Sources of Bias in Teenagers' College Expectations

Benjamin W. Cowan[†]

April 7, 2015

Abstract

Objective: Though many studies have found that teenagers over-estimate their chances of college completion, the sources of this bias are still not well understood. *Methods*: This paper compares individuals' college expectations as teenagers with their subsequent college outcomes using data from the 1997 National Longitudinal Survey of Youth (NLSY97). I analyze how a rich set of youth characteristics correlate with the discrepancy between expectations and realizations. *Results*: Teenagers' expectations are highly predictive of future college completion, but they are also systematically positively biased (overly optimistic). I find that scholastic aptitude–as proxied by youths' AFQT scores–is highly negatively correlated with expectation bias. Once test scores are accounted for, family income, parental education, race/ethnicity, and sex have little or no effect on bias in teenagers' college expectations. *Conclusions*: The relationship between scholastic aptitude and college success may not be well understood by some youths, contributing to inflated expectations among those with lower test scores.

JEL Codes: D84, I24

Keywords: College expectations; adolescence; bias; test scores; income

1   Introduction

In an influential study, Manski [1993] shows that the set of variables youths condition on when forming expectations of their return to future schooling investments has important ramifications for educational behavior. Manski [2004] and others have argued for more research into the formation of subjective expectations in order to allow economists to scrutinize their assumptions about agents' expectations in their models.[1] In this vein, recent literature on the educational expectations of young people documents that teenagers tend to overestimate their chances of obtaining high levels of education (such a college degree), and that the gap between expectation and reality has grown over time (Jacob and Linkow, 2011; Reynolds, Stewart, MacDonald, and Sischo, 2006). This paper analyzes how the effects of family and individual characteristics on educational attainment compares to their effects on educational expectations in order to gain insight into the sources of bias in teenagers' expectations.

Given that it has been demonstrated that factors such as family income, parental education, and scholastic aptitude are robustly associated with schooling outcomes, a natural question is whether youths' expectations over future (uncertain) schooling events reflect these effects. In this study, I use data from the 1997 National Longitudinal Survey of Youth (NLSY97) on 15-18 year-old respondents' self-assessed probabilities of obtaining a 4-year college degree by age 30 combined with their college outcomes attained by age 30. Teenagers' college expectations are strongly correlated with future college completion even after controlling for a rich set of individual and family characteristics. However, I also find evidence

---

[1]Research on subjective educational expectations has a long tradition in the social sciences. For a review of this literature, see Jacob and Linkow [2011].

of systematic positive bias in expectations–on average, youths tend to be too optimistic about their chances of graduating from college. This is similar to the findings of Jacob and Linkow [2011], who employ data sources other than the NLSY97; it has also been suspected (but not formally tested) by others with respect to the NLSY97 data (see Walker, 2001; Reynolds and Pemberton, 2001).[2] I go beyond this initial result by showing that this bias is strongly correlated with a measure of youths' scholastic aptitude (namely, their Armed Forces Qualifying Test, or AFQT, score). Youths with poorer AFQT scores generally fail to lower their expectations to better match their objective chances of college completion. I also find that after controlling for AFQT scores, youths' family income, parental education, race/ethnicity, and sex have little to no effect on expectation bias.

An understanding of why youths make systematic mistakes in their expectations has potentially important policy implications.[3] First, since youths generally have inflated expectations about the likelihood of college, we might worry about the causes of such misinformation.[4] One possibility is that bias in expectations is mainly dependent on socioeconomic status and other characteristics associated with where youths live and go to school (since in some areas information about college is likely to be poor).[5] My results imply that many factors that are correlated with socioeconomic status and geography–income, mother's education, race, urban status–do not independently contribute to a lack of accuracy in ex-

---

[2]Relatedly, Zafar [2011] finds that college students are overly optimistic about their future academic performance.

[3]Several recent papers outside of the subjective expectations literature also find that youths' information about college is lacking (see, for example, Hoxby and Avery, 2012 and Bettinger, Long, Oreopoulos, and Sanbonmatsu, 2012).

