]$; some respondents may expect a likelihood higher, and some lower, but on average their beliefs would equal 0.5. Testing that the mean of these subjective beliefs is 0.5 yields a p-value less than 0.001.

However, if my sample were selected based on traits which predict a higher-than-average likelihood of success (and in fact, my respondents indicate that this is the case by assigning a higher probability of own-success than population-success), then we would not expect that $\mathbb{E}[\mathscr{E}_i[X_i]] = \mathbb{E}[X_i]$ and rejecting this equality would not tell us anything about the rationality of beliefs. However, I also ask respondents about the probability that the typical transfer student will graduate, meaning that I directly elicit their subjective belief of $\mathbb{E}[X_i]$, which I denote $\mathscr{E}_i[\mathbb{E}[X_i]]$. $\mathscr{E}_i[\mathbb{E}[X_i]]$ may be different from $\mathbb{E}[X_i]$ because of measurement error in the survey or information frictions, but if the population as a whole has an unbiased view of the true distribution of X_i, then it should be the case that:

$$\mathbb{E}[\mathscr{E}_i[\mathbb{E}[X_i]]] = 0.5 = \mathbb{E}[X_i]$$

Although the t-statistic is marginally closer to 0, this test again rejects the null of rationality with a p-value smaller than 0.001. In the case of subjective beliefs about the population level probability of success for transfer students, it appears we have grounds to reject rational expectations.

I next tackle the slightly more difficult case of tuition expectations. A simple test would again focus on means, either using the sticker price of $10,400 (or $12,000 with fees) mentioned above, which would give an upper bound, or using an estimate of out-of-pocket costs (which would be strictly lower). Even with the strict upper bound of $12,000, both own- and population-expected tuition costs would fail a test of rational expectations with p-value smaller than 0.001. While testing with the distribution of own-beliefs is subject to the same caveats about selection described above, the test using population-beliefs should not be, and the results give us reason to think that beliefs about the average tuition cost are biased upwards.

Economists in more recent years have begun developing more sophisticated tests for rational expectations, in particular ones better able to control for covariates and which incorporate restrictions on higher order moments of a distribution. D'Haultfoeuille et al. (2021) introduce one such test, which relies on the insight that realized outcomes must be a mean-preserving spread of subjective expectations, thus implying a restriction on the variance of the distribution as well as the mean.

In order to leverage the power of these tests to control for covariates, and thus to potentially provide a better test for whether the distribution of own-beliefs in my data is consistent with rational expectations, I compare the subjective expectations of my sample against data from the High School Longitudinal Study (2009). This is a survey which tracked a large sample of students who began 9th grade in 2009, and thus were entering college in 2013 (if going directly from undergrad to higher education). In order to get the samples which match up as closely as possible, I restrict my survey to respondents between 18 and 22 who gave expected tuition at in-state public schools, and in the HSLS I look at the net tuition paid by those enrolled in in-state public schools (details on how I processed the HSLS can be found in appendix C).

Table 8 shows the p-values from that test, both without covariates and while controlling for high school GPA (which should influence merit-based aid) and family income (which should influence need-based aid). As an extra robustness check, I also test for rationality in the case that my survey respondents were mistakenly answering for the sticker price rather than the net (out-of-pocket) price.

Covariates	Sticker	Net
No	0.012	0.00
Yes	X	0.00

Table 8: Test for RE of tuition costs

Results from this test again give us reason to doubt that beliefs are rational. When comparing net price, as the survey question designed to elicit, I reject the null of rational expectations with $p \leq 0.01$.

The robustness check also suggests that even if individuals were misinterpreting the question (and thus overestimating tuition costs), rational expectations would not hold.[16]

Although this is preliminary evidence only, on the two variables where there are clear data-based benchmarks against which to compare subjective beliefs I am able to reject a null of rational expectations.

4.4 Choice probabilities

The final component of the survey data is the choice probabilities. Respondents provide the probability that they earn no degree (dropout), a 2-year degree (AA), and transfer to 4-year school, either with or without an AA. Figure 11 shows both the average for each probability (on the right) as well as the percentage of the sample which rated each path as their most likely outcome (on the left). We can see that students overwhelmingly intend to transfer, although there is a sizable minority which intends to pursue a terminal AA. Within those students who intend to transfer, the vast majority intend to do so after earning a 2-year degree. As might be expected for a group investing in their education, respondents assign a very low probability to the case that they leave the community college with no degree and without transferring.

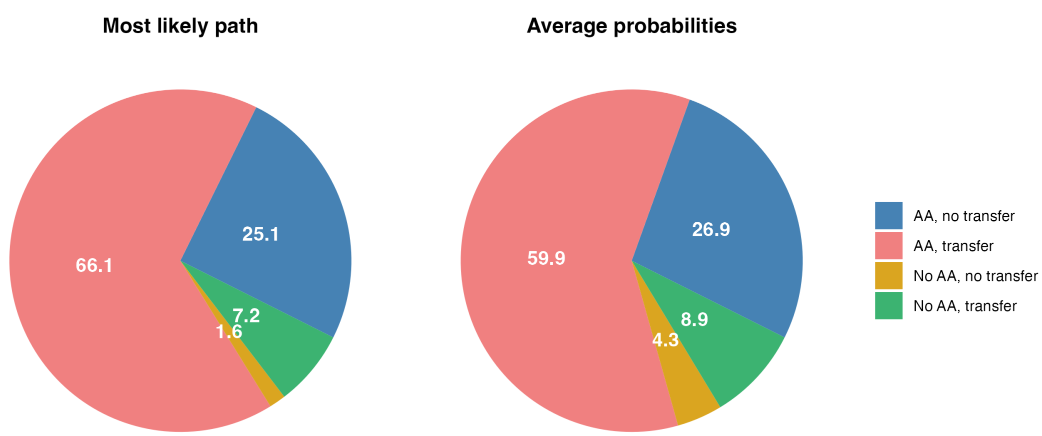


Figure 11: Education path probabilities

We will now put the pieces of the descriptive analysis together by linking subjective beliefs to the reported outcome probabilities in a simple, multinomial logit.

[16] I do not run the test with covariates using sticker price, since if respondents were indeed answering with their expected sticker price, then personal characteristics (which influence aid only) should not matter.

5 A simple model of community college students' choices

In this section I will present results from a multinomial logit regression of the probability of each education pathway on fixed characteristics and expected outcomes. First, I will motivate this framework using a simple, static choice model, before presenting estimates and discussing a few implications of my results.

5.1 Static choice model

Community college students make a joint decision about whether to complete their two-year degree and whether to transfer to a 4-year institution. Index these choices as $d_i \in \{1, 2, 3, 4\}$, where:

1. No 2-year degree, no transfer (dropout)
2. 2-year degree, no transfer (terminal AA)
3. No 2-year degree, transfer
4. 2-year degree, transfer

Utility depends on d_i and on the highest degree ultimately achieved, which may be high school, 2-year, or 4-year; denote this outcome by $D_i(d_i) \in \{HS, AA, BA\}$. Receipt of a BA is uncertain: if an individual chooses to transfer, they only receive a BA with probability p_i. However, they will incur a tuition cost of T_i regardless of the outcome, making the decision to transfer risky. Utility can be written as:

$$u(d_i = 1) = v(HS) + \nu_{i1} \tag{2a}$$

$$u(d_i = 2) = v(AA) + \nu_{i2} \tag{2b}$$

$$u(d_i = 3) = (1 - \mathbb{1}[D_i(3) = BA]) \times v(HS) + \mathbb{1}[D_i(3) = BA] \times v(BA) - \gamma \log(T_i) + \nu_{i3} \tag{2c}$$

$$u(d_i = 4) = (1 - \mathbb{1}[D_i(4) = BA]) \times v(AA) + \mathbb{1}[D_i(4) = BA] \times v(BA) - \gamma \log(T_i) + \nu_{i4} \tag{2d}$$

Where $v(\cdot)$ is the portion of utility which comes from $D_i(d_i)$, and ν_{ij} is an idiosyncratic choice-specific component unknown to the econometrician but seen by the student prior to making their choice. γ is the utility effect of tuition costs.

Besides uncertainty in degree completion, neither tuition costs conditional on transfer T_i nor the labor market returns to each degree known prior to making the decision. Denote earnings with highest degree D_i as $Y_i(D_i)$. Then the three key sources of uncertainty are:

1. Probability of success at a 4-year institution, p_i
2. Tuition costs at 4-year school, T_i, which follow distribution $F_{i,T}$

3. Labor market returns to a 4-year degree, $Y_i(D_i)$, which follow distribution $F_{i,y}$

Choices are made before these variables are realized, so students make their choice by taking an expectation of (2a) - (2d) over these variables. As in section 4, denote the expectations operator as $\mathscr{E}_i$, where the subscript indicates that the expectation is taken over an individual-specific distribution and we use $\mathscr{E}$ rather than $\mathbb{E}$ to emphasize that expectations need not be rational:

$$\begin{aligned} \mathscr{E}_i\left[Y_i(D_i)\right] &= \int Y_i(D_i) dF_{i,y} \\ \mathscr{E}_i\left[\log(T_i)\right] &= \int \log(T_i) dF_{i,T} \end{aligned} \tag{3}$$

Putting these pieces together, and assuming that p_i is the same for someone with and without a 2-year degree, students consider their subjective expected utility from each choice:

$$\mathscr{E}_i\left[u(d_i = 1)\right] = \mathscr{E}_i\left[v(HS)\right] + \nu_{i1} \tag{4a}$$

$$\mathscr{E}_i\left[u(d_i = 2)\right] = \mathscr{E}_i\left[v(AA)\right] + \nu_{i2} \tag{4b}$$

$$\mathscr{E}_i\left[u(d_i = 3)\right] = (1 - p_i) \times \mathscr{E}_i\left[v(HS)\right] + p_i \times \mathscr{E}_i\left[v(BA)\right] - \gamma \mathscr{E}_i\left[\log(T_i)\right] + \nu_{i3} \tag{4c}$$

$$\mathscr{E}_i\left[u(d_i = 4)\right] = (1 - p_i) \times \mathscr{E}_i\left[v(AA)\right] + p_i \times \mathscr{E}_i\left[v(BA)\right] - \gamma \mathscr{E}_i\left[\log(T_i)\right] + \nu_{i4} \tag{4d}$$

As is standard, community college students face two types of uncertainty: *resolvable* uncertainty which will be known to the decision maker at the time of the decision but is not known to the econometrician, and *unresolvable* uncertainty which will not. I specify ν_{ijk} as containing the resolvable component, although this does not preclude ν_{ijk} from also containing some nonstochastic features (as detailed below).