[4]Sabates, Harris, and Staff [2011] find that "misaligned ambitions" (educational expectations that are not in harmony with career aspirations) contribute to unemployment and lower wages in adulthood.

[5]Other studies find that the gap between expectations and outcomes among low-income individuals is not due to different expectations about the cost and benefits of higher education nor the influence of those expectations in making college plans (Avery and Kane, 2004; Rouse, 2004).

pectations. Rather, to the extent that those variables influence the development of realistic expectations, they seem to do so through the channel of youths' academic performance (as proxied by the AFQT).

The large discrepancies in expectation accuracy by AFQT percentile suggests that many teenagers' knowledge of "what it takes" to get a college degree is lacking in that they do not comprehend the crucial role academic preparation plays in college success (this misinformation may be more highly concentrated in poorer areas, where test scores are typically lower). This is consistent with Jerrim [2014], who finds that Americans who score low on international tests are likely to have unrealistically high college expectations compared to low-scoring youths in other countries (who may receive better signals from schools about their relatively low likelihood of college success). If this is the case, policymakers who wish to improve the accuracy of teenagers' college expectations should find it useful to focus on improving knowledge of the strong relationship between high-school academic ability/achievement and college completion. However, it is also possible that AFQT scores are related to expectation bias for other reasons; thus, future work is needed to more fully understand this link.

Second, these findings suggest that relaxing the assumption of *rational expectations* and considering directly how youths form college expectations through adolescence may change our interpretation of current data. That is, youths may be making different educational choices than they would if their expectations did not exhibit systematic bias. Though performing counterfactual analysis in which the thought experiment is changing youths' expectations is beyond the scope of this paper, my results suggest this is a fruitful area of current and future research.[6]

---

[6]Recent studies in economics have begun incorporating subjective expectations into the estimation of

2 Existing Work on Subjective Expectations and Determinants of College Outcomes

The NLSY97 elicits college expectations of 15-17 year-olds (in 1997) via the question, "What is the percent chance you will have a four-year college degree by the time you turn 30?" Walker [2001] provides an early analysis of the responses to this question, though it is not his main focus.[7] Of 3,565 respondents who were eligible to answer the question, 3,511 (98.5 percent) did so. The high response rate suggests college graduation is salient to these youths.

However, Walker [2001] notes disturbingly high mean and median expected chances of finishing college among these youths (73 and 85 percent, respectively). Furthermore, almost 36 percent of youths report they have a 100 percent chance of realizing graduation. Reynolds and Pemberton [2001], who examine college expectations in the NLSY97 explicitly, report a college graduation rate of only 25-30 percent among 25 to 34 year-olds in 1998. This troubling discrepancy leads Walker [2001, pg. 206] to question whether these "assessments of college graduation entail more than subjective probabilities." In other words, perhaps NLSY97 youths report their college goals or aspirations rather than their expectations in response to the survey question.

Reynolds and Pemberton [2001] note that two-thirds of high school graduates were enrolled in college in 1997 and contend by that yardstick NLSY97 youths appear much more realistic.[8] Thus, another interpretation of the evidence is that respondents have trouble dif-

---

structural economic models. These include Stinebrickner and Stinebrickner [2012], Wiswall and Zafar [2011], Attanasio and Kaufmann [2012], and Arcidiacono, Hotz, and Kang [2012].

[7]Rather, he focuses on expectations regarding teenage parenthood.

[8]They note this fraction is very similar to the percentage of NLSY97 youths who gave themselves more than a 50 percent chance of receiving a degree.

ferentiating between college completion and college enrollment; this seems especially likely if youths are relying on their high school experience to assess how difficult advancing in college will be. Additionally, I later show that expectations are highly correlated with college completion, which suggests that youths largely take the NLSY97 survey question seriously. Similarly, Domina, Conley, and Farkas [2011] find that expectations are an important predictor of student effort in high school.

Reynolds and Pemberton [2001] are the first to analyze college expectations in the NLSY97; their goal is to isolate determinants of college expectations in the NLSY97 and its predecessor, the NLSY79 (and to examine how youths' college expectations have changed over time). However, their study was both too early to include youths' AFQT scores (which became available in 1999) as well as observations on these youths' college outcomes attained later in adulthood (2010-11). Thus, they cannot compare respondents' expectations with their outcomes or establish whether certain variables are correlated differently with expectations than they are with outcomes. These are the goals of this paper.

More recently, Jacob and Linkow [2011] use several longitudinal data sources (not including the NLSY97) to analyze how youths' college expectations and college outcomes have changed over time. The authors also examine sources of "misalignment" in expectations, which is similar to the goal of this study.[9] However, the expectations in the surveys used are elicited via asking youths the highest level of education they expect to obtain (less than a high-school diploma, high-school diploma, some college, or college degree or higher) as opposed to a subjective probability (with answers ranging between 0 and 100) over a particular

---

[9]The definition of expectation alignment is the predicted probability (based on teenage characteristics) that the youth obtains at least the level of education (s)he expects as a senior in high school.

event (completing a college degree by age 30), as in the NLSY97.[10] The disadvantage of the expectation type used in Jacob and Linkow [2011] is that youths may answer the question similarly (e.g. "some college" as the most likely educational outcome) but place very different probabilities on each educational outcome (e.g. "some college" might be way more likely than a college degree, or it might only be slightly more likely). Hence, it is difficult to ascertain exactly how confident a respondent is in reaching a particular level of schooling, which makes it harder to judge discrepancies in expectations and outcomes. The format of the expectation question in the NLSY97 is more conducive to analyzing expectation bias over large samples.[11]

3   Data

In this project, I use the NLSY97 dataset, an ongoing longitudinal survey that includes annual data since 1997 on young adults who were born between 1980 and 1984 (the most recent wave of available data is from 2011). The NLSY97 began with 8,984 youths including a black and Hispanic oversample. The data contains detailed information on youths' educational and labor-market experiences in addition to many other variables.

My analysis utilizes data on NLSY97 respondents who were born in 1980 or 1981 (younger respondents were not asked about their college expectations). Expectations were elicited in the initial 1997 survey (when youths were 15 to 18 years old), and I measure college outcomes

---

[10]Kim, Sherraden, and Clancy [2013] analyze a similar expectation measure to the one in Jacob and Linkow [2011] but for mothers of young children (rather than youths themselves). Their main focus is on differences in expectations across race/ethnicity and socioeconomic status.

[11]In addition, Jacob and Linkow [2011] do not analyze the individual effects of family income and parental education on expectation misalignment; rather, these variables (plus parental occupation) are analyzed via a composite measure of socioeconomic status.

in 2010 and 2011 (the years in which the 1980 and 1981 cohorts turn 30, respectively).[12] As stated in Section 2, the measure of expectations is a youth's self-assessed probability of obtaining a four-year college degree by age 30. In line with the question on expectations, my college outcome measure is a binary variable that takes a value of "1" if the individual has obtained at least 16 years of education (corresponding with attainment of a 4-year college degree) by age 30 and is "0" otherwise. For my purposes, I will call this "college degree" (CD).

The set of right-hand side variables used in the analysis includes youths' Armed Forces Qualifying Test (AFQT) scores. These represent a measure of a youth's performance on the word knowledge, paragraph comprehension, arithmetic reasoning, and numerical operations portions of a broader test called the Armed Services Vocational Aptitude Battery (ASVAB), which is normally used by the military. As part of the administration of the survey, NLSY97 youths were given the opportunity to take the ASVAB (roughly 80 percent of respondents took the exam). AFQT scores have been used extensively in the labor economics literature as a proxy for scholastic aptitude.[13]