Note that we can rewrite eqs. (4c) and (4d) as:

$$\mathscr{E}_i\left[u(d_i = 3)\right] = \mathscr{E}_i\left[u(HS)\right] + p_i \times \mathscr{E}_i\left[u(BA) - u(HS)\right] - \gamma \mathscr{E}_i\left[\log(T_i)\right] + \nu_{i3} \tag{5a}$$

$$\mathscr{E}_i\left[u(d_i = 4)\right] = \mathscr{E}_i\left[u(AA)\right] + p_i \times \mathscr{E}_i\left[u(BA) - u(AA)\right] - \gamma \mathscr{E}_i\left[\log(T_i)\right] + \nu_{i4} \tag{5b}$$

which more clearly reflects the fact that transfer is a risky investment bringing a possible return of $u(BA) - u(D_i')$ over a baseline level of utility from D_i' for $D_i' \in \{HS, AA\}$, while paying an expected (utility) cost of $\gamma \mathscr{E}_i\left[\log(T_i)\right]$. Thus, whether the student expects the investment to be worthwhile depends on the expected increase in utility in the case of success, the probability of success, the expected cost of tuition, and the idiosyncratic component.

We can go further and reach a tractable specification for the empirical exercise by assuming that pecuniary and non-pecuniary returns are separable in $v(\cdot)$ and specifying a simple linear utility function:

$$v(D_i) = X_i'\beta_{D_i} + e^{-b(s-18)} \times \lambda_{D_i} Y_i(D_i) \tag{6}$$

Where s is current age and $e^{-b(s-18)}$ is an age effect which means that older people may value the pecuniary returns to a degree differently than young people.

Parameters vary freely for each degree-type. $X_i'\beta_{D_i}$ gives non-pecuniary utility which depends on observable characteristics X_i and contains a common intercept. Pecuniary returns are captured by $e^{-b(s-18)} \times \lambda_{D_i} Y_i(D_i)$. $Y_i(D_i)$ is the only stochastic component in $v(\cdot)$. This is a reduced form utility function and is mis-specified in the sense that we are projecting what is properly a dynamic problem onto a static one (more on this later). The model will fit the data well in the case that the static model is able to capture the relevant continuation payoff from the dynamic problem using the β_{D_i}, b, λ_{D_i}, and γ parameters.

Assuming the functional form of eq. (6), we can further rewrite eqs. (5a) and (5b) as:

$$\begin{aligned}
\mathscr{E}_i\left[u(d_i=3)\right] = {} & X_i'\beta_{HS} + e^{-b(s-18)} \times \lambda_{HS}\mathscr{E}_i\left[Y_i(HS)\right] + p_i \times X_i'(\beta_{BA} - \beta_{HS}) + \\
& p_i \times e^{-b(s-18)} \times \mathscr{E}_i\left[\lambda_{BA}Y_i(BA) - \lambda_{HS}Y_i(HS)\right] - \gamma\mathscr{E}_i\left[\log(T_i)\right] + \nu_{i3} \qquad (7a) \\
\mathscr{E}_i\left[u(d_i=4)\right] = {} & X_i'\beta_{AA} + \lambda_{AA}\mathscr{E}_i\left[Y_i(AA)\right] + \underbrace{p_i \times X_i'(\beta_{BA} - \beta_{AA})}_{\equiv \varrho_i^{AA}\text{: expected non-pecuniary returns}} + \\
& \underbrace{e^{-b(s-18)} \times p_i \times \mathscr{E}_i\left[\lambda_{BA}Y_i(BA) - \lambda_{AA}Y_i(AA)\right]}_{\equiv \rho_i^{AA}\text{: expected pecuniary returns}} - \gamma\mathscr{E}_i\left[\log(T_i)\right] + \nu_{i4} \qquad (7b)
\end{aligned}$$

Here, ϱ_i^{AA} is the *non*-pecuniary premium for a BA over an AA, and ρ_i^{AA} the expected *pecuniary* premium for a BA over AA (with a similar interpretation for A compared to HS). It will be of interest to explore the relative importance of ϱ_i^{AA} and ρ_i^{AA} in explaining the reported choice probabilities.

With expected earnings at three points in the future, I construct $Y_i(D_i)$ as an index based on the expected log of future earnings:

$$Y_i(D_i) \equiv \log\left(Y_i(D_i|s+5) + \beta^5 Y_i(D_i|s+10) + \beta^{15} Y_i(D_i|s+20)\right)$$

Where β discounts future earnings and is set to 0.95. For fixed characteristics, I include the log of high school GPA (which will capture complementarity between skills and degree), an indicator for whether one's parent had a BA or higher (to capture SES), and an indicator for having children.

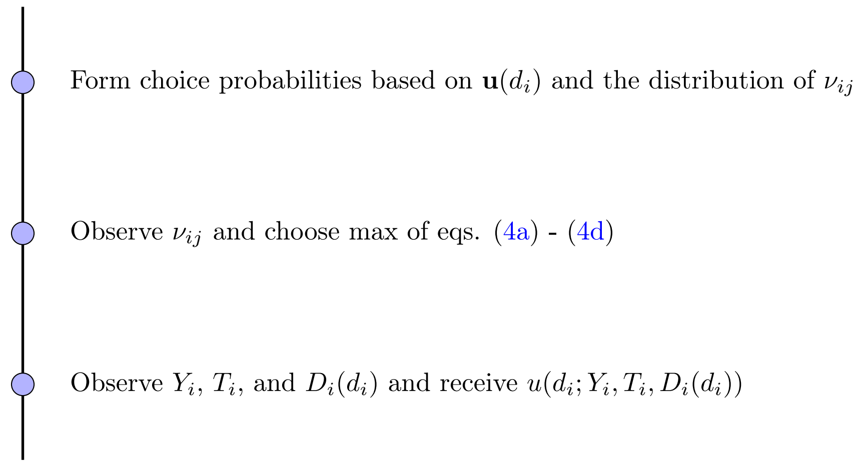


Figure 12: Information timeline of respondents

Timeline

It is worth emphasizing the difference between my setup and a standard discrete choice problem. Economists typically work with observed choices, rationalizing them by modeling the decision process of individuals who know some things (the ν_{ij} components), remain uncertain about others (tuition and earnings in $\mathbf{u}(d_i)$), and combine these to calculate their expected utility. In this case, rather than working with observed choices, I am using reported choice probabilities, and the information set of my respondents is different; respondents know neither ν_{ij} nor the uncertain components of $\mathbf{u}(d_i)$, and so they integrate over both when forming their expected utility. Figure 12 illustrates the process.

Empirical exercise

So far I have said little about ν_{ij}, which capture the choice-specific impacts of unobservables. For the empirical exercise, I focus on the case where ν_{ij} is a choice-specific "taste shock" which is uncorrelated with observables. For that purpose, it will be useful to define the portion of the utility net of ν_{ij} as:

$$\mathbf{u}(d_i) = u(d_i) - \nu_{ij} \tag{8}$$

Defining F_ν as the joint distribution of the 4 shocks, the probability that $d_i = j'$ is given by:

$$\begin{aligned} \pi_i(d_i = j') &= P(\mathbf{u}(j') + \nu_{ij'} > \mathbf{u}(j) + \nu_{ij},\ \forall j \neq j') \\ &= P(\nu_{ij'} - \nu_{ij} > \mathbf{u}(j) - \mathbf{u}(j'),\ \forall j \neq j') \\ &= \int_{\nu_i 1, \ldots, \nu_i 4} \mathbb{1}\left[\nu_{ij'} - \nu_{ij} > \mathbf{u}(j) - \mathbf{u}(j'),\ \forall j \neq j'\right] dF_\nu \end{aligned} \tag{9}$$

So the probability of each choice depends on the joint probability of three events integrated over four random draws of ν_{ij}. The joint distribution of the ν_{ij} thus defines the probabilities. If ν_{ij} are iid and follow a type I extreme value distribution with scale parameter σ, then choice probabilities take the familiar logit form, and are given by:

$$\hat{\pi}_i(d_i = j') = \frac{\exp(\mathbf{u}(d_i = j)/\sigma)}{\sum_{d_i} \exp(\mathbf{u}(d_i)/\sigma)} \tag{10}$$

This is a common specification in discrete choice models, used in a similar context in Arcidiacono et al. (2012).[17] The scale parameter σ is a nuisance parameter which is generally not identified.

Using the subjective expectations data, I am able to plug in individual beliefs directly into the RHS of eqns. (4a) - (4d), put these values into eq. (10), and then find the preference parameters such that the model-implied choice probabilities $\hat{\pi}_i(d_i = j)$ best match the reported choice probabilities $\pi_i(d_i = j)$. This allows me to identify preferences without any assumptions on the form that beliefs take (albeit with a strong parametric assumption on the heterogeneous component). The model-implied choice probabilities are:

$$\hat{\pi}_i(d_i = j) = \frac{\exp(\mathbf{u}(d_i = j))}{\sum_{d_i} \exp(\mathbf{u}(d_i))} \tag{11}$$

Following Papke and Wooldridge (1996), the log-likelihood contribution from each individual is:

$$l_i(\theta) = \sum_{d_i} \left[\pi_i(d_i) \log(\hat{\pi}_i(d_i)) + (1 - \pi_i(d_i)) \log(1 - \hat{\pi}_i(d_i))\right] \tag{12}$$

We will find the parameter vector $\theta^* = [\beta_{HS}, \beta_{AA}, \beta_{BA}, \gamma]$ which maximizes the log-likelihood function for our data:

$$\theta^* = \arg\max_\theta \sum_i l_i(\theta) \tag{13}$$

It is worth restating the strong set of assumptions underpinning this exercise:

1. Parametric restriction on unobservables ν_{ij}
2. Probability of success at a 4-year school is independent of 2-year degree status
3. Utility function is linear in observable characteristics

Of course, other restrictions are lurking under the hood, in particular around the dynamic nature of the problem, financial constraints, and the possibility of dropout from community college. I will discuss these in section 6 when I present the full model.

[17] Wiswall and Zafar (2015), on the other hand, are able to relax this assumption by using an information treatment to difference out the unobserved component.

5.2 Results

Parameter estimates are shown in table 9. While the estimates on the fixed characteristics are relatively noisy, most of the parameters are precisely estimated in the expected direction; respondents value earnings positively, receive negative utility from tuition, and at baseline get more non-pecuniary utility from a 4-year degree than a 2-year degree. The only surprising result is an aggressive discount parameter in the age function; this may be due to the functional form, and in future work I will experiment with alternative specifications.