Other control variables used throughout my analysis include family income in 1997 (measured in 1996 dollars), mother's education variables, number of siblings, race/ethnicity indicators (with "black" and "Hispanic" as separate included categories and "white" as the omitted category), sex, and binary variables indicating whether a respondent lives in the

---

[12]The vast majority of youths were 15-17 years old at the time of the 1997 survey, but since a few youths could only be interviewed in early 1998, they had already turned 18. It is reasonable to expect that college expectation accuracy improve with the age at which expectations are elicited (as college draws nearer). When I examine expectation bias for 15, 16, and 17/18 year-olds separately, I find that average bias is lower for older youths (0.42 for 15 year-olds vs. 0.38 for 17 and 18 year-olds) but that the effect of AFQT score on expectation bias is very similar across elicitation age.

[13]Neal and Johnson [1996, pg. 871] state that the AFQT is a "racially unbiased measure of basic skills that helps predict actual job performance."

south, whether (s)he lives in an urban area (as opposed to a rural area), whether (s)he resides with both biological parents, and whether (s)he is in the 1980 birth cohort (as opposed to the 1981 cohort).[14] Summary statistics for all variables used in the paper are contained in Table 1. Only the oldest two (out of five) NLSY97 cohorts are used in this study, as they were the only ones for whom the measure of expectations used in this paper was solicited. After excluding those with missing values for schooling outcomes at age 30 (due to sample attrition), parental income in adolescence, AFQT score, and other key variables, the study sample used throughout the paper contains 1,550 observations.

4   Results

4.1   Expectations as a Predictor of College Outcomes

Figure 1 shows the percentage of respondents falling into each of three expectation groups: those whose self-assessed probabilities of college completion were between zero and 40 percent (low), those whose expectations were between 41 and 80 percent (medium), and those with expectations in the 81 to 100 percent range (high). The percentage of respondents in each expectation category is also broken out by AFQT tercile.[15] This figure reiterates the fact that these youths are highly optimistic about college–over half believe their chances of graduating are at least 80 percent. Higher AFQT scores are generally associated with higher expectations, but even nearly 40 percent of teenagers in the lowest AFQT tercile believe

---

[14]This set of independent variables is very similar to those used by Carneiro and Heckman [2002] and Belley and Lochner [2007].

[15]AFQT scores in NLSY97 are reported as age-adjusted percentiles (based on the national distribution of scores). Thus, AFQT Tercile 1 comprises all those whose percentile score was between 0 and 33; AFQT Tercile 2 includes all those in the 34-66 range; and respondents in the 67-100 range make up AFQT Tercile 3.

college graduation is fairly certain.[16]

Figure 2 shows the percentage of respondents who had attained at least four years of post-secondary schooling at age 30 by 1997 college expectation group. Among the entire sample and each AFQT tercile separately, expectations are clearly positively correlated with obtaining a college degree. Thus, it appears that youths expectations are at least predictive of college outcomes. However, it is also apparent that expectations are biased upward; only 24 percent of kids who put their expectations in the middle range (41-80 percent) have attained a college degree, and only 52 percent of those in the highest range (81-100 percent) have done so. Given the large fraction of low and middle AFQT individuals who are in the highest expectation group, it is not surprising the bias is mainly concentrated among those in the lower end of the AFQT distribution.

I investigate the relationship between expecations and outcomes more formally in Table 2. Each of the left-hand columns under each sample heading shows the results of an OLS regression of the binary indicator for obtaining a college degree by age 30 on teenage college expectations (of realizing the outcome associated with the dependent variable) and a constant. Bernheim [1988, 1989, 1990] and Bernheim and Levin [1989] test the null hypothesis of *rational expectations* by examining the joint hypothesis that the expectation coefficient is equal to 1 (meaning a 1% increase in expectations is related to a 1% increase in the probability of realization, on average) and the constant is zero. This is an implication of *rational expectations* because if youths understood the true sources of randomness that affect their college outcomes, they would not make systematic mistakes (i.e. they would "get it right" on

---

[16]The averages displayed in both Figures 1 and 2 are weighted (with 1997 sampling weights) so that they are nationally representative for that year.

average). Table 2 reports that this hypothesis is rejected at the 1% level for the whole sample as well as each AFQT tercile sample separately. That the coefficient on expectations is less than 1 in every sample reiterates the claim that youths' four-year college degree expectations are overly optimistic (though the degree of bias clearly shrinks as AFQT percentile rises).