Table 9: Parameter estimates

Variable	Estimate
2-year degree:	
Intercept	1.354* (0.912)
Parent has BA	0.941** (0.440)
Log HS GPA	0.475 (0.761)
Has children	0.857 (1.010)
4-year degree:	
Intercept	2.761*** (1.163)
Parent has BA	1.314*** (0.513)
Log HS GPA	0.139 (0.952)
Has children	0.867 (1.133)
Age	0.196* (0.152)
Earnings	0.631*** (0.221)
Log(tuition)	-0.252*** (0.079)

Notes: Standard errors in parentheses.
Reference category is high school only.
Significance: * p<0.10, ** p<0.05, *** p<0.01

Table 10 shows the average marginal effect from a few key variables. Effects are given as the average percentage-point increase in the probability of earning a high school degree only, an AA only, or choosing to transfer. Rows 1 and 2 show the effect of increasing the expected premium, relative to high school only, by 1%; so the effect of -0.223 for an increase in the BA premium means that, on average, respondents decrease the probability of earning a HS degree only by 0.223%. Row 3 shows the effect of decreasing tuition by 1%. Surprisingly, this has a roughly similar effect to increasing the BA premium by 1%. Finally, row 4 shows the effect of increasing the probability of successfully graduating by 1 percentage point; this results in almost a 0.5% increase in the probability of transfer.

Finally, table 11 shows a partial decomposition of $\mathscr{E}_i\left[u(d_i = 3)\right]$ and $\mathscr{E}_i\left[u(d_i = 4)\right]$. Row 1 shows the portion of subjective expected utility related to expected pecuniary returns, ρ_i^{HS} and ρ_i^{AA}, while row 2 shows the portion of subjective expected utility related to expected *non*-pecuniary returns, ϱ_i^{HS} and ϱ_i^{AA}. The remainder of utility is driven by the default option in the case that the respondent is not successful

Table 10: Average marginal effects

Variable	HS Only	AA Degree	Transfer
AA premium	-0.001*** (0.000)	0.000*** (0.000)	-0.001*** (0.000)
BA premium	-0.223*** (0.074)	-0.029*** (0.009)	0.083*** (0.028)
4-year tuition	-0.173*** (0.055)	-0.023*** (0.007)	0.078*** (0.025)
Graduation prob.	-1.006*** (0.143)	-0.470*** (0.039)	0.494*** (0.079)

Notes: Bootstrapped standard errors in parentheses.
Effects are average percentage changes in choice probabilities, given in percentage points.
Significance: * p<0.10, ** p<0.05, *** p<0.01

in pursuing a BA, either $v(HS)$ or $v(AA)$. Of interest is that in both cases, non-pecuniary returns are much more important than pecuniary returns, of similar orders of magnitude (although pecuniary returns are slightly more important when transferring without an AA).

Table 11: Importance of pecuniary and non-pecuniary expected returns

Return Type	BA to HS	BA to AA
Pecuniary	0.053*** (0.011)	0.028*** (0.007)
Non-pecuniary	0.583 (0.475)	0.173*** (0.066)

Notes: Bootstrapped standard errors in parentheses.
Values represent proportion of total expected utility from $d_i = 3,\ 4$.
Significance: * p<0.10, ** p<0.05, *** p<0.01

5.3 Limitations

Although the results are suggestive, the static model suffers from some significant shortcomings. First, this current model is identified off of a strong parameteric assumption on unobservables; relaxing this assumption by modeling that heterogeneity (in a control function type approach) or using panel data would improve the credibility of the results. Second, the setup uses the elicited probabilities of each education path as *choice* probabilities; but these probabilities may take into account shocks which could force non-voluntary dropout at the community college level, or financial constraints which would prevent transfer; that is to say, they may not fully reflect *choices* individuals make, but also include forcing events. This would lead to biased estimates of preferences, since a low subjective probability which actually reflects an expected financial constraint would be rationalized in the model by some combination of preferences. Finally (and relatedly) this model does not capture the inherent dynamics in the problem. Students invest in their education in order to get a long-term return, and modeling that future utility in a richer way than the static-projection in the above model is necessary to understand the tradeoffs they make. In the next section, I will present a full dynamic model which will address these shortcomings.

6 Dynamic model

In this section, I present a dynamic model of community college student choice which addresses the shortcomings of the static model, and can be estimated from the survey data. As before, the primary motivation is to understand the role of uncertainty and subjective (potentially biased) beliefs of current community college students with respect to:

1. Tuition costs at 4-year school
2. Graduation probabilities from 2- and 4-year school
3. Wage premia from 2- and 4-year degrees

6.1 Setting

The model is in discrete time with a finite horizon. Periods correspond to 6-month semesters[18] and are discounted by a factor β. The model ends at period T with a terminal value payoff, so periods are:

$$t \in [t_1, t_1 + 1, \dots, T]$$

Agents are current community college students who are defined by a state Ω_t and make decisions $\omega_t = \{c_t, d_{t+1}, h_{t+1}\}$, where d_{t+1} is next-period enrollment, h_{t+1} next period labor supply, and c_t is consumption. Possible values of d_t depend on current enrollment status S_t^j and on accumulated credits K_t:

1. If $S_t^c = PT$ and $K_t < 3.5$ or $S_t^c = FT$ and $K_t < 3$ (enrolled CC student, not about to graduate) then $d_{t+1} \in \{0, S_{t+1}^c, S_{t+1}^u\}$, that is, they can choose no enrollment (dropout), enrollment in CC, at PT or FT status (persistence), or enrollment in a 4-year program at PT or FT status (transfer without AA)

2. If $S_t^c = PT$ and $K_t = 3.5$ or $S_t^c = FT$ and $K_t = 3$ (enrolled CC student, about to graduate) then $d_{t+1} \in \{0, S_{t+1}^u\}$, that is, they can choose no enrollment (terminal AA), or enrollment in a 4-year program at PT or FT status (transfer with AA)

3. If $S_t^u = PT$ and $K_t < 7.5$ or $S_t^u = FT$ and $K_t < 7$ (enrolled 4-year student, not about to graduate) then $d_{t+1} \in \{0, S_{t+1}^u\}$, that is, they can choose no enrollment (dropout), or continued enrollment in a 4-year program at PT or FT status (persistence)

No one else faces an education decision. The decision to transfer and to leave school entirely are absorbing, so any individual not enrolled in period t is constrained to $d_{t+1} = 0$ and only chooses labor supply and consumption.

[18] I abstract from intersession periods during summer and winter breaks. An alternative, as in Keane and Wolpin (2001) or Arcidiacono et al. (2023), would be to include a shorter summer session.

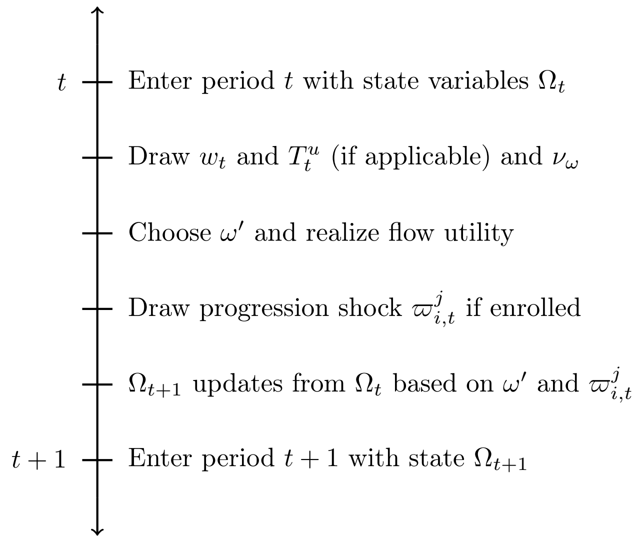
In each period agents face uncertainty around the next period's wage, w_{t+1}. Community college students contemplating transfer are uncertain about tuition costs at 4-year school, T_t^u; for the student who has already transfered, however, T_t^u stays constant during their enrollment. Students are also uncertain about whether their current semester will be successful; with probability π_t^j they draw a negative success shock $\varpi_t^j = 1$ and are not able to continue their education, where $j = c$ indicates 2-year school (**c**ommunity college) and $j = u$ indicates 4-year school (**u**niversity). Finally, they draw an idiosyncratic shock associated with each discrete choice, ν_{ω_t}.

6.2 Timeline

The timeline of the model is illustrated in figure 6.2. First, workers realize their wage w_t, first semester transfer students draw their tuition cost T_{t+1}^u, and all agents draw taste shocks ν_ω for each discrete choice they may make in t. The agent then makes their choice: c_t and h_{t+1} for all agents, and d_t for students. There is an intensive margin for labor supply and enrollment, so these variables can take on values of 0, part-time (0.5), and full-time (1).

The agent then realizes their flow utility from c_t, enrollment d_t, and labor supply h_t (described in subsection 6.9). Finally, agents enrolled at school j draw the success shock ϖ_t^j which indicates whether enrollment in period t was successful. Their state variables then update in the following manner:

- **Age**: age $s_{t+1} = s_t + .5$
- **Credits**: college credit K_{t+1} updates according to equation 14
- **Borrowing**: student loan balance B_{t+1} updates according to equation 17
- **Degree**: highest degree D_{t+1} updates to AA or BA if the agent has crossed the credit threshold as described in subsection 6.4
- **Enrollment**: enrollment status S_{t+1}^c and S_{t+1}^u update according to the choice d_{t+1} made in period t, conditional on the success shock ϖ_t^j



Model Timing

6.3 Initial conditions

The initial state (Ω_{t_1}) consists of:

1. Wage $w_{t_1}^*$, which is missing if $h_{t_1} = 0$
2. Community college tuition $T_{t_1}^c$
3. Age s_{t_1}
4. Highest degree D_{t_1}, which is high school (HS) for the whole sample
5. Accumulated college credits K_{t_1}
6. Assets A_t which are net of student loan balance B_{t_1}
7. Parental transfer function $Tr(\Omega_t)$ which determines how parental transfers evolve as the agent progresses through school

All of these variables will be directly observed in the survey, along with the initial period labor supply choice h_{t_1} and community college enrollment $S_{t_1}^c$, both of which reflect decisions made in t_0.

6.4 Degree progression

Progression in school is uncertain in two ways. First, agents may voluntarily choose to leave; in the model, this could be due to a particularly high wage draw or to the relative values of the taste shocks. Second, agents face the possibility of *failure* in each semester, where they will not be allowed to continue their education.

With probability $\pi^j(\Omega_t)$ the agent draws a success shock $\varpi_t^j = 1$ and education in period t at school of type j is not successful, where dependence on Ω_t allows success probabilities to change over time. Specifically, the closer a student is to earning their degree, the less uncertainty there is that they successfully complete a semester (as in Athreya and Eberly (2021)), so we would expect that $\pi^j(\Omega_t) \leq \pi^j(\Omega_{t+1})$, and in the final semester before receiving a degree there is no uncertainty remaining. There may also be a dependency on age, PT vs. FT status, and other factors. If $\varpi_t^j = 0$, then the agent is successful and they accumulate additional credit depending on their enrollment status:

$$K_{t+1} = \begin{cases} K_t + 0.5 & \text{if } S_t^j = 0.5 \\ K_t + 1 & \text{if } S_t^j = 1 \end{cases} \tag{14}$$

Degree receipt itself is deterministic. Highest degree D_{t+1} updates to AA (a 2-year degree) when $S_t^c \neq 0$ and $K_{t+1} \geq 2$, representing the student who has accumulated 2 years of full-time equivalent credits and received an AA. Highest degree D_{t+1} updates to BA (a 4-year degree) when $S_t^u \neq 0$ and $K_{t+1} \geq 4$, representing the 4-year college student who has accumulated 4 years of full-time equivalent.

6.5 Additional uncertainty

Net tuition[19] at 4-year school is stochastic with a CDF $F_{i,T^u}(\cdot)$. The i subscript makes clear that each agent draws from an individual-specific distribution which may depend on the observed state $\Omega_{i,t}$ as well as on other unobserved characteristics. Agents draw from this distribution in period t prior to choosing d_t. 4-year costs are then constant at $T_{i,t+1}^u$ for the rest of enrollment.

Wages are drawn from a distribution which depends on age and highest degree, with CDF $F_{i,w}(\cdot; s_t, D_t)$. The agent draws $w_{i,t}$ prior to making their labor supply decision. The distribution explicitly varies with observed characteristics (age s_t and highest degree D_t), and the i subscript again makes clear that each individual draws from a specific distribution which may depend on observable characteristics $\Omega_{i,t}$ as well as other individual-specific characteristics not be observed by the econometrician.

Finally, agents face *resolvable uncertainty* not observed by the econometrician. This is unknown at the time that individuals report their choice probabilities, but will be observed before their decision is made and is represented by an additive "taste shock" to the utility of each discrete choice in each period. Taste shocks are assumed to be i.i.d. and follow a Gumbel distribution with location 0 and scale parameter σ, so that the choice probabilities will follow the classic logit form.

[19] Net (or out-of-pocket) tuition is given as the difference between the sticker price and total grants and aid awarded.

I assume that the ν shocks are type-1 extreme value. In addition, the school success shock follows a Bernoulli distribution, since the agent either progresses or does not. I specify a parametric distribution for $F_{i,w}(w; s_t, D_t)$ or $F_{i,T^u}(T^u)$ based on elicited PDFs in the manner described in appendix B.

6.6 Parental transfers

Enrolled students may receive some income in the form of parental transfers, Tr_t. Each agent begins the model with a parental transfer function $Tr(\cdot)$, which shows how parental transfers change based on the student's state Ω_t. I observe the level of parental transfers in t_1 and assume that they evolve according to an exogeneous process which will depend on student's age and background characteristics (to be estimated from external data) so that parental support may depend on a student's age and family characteristics. Once an agent leaves school, parental transfers go to 0.

$$PT_{i,t} = Tr_i(\Omega_{i,t}) \tag{15}$$

6.7 Borrowing

Agents are able to take out student loans to finance their education and meet a consumption floor for students, $\underline{c}^s$. This process is automatic, in the sense that for the enrolled student, any tuition costs which are not met by income and parental transfers are covered by student loan borrowing, and agents take this into account when making their enrollment decisions.[20] The interest rate process follows discussion from Lochner and Monge-Naranjo (2008). Student loans do not accrue interest while agents are enrolled, $S_t^c + S_t^u \neq 0$. When agents are not enrolled ($S_t^c + S_t^u = 0$), student loans grow with an interest rate of $\mathbf{R}$ but decrease by student loan payments which are given by the function $b(B_t, \Omega_t)$. The interest rate and loan repayment amount will be calculated outside of the model, either according to program rules or estimated in a reduced form way from external data (such as the ELS).[21]

Labor earnings are given by:

$$Y_t = w_t \times 1020 \times h_t \tag{16}$$

With $\dfrac{52 \times 40}{2} = 1,020$ representing full-time hours for 6 months, which are multiplied by $\frac{1}{2}$ if $h_t = 0.5$

[20] This assumes agents always have access to credit; this will be relaxed in the future.
[21] Johnson (2013) specifies a linear function for debt payments.

and are 0 if $h_t = 0$. The student loan balance B_{t+1} updates according to:

$$B_{t+1} = \begin{cases} B_t + \max(0, (T_t^j + \underline{c}^s - w_t \times 1020 \times h_t - PT_t(\Omega_t))) & \text{if } S_t^c + S_t^u \neq 0 \\ \mathbf{R} \times (B_t - b(B_t, \Omega_t)) & \text{if } S_t^c + S_t^u = 0 \end{cases} \tag{17}$$

6.8 Consumption

The budget constraint for students is determined by their income, parental transfers, borrowing, and accumulated assets. Define by A_t assets net of borrowing and let M_t^s be the pre-borrowing market resources from which a student can consume:

$$M_t^s = A_t + w_t \times 1020 \times h_t + Tr(\Omega_t) - T_t^c \mathbb{1}(S_t^c \neq 0) - T_t^u \mathbb{1}(S_t^u \neq 0) \tag{18}$$

Students choose $c_t \leq M_t$ if $M_t \geq \underline{c}^s$, or consume $\underline{c}^s$ and accumulate additional student loans following eq. (17). Assets then grow according to:

$$A_{t+1} = \begin{cases} R \times (M_t - c_t) & \text{if } M_t - c_t \geq 0 \\ M_t - c_t & \text{if} M_t - c_t < 0 \text{ and } S_t^c + S_t^u \neq 0 \\ \mathbf{R} \times (M_t - c_t) & \text{if} M_t - c_t < 0 \text{ and } S_t^c + S_t^u = 0 \end{cases} \tag{19}$$

Where it may be that $R \neq \mathbf{R}$.

The budget constraint for the non-student is based on income, accumulated assets, and loan payments (where former students may grow savings while paying off loans). Market resources are:

$$M_t^{ns} = A_t + w_t \times 1020 \times h_t - b(B_t, \Omega_t) \tag{20}$$

Non-students can choose $c_t \leq M_t^{ns}$ if $M_t^{ns} \geq \underline{c}^{ns}$; otherwise, they consume M_t^{ns} and receive an automatic benefit bringing their consumption up to $\underline{c}^{ns}$. Assets then grow according to $A_{t+1} = R \times (M_t - c_t) + B_{t+1}$.

6.9 Flow utility

Utility comes from consumption, school enrollment, and labor supply, and for the latter two choices varies based on the intensive margin (part-time vs. full-time). Utility is separable in each component, and utility from consumption follows a CRRA form with curvature γ. The general form is:

$$
\begin{aligned}
u_t(\omega') = \eta \frac{c_t^{1-\gamma} - 1}{1-\gamma} + \sum_{j=0}^{2} \phi_j^c \mathbb{1}(S_t^c = j) \\
+ \sum_{k=0}^{2} \phi_k^u \mathbb{1}(S_t^u = k) + \sum_{l=0}^{2} \phi_l^h \mathbb{1}(h_t = l) + \nu_{\omega'}
\end{aligned}
\tag{21}
$$

Note that utility in period t depends directly on today's c_t decision, but enrollment choice d_t and labor supply h_t affect period t utility only through the taste shock, and affect utility in period $t+1$ through their effects on state Ω_{t+1} which lead to direct utility effects from school enrollment through the ϕ^c and ϕ^u parameters and indirect effects through their impacts on consumption c_{t+1}. This is a standard distinction between the pecuniary and non-pecuniary rewards to schooling, as in Todd and Zhang (2020).

6.10 Agent's problem

I have now defined a state space consisting of age, highest degree, accumulated credits, assets, parental transfers, and school enrollment status; a choice set with decisions about current period labor supply and next period enrollment; flow utility; and future uncertainty around wages, tuition, and school progression. With all of these elements, we are ready to specify the agent's problem. Denote by $F_i(\Omega_s|\Omega_t)$ the CDF of future states Ω_s, which may take on values in $\mathbf{\Omega}$, conditional on current state Ω_t. This depends on the joint distribution of all shocks described in subsection 6.5.

In period t agents choose ω_t to maximize their expected lifetime utility. We can write the agent's problem in Bellman equation form as:

$$
\begin{aligned}
V_t(\Omega_t) &= \max_{\omega_t} \left(u(\omega_t) + \beta \mathbb{E}_i \left[V_t(\Omega_{t+1}) \right] \right) \\
&= \max_{\omega_t} \left(u(\omega_t) + \beta \int_{\mathbf{\Omega}} V_{t+1}(\Omega_{t+1}) \ dF_i(\Omega_{t+1}|\Omega_t) \right)
\end{aligned}
\tag{22}
$$

6.11 Extensions

The model presented in this section addresses a number of the shortcomings of the simple model from section 5; in particular, it allows for the outcome probabilities to be the cumulative result of both choices and shocks and captures the long-run nature of investments in human capital. There are several further extensions, some of which are possible now and others of which will be implemented with the second round of data. These include a more realistic and explicit approach to unobserved heterogeneity and a model of belief updating.

7 Conclusion

In this paper I have presented novel survey data on the beliefs of community college students; to my knowledge, this is the first survey of its kind to offer evidence on the beliefs of students about outcomes from transfer. Students overwhelmingly intend to transfer to 4-year programs, and expect significant wage premia to earning 4-year degrees. However, there is suggestive evidence that beliefs about tuition costs and graduation probabilities deviate from rational expectations. Using a multinomial logit, I show how the reported probabilities of pursuing 2- and 4-year degrees are influenced by the expected outcomes from each. Tuition costs emerge as particularly important – a somewhat promising sign, since respondents seem to overestimate tuition costs. Non-pecuniary returns, rather than wage premia, seem to drive decisions.

However, there are several next steps which I am actively working on. In section 6 I present a dynamic model which addresses some of the many shortcomings of the static model. In a future version of this paper I will provide details on the estimation and results, and then conduct counterfactual analyses to measure the welfare costs of biases and evaluate the efficacy of various financial aid policies in the presence of biased beliefs.

More ambitious work will be possible after collecting a second round of survey data; this will happen in fall 2025, supported by a grant from the Russell Sage Foundation, and will allow me to measure the accuracy of first-round beliefs and to study how students learn and update their expectations. This research contributes to our understanding of how subjective beliefs shape educational decisions and highlights the importance of addressing information frictions in higher education policy, but more work is needed to understand the data, and corresponding behavior, as thoroughly as possible.

References

Abbiati, G. and Barone, C. (2017). Is university education worth the investment? The expectations of upper secondary school seniors and the role of family background. *Rationality And Society*, 29(2):113–159.

Arcidiacono, P., Aucejo, E. M., Maurel, A., and Ransom, T. (2023). College Attrition and the Dynamics of Information Revelation. *https://papers.ssrn.com › sol3 › papershttps://papers.ssrn.com › sol3 › papers*.

Arcidiacono, P., Aucejo, E. M., and Spenner, K. (2012). What happens after enrollment? An analysis of the time path of racial differences in GPA and major choice. *IZA Journal of Labor Economics*, 1(1):1–24.