In spite of failing this test of *rational expectations*, youths' college expectations are strongly related to college outcomes. Expectations alone explain about 14% of the variation in college degree attainment in the full sample. Among the entire sample, a 1% increase in a youth's self-assessed likelihood of college graduation is associated with a 0.6% increase in the probability of attaining a college degree by age 30. When a rich set of family and individual characteristics is added to the model, the coefficient is roughly cut in half, but it remains statistically significant at the 1% level. The same basic pattern holds for every AFQT tercile. These results show that youths' expectations regarding future schooling outcomes have empirical content. This raises the possibility that one channel by which policymakers may be able to change college outcomes is manipulating the college expectations of young adults. Determining whether this is the case is difficult because the positive correlation between expectations and outcomes may not be causal (rather, it may simply reflect the effects of unobserved factors that contribute to schooling success and are correlated with expectations). Some previous work has shown that youths' expectations are influenced by *objective* differences in the return to college (e.g. Cowan, 2011), but more work is needed to untangle the mechanisms by which expectations and outcomes are related (Jacob and Linkow, 2011).[17]

---

[17]I have demonstrated that expectations are correlated with realized schooling, but do they reflect factors that go beyond those that affect actual schooling? To answer this question, I examined how the teenage college expectations used in the paper correlated with wages at age 30 controlling for realized schooling and the other background variables used in the paper. The effect of expectations is statistically significant (at

4.2   Youths' Characteristics and Expectation Accuracy

I now turn my attention to comparing the effect of certain key variables on expectations to their effect on college outcomes attained later in adulthood. The goal is to pinpoint which characteristics of youths are correlates of expectation accuracy–in other words, which characteristics have a significantly different effect on expectations than they do on college outcomes? To accomplish this task, I regress my measure of college success as well as youths' expectations on an identical set of covariates. These two equations take the following form:

$$CD_i = X_i\beta_1 + \varepsilon_{1i}, \tag{1}$$

$$CE_i = X_i\beta_2 + \varepsilon_{2i}, \tag{2}$$

where CD_i is the (binary) variable indicating attainment of a degree by age 30, CE_i is the youth's subjective probability (expectation) over that event, and X_i is a k-dimensional vector of characteristics (including a constant) that is the same across equations. I test whether $\beta_1 = \beta_2$ jointly as well as whether $\beta_1^1 = \beta_2^1, ..., \beta_1^k = \beta_2^k$ for each element of X individually. The first test provides a sense of whether the set of covariates as a whole has a similar association with college expectation and outcomes, while the second allows me to pinpoint which covariates (if any) differ in their correlation with expectations versus their correlation with outcomes. The most straightforward way to implement these tests is by subtracting

---

the 5% level) but relatively modest in magnitude: a 10 percentage point increase in the expected probability of college completion is associated with about a $400 increase in annual income (if the full complement of ASVAB scores is included in the regression, this falls to about $300 per year and is no longer significant at conventional levels). These results are available upon request.

one equation from the other:

$$CE_i - CD_i = X_i(\beta_2 - \beta_1) + (\varepsilon_{2i} - \varepsilon_{1i}) \equiv X_i\beta_3 + \varepsilon_{3i}. \tag{3}$$

Regressing the difference between CE_i and CD_i on X_i and then testing whether $\beta_3 = 0$ and $\beta_3^1 = 0, ..., \beta_3^k = 0$ is equivalent to testing for differences in coefficients across Equations (1) and (2). The average difference between CE and CD in the sample is 0.39, indicating that a negative coefficient in the expectation-bias regression (Equation (3)) means that increases in the corresponding variable reduces bias, on average. Of course, no particular youth's expectations can be labeled as "biased." Rather, we can only say which characteristics are associated with expectation bias by examining how those characteristics correlate with the difference between CE and CD over a large sample of youths. This is the aim of the regression analysis.

The results of the regressions of college success, expectations, and the difference between them on the set of individual and family characteristics discussed in the last section are presented in Table 3. All models are estimated using ordinary least-squares (OLS). I focus first on family income. Belley and Lochner [2007] use the NLSY97 to show that family income has become a more important determinant of college outcomes over the past 20-30 years.[18] Looking at the realization of a college degree in the first column of Table 3, the well-documented positive family income-college completion gradient shows up in separating the highest income quartile from the other quartiles: those youths are 12 percentage points

---

[18]Carneiro and Heckman [2002] perform a similar analysis using the older NLSY79 data on youths attending college in the late 1970's and early 1980's and find weaker effects of family resources on schooling outcomes.

more likely to graduate than youths in the lowest quartile (roughly a 33 percent difference at the mean).[19] Do youths anticipate that (only) those youth in the highest part of the income distribution have a better chance of obtaining a college degree, all else equal? The second column of Table 3 shows that youths are notably good at predicting how income correlates with college completion, as the effects on expectations are quite similar to those on college graduation. As a result, family income category has a negligible effect on the difference between college expectations and completion, which implies it has little effect on expectation bias.[20] Indeed, a test that the income quartile coefficients are jointly zero in the $CE - CD$ regression (Equation (3)) cannot be rejected at conventional levels.

AFQT scores are a different story. Compared to the lowest tercile, youths in AFQT Tercile 2 and Tercile 3 have a 15 and 42 percentage point higher chance of attaining a college degree, respectively. Expectations, however, are only 6 and 17 percentage points higher for these groups, on average. This discrepancy makes AFQT score the most economically significant determinant of expectation accuracy: youths in the highest tercile have a value of $CE - CD$ that is 26 percentage points (66 percent) smaller than those in the lowest tercile. While most youths are optimistic about college, those with high scholastic aptitude (as proxied by AFQT) are much more likely to actually receive a four-year degree.[21]

---

[19]These family-income effects closely resemble those reported in Belley and Lochner [2007], who also use NLSY97.

[20]Reynolds and Pemberton [2001] find a positive but insignificant effect of income (in levels) on expectations when they control for 8th grade GPA (but not AFQT score). Other variables' coefficient estimates from their regressions resemble their counterparts from my expectation regression reasonably well.

[21]Jacob and Linkow [2011] similarly find that youths' math and reading test scores are positively related to expectation "alignment." In contrast to my results, they find that youths' socioeconomic status is also positively related to alignment. However, once students' individual characteristics are controlled for, students' school characteristics explain very little of the variation in alignment.

Though a much larger percentage of individuals finish high school (86%) and obtain some college (63%) than complete college (37%), there is a stronger relationship between AFQT tercile and each of these outcomes than there is with subjective college expectations. However, the relationship between AFQT scores and high-school graduation is somewhat weaker than it is between college attendance/completion, suggesting

Other variables in the regression have notably similar effects on expectations and outcomes (and thus little effect on the difference between them). Both black and female youths are more likely to complete college, all else equal, and youths anticipate these effects almost perfectly.[22] Similarly, youths whose mothers have college experience properly rate their elevated chances of college completion relative to youths with less well-educated mothers. Other than AFQT, the only statistically significant predictors of expectation accuracy are residing in the southern region of the country and living with both biological parents. In the former case, the objective chances of degree attainment are the same by region but southern youths believe they are slightly more likely to graduate. In the latter case, the positive effect on college completion of living with both biological parents is underestimated by youths in their expectations.[23]