Arcidiacono, P., Hotz, V. J., and Kang, S. (2010). Modeling College Major Choices using Elicited Measures of Expectations and Counterfactuals. (15729).

Armantier, O., Topa, G., van der Klaauw, W., and Zafar, B. (2016). An overview of the survey of consumer expectations. *ERN: Collecting*.

Ashworth, J. and Ransom, T. (2019). Has the college wage premium continued to rise? Evidence from multiple U.S. surveys. *Economics of education review*, 69:149–154.

Athreya, K. and Eberly, J. (2021). Risk, the College Premium, and Aggregate Human Capital Investment. *American Economic Journal: Macroeconomics*, 13(2):168–213.

Attanasio, O. and Kaufmann, K. (2009). Educational Choices, Subjective Expectations, and Credit Constraints. (15087).

Bahr, P. R., Dynarski, S., Jacob, B., Kreisman, D., Sosa, A., and Wiederspan, M. (2015). Labor Market Returns to Community College Awards: Evidence from Michigan. A CAPSEE Working Paper. *Center for Analysis of Postsecondary Education and Employment*.

Baker, R., Bettinger, E., Jacob, B., and Marinescu, I. (2018). The Effect of Labor Market Information on Community College Students' Major Choice. *Economics of education review*, 65:18–30.

Barone, C., Schizzerotto, A., Abbiati, G., and Argentin, G. (2017). Information Barriers, Social Inequality, and Plans for Higher Education: Evidence from a Field Experiment. *European sociological review*, 33(1):84–96.

Becker, G. S. (1964). Human Capital: A Theoretical and Empirical Analysis, with Special Reference to Education.

Belfield, C. R. and Bailey, T. (2011). The Benefits of Attending Community College: A Review of the Evidence. *Community College Review*, 39(1):46–68.

Ben-Porath, Y. (1967). The Production of Human Capital and the Life Cycle of Earnings. *The journal of political economy*, 75(4):352–365.

Berkes, J., Peter, F., Spiess, C. K., and Weinhardt, F. (2022). Information provision and postgraduate studies. *Economica*, 89(355):627–646.

Bettinger, E. P., Long, B. T., Oreopoulos, P., and Sanbonmatsu, L. (2012). The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment*. *The quarterly journal of economics*, 127(3):1205–1242.

Betts, J. R. (1996). What Do Students Know about Wages? Evidence from a Survey of Undergraduates. *The Journal of human resources*, 31(1):27–56.

Black, S. E., Denning, J. T., Dettling, L. J., Goodman, S., and Turner, L. J. (2020). Taking It to the Limit: Effects of Increased Student Loan Availability on Attainment, Earnings, and Financial Well-Being. (27658).

Blass, A. A., Lach, S., and Manski, C. F. (2008). Using Elicited Choice Probabilities to Estimate Random Utility Models: Preferences for Electricity Reliability. (14451).

Bleemer, Z. and Zafar, B. (2018). Intended college attendance: Evidence from an experiment on college returns and costs. *Journal of public economics*, 157:184–211.

Blesch, M., Eisenhauer, P., Haan, P., Ilieva, B., Schrenker, A., and Weizsäcker, G. (2023). Biased wage expectations and female labor supply. Technical report, Discussion Paper.

Boneva, T., Golin, M., and Rauh, C. (2022). Can perceived returns explain enrollment gaps in postgraduate education? *Labour economics*, 77:101998.

Bordalo, P., Gennaioli, N., and Shleifer, A. (2018). Diagnostic Expectations and Credit Cycles. *The Journal of finance*, 73(1):199–227.

Cameron, S. V. and Heckman, J. J. (2001). The Dynamics of Educational Attainment for Black, Hispanic, and White Males. *The journal of political economy*, 109(3):455–499.

Carroll, C. and Wang, T. (2023). Chapter 25 - Epidemiological expectations. In Bachmann, R., Topa, G., and van der Klaauw, W., editors, *Handbook of Economic Expectations*, pages 779–806. Academic Press.

Carroll, C. D. (1996). Buffer-Stock Saving and the Life Cycle/Permanent Income Hypothesis. (5788).

Carroll, C. D., Crawley, E., Slacalek, J., Tokuoka, K., and White, M. N. (2020). Sticky Expectations and Consumption Dynamics. *American Economic Journal: Macroeconomics*, 12(3):40–76.

Cunha, J. M., Miller, T., and Weisburst, E. (2018). Information and College Decisions: Evidence From the Texas GO Center Project. *Educational evaluation and policy analysis*, 40(1):151–170.

Daniel, E., Ezra, G., and Hunt, J. B. (2009). Mental Health and Academic Success in College. *The B.E. journal of economic analysis & policy*, 9(1):1–37.

Delavande, A. (2008). Pill, patch, or shot? Subjective expectations and birth control choice. *International economic review*, 49(3):999–1042.

Delavande, A., Del Bono, E., and Holford, A. (2022). Academic and non-academic investments at university: The role of expectations, preferences and constraints. *Journal of econometrics*, 231(1):74–97.

Delavande, A. and Zafar, B. (2019). University Choice: The Role of Expected Earnings, Nonpecuniary Outcomes, and Financial Constraints. *The journal of political economy*, 127(5):2343–2393.

DeLuca, S., Papageorge, N. W., Boselovic, J. L., Gershenson, S., Gray, A., Nerenberg, K. M., Sausedo, J., and Young, A. (2021). "When Anything Can Happen": Anticipated Adversity and Postsecondary Decision-Making. (29472).

D'Haultfoeuille, X., Gaillac, C., and Maurel, A. (2021). Rationalizing rational expectations: Characterizations and tests. *Quantitative economics*, 12(3):817–842.

Dinkelman, T. and A., C. M. (2014). Investing in Schooling In Chile: The Role of Information about Financial Aid for Higher Education. *The review of economics and statistics*, 96(2):244–257.

Dominitz, J. and Manski, C. (1996). Eliciting Student Expectations of the Returns to Schooling. *The Journal of human resources*, 31(1):1–26.

Dynarski, S. (2000). Hope for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance. *National tax journal*, 53(3.2):629–661.

Dynarski, S., Libassi, C. J., Michelmore, K., and Owen, S. (2021). Closing the Gap: The Effect of Reducing Complexity and Uncertainty in College Pricing on the Choices of Low-Income Students. *The American economic review*, 111(6):1721–1756.

Dynarski, S., Page, L. C., and Scott-Clayton, J. (2022). College Costs, Financial Aid, and Student Decisions. (30275).

Ehlert, M., Finger, C., Rusconi, A., and Solga, H. (2017). Applying to college: Do information deficits lower the likelihood of college-eligible students from less-privileged families to pursue their college intentions?: Evidence from a field experiment. *Social science research*, 67:193–212.

Engelberg, J., Manski, C. F., and Williams, J. (2009). Comparing the Point predictions and subjective probability distributions of professional forecasters. *Journal of business & economic statistics: a publication of the American Statistical Association*, 27(1):30–41.

Erdem, T., Keane, M. P., Öncü, T. S., and Strebel, J. (2005). Learning About Computers: An Analysis of Information Search and Technology Choice. *Quantitative Marketing and Economics*, 3(3):207–247.

Fuster, A. and Zafar, B. (2023). Chapter 4 - Survey experiments on economic expectations. In Bachmann, R., Topa, G., and van der Klaauw, W., editors, *Handbook of Economic Expectations*, pages 107–130. Academic Press.

Giustinelli, P. (2022). Expectations in Education: Framework, elicitation, and evidence.

Goldin, C. and Katz, L. F. (2008). *The Race between Education and Technology*. Harvard University Press.

Gong, Y., Stinebrickner, R., and Stinebrickner, T. R. (2019). Uncertainty about Future Income: Initial Beliefs and Resolution During College. (25485).

Grosz, M. (2020). The Returns to a Large Community College Program: Evidence from Admissions Lotteries. *American Economic Journal: Economic Policy*, 12(1):226–253.

Heckman, J., Lochner, L., and Cossa, R. (2002). Learning-By-Doing Vs. On-the-Job Training: Using Variation Induced by the EITC to Distinguish Between Models of Skill Formation. (9083).

Horn & Skomsvold (2011). Community College Student Outcomes: 1994–2009.

Hoxby, C. and Avery, C. (2013). The Missing "One-Offs": The Hidden Supply of High-Achieving, Low-Income Students. *Brookings papers on economic activity*, 44(1 (Spring)):1–65.

Hoxby, C. M. and Turner, S. (2015). What High-Achieving Low-Income Students Know about College. *The American economic review*, 105(5):514–517.

Hussey, A. J. and Swinton, O. H. (2011). Estimating the Ex Ante Expected Returns to College. *The American economic review*, 101(3):598–602.

Jacobson, L., LaLonde, R., and G. Sullivan, D. (2005). Estimating the returns to community college schooling for displaced workers. *Journal of econometrics*, 125(1):271–304.

Jäger, S., Roth, C., Roussille, N., and Schoefer, B. (2021). Worker Beliefs About Outside Options. *https://www.nber.org › papershttps://www.nber.org › papers*, (29623).

Jensen, R. (2010). The (perceived) returns to education and the demand for schooling*. *The quarterly journal of economics*, 125(2):515–548.

Jepsen, C., Troske, K., and Coomes, P. (2014). The Labor-Market Returns to Community College Degrees, Diplomas, and Certificates. *Journal of labor economics*, 32(1):95–121.

Johnson, M. T. (2013). Borrowing Constraints, College Enrollment, and Delayed Entry. *Journal of labor economics*, 31(4):669–725.

Kantrowitz (2017). Who Pays the Full Sticker Price for a College Education? http://www.studentaidpolicy.com/who-pays-full-sticker-price-for-a-college-education.html. Accessed: 2023-11-27.

Kaufmann, K. (2014). Understanding the income gradient in college attendance in Mexico: The role of heterogeneity in expected returns. *Quantitative economics*, 5(3):583–630.

Keane, M. P. and Wolpin, K. I. (2001). The Effect of Parental Transfers and Borrowing Constraints on Educational Attainment. *International economic review*, 42(4):1051–1103.

Koşar, G. and O'Dea, C. (2022). Expectations Data in Structural Microeconomic Models. *Federal Reserve Bank of New York Staff Reports*.

Levine, P. B., Ma, J., and Russell, L. C. (2022). Do college applicants respond to changes in sticker prices even when they don't matter? *Education finance and policy*, pages 1–30.

Lochner, L. and Monge-Naranjo, A. (2008). The nature of credit constraints and human capital. Technical Report w13912, National Bureau of Economic Research, Cambridge, MA.

Manski, C. F. (1993). Identification of Endogenous Social Effects: The Reflection Problem. *The Review of economic studies*, 60(3):531–542.

Manski, C. F. (2004). Measuring Expectations. *Econometrica: journal of the Econometric Society*, 72(5):1329–1376.

Marcotte, D. E., Bailey, T., Borkoski, C., and Kienzl, G. S. (2005). The Returns of a Community College Education: Evidence from the National Education Longitudinal Survey. *Educational evaluation and policy analysis*, 27(2):157–175.

Marx, B. M. and Turner, L. J. (2019). Student Loan Nudges: Experimental Evidence on Borrowing and Educational Attainment. *American Economic Journal: Economic Policy*, 11(2):108–141.

Miller, G., de Paula, Á., and Valente, C. (2025). Subjective expectations and demand for contraception. *Journal of Econometrics*, 249(105997):105997.

Mulhern, C. (2021). Changing College Choices with Personalized Admissions Information at Scale: Evidence on Naviance. *Journal of labor economics*, 39(1):219–262.

Oreopoulos, P. and Dunn, R. (2013). Information and College Access: Evidence from a Randomized Field Experiment. *The Scandinavian journal of economics*, 115(1):3–26.

Ortagus, J. C., Skinner, B. T., and Tanner, M. J. (2021). Investigating Why Academically Successful Community College Students Leave College Without a Degree. *AERA Open*, 7:23328584211065724.

Papke, L. E. and Wooldridge, J. M. (1996). Econometric methods for fractional response variables with an application to 401(k) plan participation rates. *Journal of applied econometrics (Chichester, England)*, 11(6):619–632.

Peter, F., Spiess, C. K., and Zambre, V. (2021). Informing students about college: Increasing enrollment using a behavioral intervention? *Journal of economic behavior & organization*, 190:524–549.

Rau, T., Rojas, E., and Urzúa, S. (2013). Loans for Higher Education: Does the Dream Come True? (19138).

Solis, A. (2017). Credit Access and College Enrollment. *The journal of political economy*, 125(2):562–622.

Stacy Berg Dale and Krueger, A. B. (2002). Estimating the Payoff to Attending a More Selective College: An Application of Selection on Observables and Unobservables. *The quarterly journal of economics*, 117(4):1491–1527.

Stevens, A. H., Kurlaender, M., and Grosz, M. (2015). Career Technical Education and Labor Market Outcomes: Evidence from California Community Colleges. (21137).

Stinebrickner, R. and Stinebrickner, T. (2008). The Effect of Credit Constraints on the College Drop-Out Decision: A Direct Approach Using a New Panel Study. *The American economic review*, 98(5):2163–2184.

Stinebrickner, T. and Stinebrickner, R. (2012). Learning about Academic Ability and the College Dropout Decision. *Journal of labor economics*, 30(4):707–748.

Stinebrickner, T. R. and Stinebrickner, R. (2013). *Academic Performance and College Dropout: Using Longitudinal Expectations Data to Estimate a Learning Model*. National Bureau of Economic Research.

Tatiana Velasco, John Fink, Mariel Bedoya-Guevara, Davis Jenkins & Tania LaViolet (2024). Tracking Transfer: Community College and Four-Year Institutional Effectiveness in Broadening Bachelor's Degree Attainment. https://ccrc.tc.columbia.edu/publications/tracking-transfer-community-college-and-four-year-institutional-effectiveness-in-broadening-bachelors html. Accessed: 2025-7-27.

Todd, P. E. and Zhang, W. (2020). A dynamic model of personality, schooling, and occupational choice. *Quantitative economics*, 11(1):231–275.

Valletta, R. G. (2016). Recent Flattening in the Higher Education Wage Premium: Polarization, Skill Downgrading, or Both? (22935).

van der Klaauw, W. (2012). On the Use of Expectations Data in Estimating Structural Dynamic Choice Models. *Journal of labor economics*, 30(3):521–554.

Wang, T. (2024). How Do Agents Form Macroeconomic Expectations? Evidence from Inflation Uncertainty. Technical Report 2024-5.

Wang, Y. (2014). Dynamic Implications of Subjective Expectations: Evidence from Adult Smokers. *American economic journal. Applied economics*, 6(1):1–37.

Westphal, M., Kamhöfer, D. A., and Schmitz, H. (2022). Marginal College Wage Premiums Under Selection Into Employment*. *The Economic Journal*, 132(646):2231–2272.

Wiswall, M. and Zafar, B. (2015). Determinants of College Major Choice: Identification using an Information Experiment. *The Review of economic studies*, 82(2 (291)):791–824.

Wiswall, M. and Zafar, B. (2021). Human Capital Investments and Expectations about Career and Family. *The journal of political economy*, 129(5):1361–1424.

Zafar, B. (2011). CAN SUBJECTIVE EXPECTATIONS DATA BE USED IN CHOICE MODELS? EVIDENCE ON COGNITIVE BIASES. *Journal of Applied Econometrics*, 26(3):520–544.

A Data supplement

A.1 Filtering responses

I received 1,492 attempted survey responses. The initial count for successfully completed responses is 477. Due to an initial flaw in the survey settings, one individual submitted 13 of these, lowering that number to 464;[22] after seeing this, I changed the security settings on the survey to make multiple responses more difficult. While no one else attempted multiple completed responses, concern that some ineligible individuals may have restarted the survey and changed their answers to the screening questions motivated the construction of a simple algorithm for detecting repeat attempts. I isolate all instances where a respondent was ruled ineligible based on age or enrollment status and another attempt was made from the same IP address within two minutes. In this case, I drop the second response, on the assumption that this is likely to have been an ineligible respondent misrepresenting themselves. This leads me to filter out an additional 18 responses, leaving 446 valid submissions.

The survey included two attention checks where respondents were asked to set a dummy slider to a specific number. The more obvious attention check was an independent question, and was failed by 16 people. The second attention check was a slider within a larger question, likely a more subtle attention check than the first; this was failed by 51 respondents. I drop anyone who failed either of the two attention checks, which meant removing 63 additional responses, leaving me with 379. Finally, I drop an additional 60 responses with unrealistic wage expectations: either wage growth larger than 10x between two ages, or wage expectations in the bottom or top 2%. After all of these filters, I am left with **319** high quality responses from unique eligible individuals who passed attention checks.

A.2 Demographics

Table 12 shows the full set of demographic characteristics for the sample and, where possible, for the student body at large.

As discussed in section 4, I allow respondents to check all racial categories with which they identify. In order to compare my statistics against those from AACC, which allow only one category per-person, I use the following procedure:

1. If a respondent only checks one category (83% of the sample), they are simply considered in that category.

[22] That individual was not compensated and none of their responses are used.

2. If a respondent identifies with more than one race and identifies as Hispanic, I assume this is the most salient identity.
3. If a respondent identifies with more than one race and does not identify as Hispanic, I distribute their weight proportionally to all categories they identify with (including multi-race).

The implicit assumption behind (3) is that if non-Hispanic respondents were forced to choose a most salient identity, they would do so randomly from the list of identities they chose. In the second-round survey work, I will be able to ask questions about salience which will allow me to relax this. In any case, this procedure is only important for comparisons between the population statistics from the registrar and my sample; in the regression analyses, there is no problem in using indicator variables for multiple overlapping racial identities. Table 13 shows the differences in statistics on race before and after this procedure is followed.

B Fitting distributions

As discussed in subsection 4.2, I follow Engelberg et al. (2009) in and Armantier et al. (2016) in fitting parameteric distributions to my elicited densities for earnings and tuition costs. When respondents only assign positive probability to one interval, I assume a uniform distribution; two intervals, I fit an isosceles triangular distribution; and with three or more intervals, I fit a generalized beta distribution. Here I will provide additional details about each case.

Uniform distributions

If respondents assign positive probability to only one interval,[23] $p_a \in [a_1, a_2]$, I do not have any ability to identify any parameters. Instead, I assume a uniform distribution, $X \sim U[a_1, a_2]$. The mean of X is $\frac{a_1 + a_2}{2}$, the median is the same, and the mode is not defined. If respondents assign all of their probability to the leftmost interval, $a_1 = 0$ and there is no problem fitting a uniform distribution. If they assign all of their probability to the rightmost interval, which is bounded by ∞, then I assume that the width of their distribution corresponds to the same as the widths of the three bounded intervals they are asked about (that is, 30% of their point expectation).[24]

Positive probability in non-adjacent intervals

If respondents report positive probability in 3 or more non-adjacent intervals, there is no problem fitting a generalized beta distribution, which will be discussed below.[25] If respondents report positive probability in just two non-adjacent intervals, however, I cannot fit a generalized beta distribution (as I can only identify one parameter), and an isosceles triangular distribution would have an inappropriate shape. Instead, I fit a mixture of uniform distributions, with the mixing parameter given by the mass in each interval.

Assume the respondent assigns mass p_a into interval $[a_1, a_2]$ and $p_b = 1 - p_a$ into interval $[b_1, b_2]$. Let $X_a \sim U[a_1, a_2]$ and $X_b \sim U[b_1, b_2]$. Then the variable X is distributed $X \sim p_a X_a + p_b X_b$. It has mean $\mathbb{E}[X] = p_a \frac{a_1 + a_2}{2} + p_b \frac{b_1 + b_2}{2}$.

The median is slightly more complicated. Let $h_a = \frac{a_2 - a_1}{p_a}$ and $h_b = \frac{b_2 - b_1}{p_b}$. if $p_a \geq .5$, then the median is $a^* = \frac{.5}{h_a} + a_1$. If $p_a < .5$, then the median is $a^* = \frac{\frac{1}{2} + p_a}{h_b} + b_1$.

As with the uniform distribution, the mode is not defined. In the full sample, for tuition expectations, only about 2% of the sample assigned positive probability to two non-adjacent intervals.

[23] In expectations for tuition costs, about 18% of the sample reports positive probability in only one interval.

[24] An alternative would be to assume that the width is the same as the width of the leftmost interval.

[25] In reality, this behavior is not seen in my dataset.

Triangular distributions

Now consider the case where a respondent assigns positive probability to two adjacent intervals, that is, $p_1 \equiv P(a \in [a_1, a_2]) > 0$ and $p_2 \equiv P(a \in [a_2, a_3]) > 0$; for tuition expectations, about 22% of the full sample fits into this group. We want to fit an isosceles triangular distribution to best match this distribution. If the intervals are of equal width and $p_1 = p_2$, then the problem is trivial; the triangle begins at a_1 and ends at a_3 and has its center (mean, median and mode) at a_2. We consider the more interesting cases where the two bins have different amounts of probability.

Right-skewed triangle

First consider the case where $p_2 > p_1$. We set a_3 as the right-most point of the triangle, and we are looking for a^*, the left-most point.

It will be useful to define the midpoint as $\bar{a} \equiv \dfrac{a^* + a_3}{2}$. The triangle has width $w = a_3 - a^*$, and since its area must integrate to 1, we know the height h is $\frac{1}{2}hw = 1 \iff h = \dfrac{2}{w} = \dfrac{2}{a_3 - a^*}$.