The last analysis I perform in this paper is to run the models described in Table 3 by each AFQT tercile separately. These results are contained in Table 4. The basic finding with respect to income in the full sample holds for each AFQT tercile separately as well: in no case can I reject (at conventional levels) the hypothesis that income coefficients are jointly zero in the $CE - CD$ regression. With respect to other variables, some interesting differences across AFQT tercile do arise. Perhaps most notably, the aforementioned large effect on expectation accuracy of living with both parents (in the full sample) is driven by what is happening among the brightest youths (highest AFQT tercile) in the data. In addition, though maternal education has a similar effect on expectations and outcomes in

---

that many low-AFQT, high-expectation youths are not making the transition from high school graduation to college attendance.

[22]Blacks are less likely than whites to graduate from college overall, but once other factors are controlled for (including family income and AFQT score), the marginal effect of being black is positive, a finding that is common in the literature (see, for example, Cameron and Taber, 2004 and Belley and Lochner, 2007).

[23]The null hypothesis that all coefficients are zero in the $CE - CD$ regression is rejected at the 1% level.

the full sample, having a mother with at least some college is strongly (negatively) correlated with expectation inaccuracy among youths in the highest AFQT tercile.[24]

5   Discussion

As described in Section 1, my finding that individuals' college attainments do not live up to their teenage expectations is similar to those of other recent studies that show the gap between expectations and reality has increased over the past several decades (Jacob and Linkow, 2011; Reynolds et al., 2006). The question then becomes why teenagers' expectations exhibit upward bias. Reynolds et al. [2006] demonstrate that the link between high-school grades and educational expectations has weakened over time, a finding that is consistent with my result that AFQT score (another measure of scholastic ability/achievement) is highly correlated with expectation accuracy. AFQT percentile has a large effect on the actual probability of college completion but a smaller effect on the self-assessed subjective probability of finishing college when one is still in high school; thus, individuals with low scores are less likely to realize their college expectations.

Why don't individuals with low test scores set their expectations to better match their objective chances of graduation? Because my empirical analyses include controls for family income, mother's education, race/ethnicity, urban status, and other variables, it is unlikely that the test-score gap in expectation accuracy results purely from poor information about college preparation in families, neighborhoods, or schools in which kids typically do poorly

---

[24]Once AFQT tercile has been controlled for (by running the regressions for each tercile separately), the null hypothesis that all coefficients are zero in the $CE - CD$ regression cannot be rejected at the 5% level for the lowest and middle AFQT terciles (but it is rejected for the highest tercile).

on national tests. Indeed, when the regressions shown in Table 3 are run for each family income quartile separately, the effect of AFQT tercile on expectation accuracy ($CE - CD$) is strong in every quartile and strongest in the higher income quartiles.[25] Rather, it seems this source of inaccuracy in educational expectations cuts across socioeconomic status (though it may be more concentrated among those in lower classes).

Recent work in sociology may provide some clues toward understanding this phenomenon. Rosenbaum [1997, 2011] argues that a "college for all" norm has emerged in the United States and may have contributed to teenagers underestimating the effects of poor academic performance on their ability to realize their educational plans. Jerrim [2014] contrasts the public-school system in the U.S. with those in others developed countries, many of which place greater emphasis on vocational training for students who do not appear well-suited to college in adolescence. This difference is perhaps one reason for why college expectations and outcomes are more aligned among low-achieving youths in other countries (Jerrim, 2014). Though more research is needed to understand the causes of biased college expectations on the part of American teens, my results suggest academic performance is likely a key piece of the puzzle.[26]

Just as striking as the fact that teenagers grossly fail to properly account for test scores in their college expectations (on average) is the finding that they do properly anticipate the role family income will play in their eventual college outcomes. In particular, youths seem to be

---

[25]These results are available upon request.