Let $f(x)$ be the PDF (eg the formula for the triangle). We want to find a^* such that $p_1 = \int_{a_1}^{a_2} f(x)dx$ and $p_2 = \int_{a_2}^{a_3} f(x)dx$. Since the midpoint $\bar{a} \in [a_2, a_3]$, and we know the area to the right of $\bar{a} = \frac{1}{2}$, $\int_{a_2}^{a_3} f(x)dx = \int_{a_2}^{\bar{a}} f(x)dx + \frac{1}{2}$.

To find $\int_{a_2}^{\bar{a}} f(x)dx$ we need the formula for the line forming the left side of the triangle. Using the points $(a^*, 0)$ and $(\bar{a}, h)$ we can see that:

$$m = \frac{h - 0}{\bar{a} - a^*} = \frac{h}{\frac{a^*}{2} + \frac{a_3}{2} - a^*} = \frac{h}{\frac{a_3}{2} - \frac{a^*}{2}} = \frac{2h}{a_3 - a^*} = \frac{2h}{w} = h^2$$

We can find the y-intercept b by comparing the point $(0, b)$ to $(a^*, 0)$ in the slope formula, seeing:

$$h^2 = \frac{b}{-a^*} \iff b = -h^2 a^*$$

So the line is described by $y = h^2 x - h^2 a^*$.

We take the integral from a_2 to $\bar{a}$:

$$
\begin{aligned}
\int_{a_2}^{\bar{a}} \left(h^2 x - h^2 a^*\right) dx &= \left(\frac{h^2 x^2}{2} - h^2 a^* x\right) |_{a_2}^{\bar{a}} \\
&= h^2 \left(\frac{\bar{a}^2}{2} - a^* \bar{a} - \frac{a_2^2}{2} + a^* a_2\right) = h^2 \left(\frac{1}{2}\left(\bar{a}^2 - a_2^2\right) - a^* \left(\bar{a} - a_2\right)\right) \\
&= h^2 \left(\frac{1}{2}\left(\bar{a} + a_2\right)\left(\bar{a} - a_2\right) - a^* \left(\bar{a} - a_2\right)\right) \\
&= h^2 \left(\bar{a} - a_2\right) \left(\frac{1}{2}\left(\bar{a} + a_2\right) - a^*\right)
\end{aligned}
$$

Noting that $\bar{a}$ is a function of a^*, fitting the triangular distribution to p_1 and p_2 means finding a^* to minimize:

$$
\begin{aligned}
L(a^*; p_1, p_2) = &\left(p_1 - \left(1 - \frac{1}{2} - h^2 \left(\bar{a} - a_2\right)\left(\frac{1}{2}\left(\bar{a} + a_2\right) - a^*\right)\right)\right)^2 + \\
&\left(p_2 - \left(\frac{1}{2} + h^2 \left(\bar{a} - a_2\right)\left(\frac{1}{2}\left(\bar{a} + a_2\right) - a^*\right)\right)\right)^2
\end{aligned}
$$

Left-skewed triangle

We will now consider the opposite case where $p_2 < p_1$. We set a_1 as the left-most point of the triangle, and we are looking for a^*, the right-most point.

Define the midpoint as $\bar{a} \equiv \dfrac{a_1 + a^*}{2}$. The triangle has width $w = a^* - a_1$, and since its area must integrate to 1, we know the height h is $\frac{1}{2} hw = 1 \iff h = \dfrac{2}{w} = \dfrac{2}{a^* - a_1}$.

Let $f(x)$ be the PDF (eg the formula for the triangle). We still want to find a^* such that $p_1 = \int_{a_1}^{a_2} f(x)dx$ and $p_2 = \int_{a_2}^{a_3} f(x)dx$. Since the midpoint $\bar{a} \in [a_1, a_2]$, and we know the area to the left of $\bar{a} = \frac{1}{2}$, the area for p_1 is $\int_{a_1}^{a_2} f(x)dx = \frac{1}{2} + \int_{\bar{a}}^{a_2} f(x)dx$.

To find $\int_{\bar{a}}^{a_2} f(x)dx$ we need the formula for the line forming the right side of the triangle. Still the points $(a^*, 0)$ and $(\bar{a}, h)$ we can see that now:

$$
m = \frac{h - 0}{\bar{a} - a^*} = \frac{h}{\frac{a^*}{2} + \frac{a_1}{2} - a^*} = \frac{h}{\frac{a_1}{2} - \frac{a^*}{2}} = \frac{2h}{a_1 - a^*} = \frac{2h}{-w} = -h^2
$$

We can find the y-intercept b by comparing the point $(0, b)$ to $(a^*, 0)$ in the slope formula, seeing:

$$
-h^2 = \frac{b}{-a^*} \iff b = h^2 a^*
$$

So the line is described by $y = -h^2 x + h^2 a^*$.

We take the integral from $\bar{a}$ to a_2:

$$\begin{aligned}\int_{\bar{a}}^{a_2} \left(-h^2 x + h^2 a^*\right) dx &= \left(-\frac{h^2 x^2}{2} + h^2 a^* x\right) |_{\bar{a}}^{a_2} \\ &= h^2 \left(-\frac{a_2^2}{2} + a^* a_2 + \frac{\bar{a}^2}{2} - a^* \bar{a}\right) = h^2 \left(\frac{1}{2}\left(\bar{a}^2 - a_2^2\right) - a^* \left(\bar{a} - a_2\right)\right) \\ &= h^2 \left(\frac{1}{2}\left(\bar{a} + a_2\right)\left(\bar{a} - a_2\right) - a^* \left(\bar{a} - a_2\right)\right) \\ &= h^2 \left(\bar{a} - a_2\right) \left(\frac{1}{2}\left(\bar{a} + a_2\right) - a^*\right)\end{aligned}$$

Noting that $\bar{a}$ is a function of a^*, fitting the triangular distribution to p_1 and p_2 means finding a^* to minimize:

$$\begin{aligned}L(a^*; p_1, p_2) = &\left(p_1 - \left(\frac{1}{2} + h^2 \left(\bar{a} - a_2\right) \left(\frac{1}{2}\left(\bar{a} + a_2\right) - a^*\right)\right)\right)^2 + \\ &\left(p_2 - \left(1 - \frac{1}{2} - h^2 \left(\bar{a} - a_2\right) \left(\frac{1}{2}\left(\bar{a} + a_2\right) - a^*\right)\right)\right)^2\end{aligned}$$

Generalized beta distributions

When respondents assign probability to 3 or more intervals, I fit a generalized beta distribution; roughly 62% of the sample falls into this category. The generalized beta distribution is a highly flexible unimodal distribution defined by 2 shape parameters, α and β, and two parameters l and r which fix the support. It has the following form:

$$f(x) = \begin{cases} 0 & \text{if } x < l \\ \frac{(x-l)^{\alpha-1}(r-x)^{\beta-1}}{B(\alpha,\beta)(r-l)^{\alpha+\beta-1}} & \text{if } l \leq x \leq r \\ 0 & \text{if } x > r \end{cases} \tag{23}$$

where $B(\alpha, \beta) = \Gamma(\alpha)\Gamma(\beta)/\Gamma(\alpha+\beta)$. Let θ bet the vector of parameters to estimate. I am able to identify one fewer parameters than I have intervals with positive probability:

- If there is positive probability in three intervals, $\theta = \{\alpha, \beta\}$.
- If there is positive probability in four intervals, $\theta = \{r, \alpha, \beta\}$ or $\theta = \{l, \alpha, \beta\}$.
- If there is positive probability in five intervals, $\theta = \{l, r, \alpha, \beta\}$.

Estimating the parameters means solving the following minimization problem:

$$\theta^* = \text{argmin}_\theta \sum_{i=1}^{5} \mathbb{1}\left[p_i > 0\right] \times \left(p_i - \int_{a_{i-1}}^{a_i} f(x;\theta)dx\right)$$

I will handle separately the case of 3, 4, and 5 intervals with positive probability.

3 intervals

Here I set l to be the lower bound of the left-most interval with positive probability and r to be the bound of the right-most interval with positive probability. If the right-most interval is not bounded above, I assume it is of the same width as the interval to its left, in a similar way to how I fit the uniform distribution when all mass is placed in an endpoint interval.

4 intervals

Here I am able to estimate one endpoint parameter. As with the triangular distribution, I will assume that l is the lower bound of the left-most interval with positive probability if that interval has more mass than the right-most interval with positive probability, and then estimate r, and vice versa in the case that the right-most outer interval has more mass than the left-most.

5 intervals

Here I am able to estimate the full vector of parameters.

Tuition expectations

I apply this procedure to all of the continuous variables in my sample. This consists of college tuition, and a set of earnings expectations, at 3 points in the life cycle and for 3 different highest degree statuses. Results for the tuition expectations suggest a high degree of correlation between the elicity expectation and both the mean and median implied by the fitted parameteric distributions (corr ≈ 0.9). The results also suggest highly symmetrical distributions; this is not surprising in the case of the uniform or triangular distributions, which are symmetric by construction, but holds for the Beta distributions as well. Figures 13 - 15 show the results for tuition expectations.

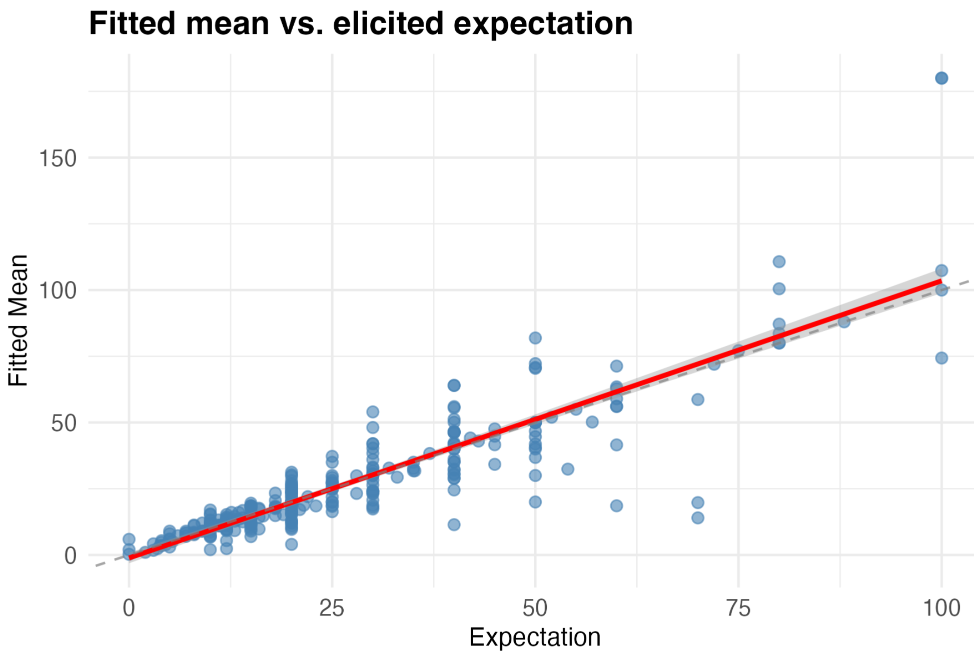


Figure 13: Correlation between tuition fitted mean and expectation

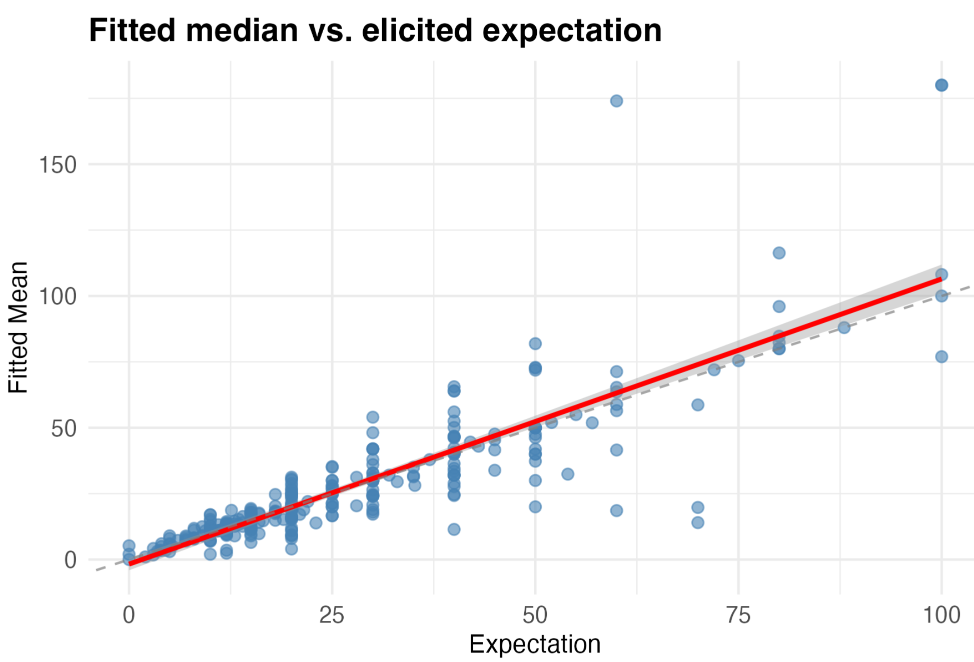


Figure 14: Correlation between tuition fitted median and expectation

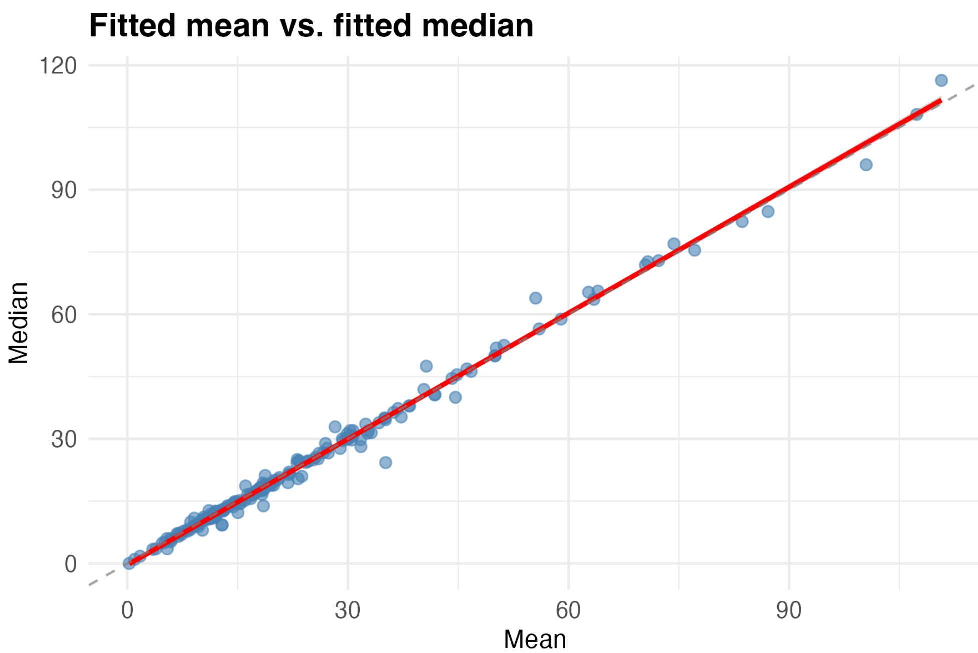


Figure 15: Correlation between tuition fitted mean and median

C Additional data

C.1 CPS

Figures 5 and 8 use statistics from the CPS (ASEC) to create a population benchmark to contextualize student expectations. In both cases, I filter out anyone younger than 22 or older than 52 (no expectations are elicited outside this age range) and any active students (since expectations are elicited conditional on working full-time, after completing education). I keep observations from 2004 to 2023; while there is nothing special about this 20-year period, it seems reasonable that students (most of whom are between 18 and 22) would benchmark their own beliefs to data from their lifetime. I weight on gender, to reflect the significantly female nature of my sample.

When examining earnings, I adjust earnings to 2024 dollars using the PCE. Because Maryland has higher wages than the US as a whole, I use only individuals from the state, which leaves me with 26,429 observations. For each age 5-year group, I calculate mean earnings by highest degree earned.

When examining labor force status, I keep 1,384,308 observations from the whole US. Unemployment status is taken directly from the CPS. Individuals working less than 13 weeks in the year (regardless of their hours) or less than 8 hours per week (regardless of their total weeks) are classified as non-workers; workers who do not fall into this group, but work less than 40 weeks in the year (regardless of hours) or work less than

30 hours per week (regardless of weeks) are classified as part-time; the remainder are classified as full-time.

C.2 HSLS

I begin with a sample of 5,871 whose first higher education institution is a 4-year program. I drop 304 records with no high school GPA, 1,107 with no information on family income, and 240 with no information on tuition, leaving me with a sample of 4,279. Of these, 65% (2,770) were enrolled in public, 4-year programs. Tuition is reported as tuition and fees charged at primary first year institution; I assume this was paid in 2013 (as year of entry is not available) and deflate all monetary variables to 2024 dollars.

Variable	Population	Sample	CI (sample)	Full sample	CI (full)
Female	-	67.4	(62.2, 72.5)	68.6	(64.3, 72.9)
Male	-	28.2	(23.3, 33.1)	27.8	(23.6, 31.9)
Nonbinary	-	4.4	(2.14, 6.6)	3.6	(1.87, 5.3)
Female (norm.)	59.9	70.5	(65.2, 75.8)	71.2	(66.3, 76.1)
Male (norm.)	40.1	29.5	(24.5, 34.6)	28.8	(24, 33.6)
Age	21.49	21.92	(21.5, 22.4)	22.02	(21.6, 22.4)
Mean HS GPA	-	3.22	(3.15, 3.29)	3.21	(3.15, 3.27)
Female	-	3.17	(3.08, 3.25)	3.17	(3.1, 3.24)
Male	-	3.33	(3.21, 3.44)	3.31	(3.21, 3.4)
Median HS GPA	-	3.36	-	3.33	-
Female	-	3.33	-	3.29	-
Male	-	3.5	-	3.4	-
Mean CC GPA	2.81	3.35	(3.15, 3.41)	3.3	(3.25, 3.36)
Female	2.86	3.33	(3.24, 3.42)	3.3	(3.22, 3.38)
Male	2.72	3.38	(3.24, 3.52)	3.33	(3.2, 3.45)
Median CC GPA	2.95	3.48	-	3.4	-
Female	3	3.42	-	3.4	-
Male	2.84	3.5	-	3.5	-
Part-time-school	62.4	42.6	(37.2, 48.1)	45.4	(40.7, 50)
Full-time-school	37.6	57.4	(51.9, 62.8)	54.6	(50, 59.3)
White	49.5	53.8	(48.3, 59.2)	50	(45.3, 54.6)
Black	20.3	19.1	(14.8, 23.4)	21.7	(17.9, 25.6)
Hispanic/Latino	17	15	(11.1, 19)	16.3	(12.8, 19.7)
Asian	5.1	6.9	(4.1, 9.6)	7.1	(4.7, 9.5)
Other	0.6	1.7	(0.31, 3.18)	1.8	(0.57, 3.04)
Multi-race	7.6	3.5	(1.45, 5.5)	3.2	(1.56, 4.8)
No work	-	27	(22.1, 31.8)	27.5	(23.4, 31.7)
Part-time work	-	24.8	(20, 29.5)	27.8	(23.6, 31.9)
Full-time work	-	48.3	(42.8, 53.8)	44.7	(40.1, 49.3)
Mean hourly wage	-	19.58	(18.3, 20.9)	19.73	(18.7, 20.8)
No FARM	-	57.4	(51.9, 62.8)	53.3	(48.6, 57.9)
Reduced-price lunch	-	10.3	(7, 13.7)	12.9	(9.8, 16)
Free lunch	-	32.3	(27.2, 37.4)	33.9	(29.4, 38.3)
50K or less	-	31	(26, 36.1)	30	(25.8, 34.3)
50-100K	-	27.3	(22.4, 32.2)	32.7	(28.4, 37.1)
100K or more	-	41.7	(36.3, 47.1)	37.2	(32.7, 41.8)
Arts and humanities	14.1	12.2	(8.6, 15.8)	12.9	(9.8, 16)
Business and economics	12.5	13.5	(9.7, 17.2)	13.1	(9.8, 16)
Medicine and health	22.9	26.3	(21.5, 31.2)	27.5	(9.8, 16)
STEM	22.3	27	(22.1, 31.8)	25.5	(9.8, 16)
Social science	7.8	6.3	(3.61, 8.9)	5.9	(3.68, 8.1)
Transfer studies	20.4	14.7	(10.8, 18.6)	15.1	(11.8, 18.5)
Has kids	-	11.9	(8.4, 15.5)	12.2	(9.1, 15.2)
No kids	-	88.1	(84.5, 91.6)	87.8	(84.8, 90.9)
Parent BA+	-	45.5	(40, 50.9)	42.9	(38.3, 47.5)
Parent HS-	-	54.5	(49.1, 60)	57.1	(52.5, 61.7)
N	6888	319	-	443	-

Table 12: Sample demographics

Category	Population	Elicited	Normalized
White	49.5	63.64	53.8
Black	20.3	25.08	19.1
Hispanic	17	15.05	15
Asian	5.1	10.03	6.9
American Indian or Alaska Native	0.4	3.13	_
Native Hawaiian or Pacific Islander	0.2	0.31	_
Middle Eastern or North African	-	1.88	_
Multi-race	7.6	16.61	3.5
Other	-	-	1.7

Table 13: Race