[26]I thank an anonymous referee for suggesting I look at whether graduation age plays a part in the relationship between AFQT scores and college completion. I indeed find that when I measure college completion at age 23 (rather than age 30), the relationship between AFQT score and graduation is weaker (results available upon request). This indicates that high-AFQT individuals are more likely to graduate at later ages, possibly because they are more likely to go back to school following economic downturns (such as the one in late 2007). Exploring this issue further is a topic for future research.

aware that those in the highest part of the income distribution are more likely to graduate. One might wonder: is this phenomenon good or bad? On one hand, we may interpret this as evidence that, all else equal, poorer teens do not have delusions about their likelihood of having success in college. Because they have an accurate view of family income as a determinant of college outcomes, they may be more able to prepare themselves for life after high school–whether it is at a four-year college or some other choice–than if they mistakenly believe income has no effect on outcomes. On the other hand, some may worry that kids who anticipate that low family income will harm their college chances may engage in behavior which in and of itself hurts their chances of becoming college-educated. In this way, low expectations of college success when individuals are young may become self-fulfilling.

Counterfactual analysis in which expectations are treated as a policy variable could shed light on the issues raised above. The answers to these questions could help us gain insights into the college-going behaviors of youths that go beyond what analysis without explicit consideration of expectations can offer.[27]

6   Conclusion

Using longitudinal data on individuals' teenage college expectations and eventual educational outcomes, I show that expectations are highly predictive of future schooling but also systematically biased upward. The most important observable determinant of expectation bias is youths' AFQT scores, a regularly used proxy for academic ability in economic stud-

---

[27]For example, Stinebrickner and Stinebrickner [2012] find that college students' subjective beliefs about future grades satisfy simple theoretical implications related to a student's decision to drop out but that beliefs constructed from typical assumptions in economics do not.

ies. Once these scores are controlled for, other factors such as family income, race/ethnicity, mother's education category, and sex have very little effect on expectation accuracy.

Two major themes emerge from these findings. The first is that incorporating subjective expectations into economic models should lead to a more precise understanding of behavior since those expectations do not satisfy the typical modeling assumptions (such as *rational expectations*). This also promises to shed light on how behavior and welfare would change if expectations were to change (e.g. through policy intervention). Second, an understanding of why teenagers overestimate their college completion chances is important inasmuch as youths fail to make optimal choices because of their (inaccurate) expectations (as has been suggested by many prior researchers). My results imply that many factors one might reasonably believe affect expectation accuracy–such as family income–only do so inasmuch as they affect kids' academic ability/achievement (as proxied by their AFQT scores). My findings also raise the possibility that the link between high-school academic performance and college success is not grasped by the typical teenager. Since AFQT scores may be correlated with other dimensions of youths' skills and personalities (many of which are unobserved in this study), more research is needed to pinpoint how this source of bias works.

References

Arcidiacono, P., V. J. Hotz, and S. Kang (2012): "Modeling college major choices using elicited measures of expectations and counterfactuals," *Journal of Econometrics*, 166, 3–16.

Attanasio, O. and K. Kaufmann (2012): "Education choices and returns on the labor and marriage markets: Evidence from data on subjective expectations," *Work. Pap., Bocconi Univ., Milan.*

Avery, C. and T. J. Kane (2004): "Student perceptions of college opportunities. The Boston COACH program," in *College choices: The economics of where to go, when to go, and how to pay for it*, University of Chicago Press, 355–394.

BELLEY, P. AND L. LOCHNER (2007): "The Changing Role of Family Income and Ability in Determining Educational Achievement," *Journal of Human Capital*, 1, 37–89.

BERNHEIM, B. D. (1988): *Social Security Benefits: An Empirical Study of Expectations and Realizations*, Stanford, Calif.: Hoover Institution Press, 312–345, Issues in contemporary retirement, rX: 1 (on Mar 24, 2007).

———— (1989): *The Timing of Retirement: A Comparison of Expectations and Realizations*, National Bureau of Economic Research Project Report series Chicago and London: University of Chicago Press, 335–355, The economics of aging.

———— (1990): *How Do the Elderly Form Expectations? An Analysis of Responses to New Information*, A National Bureau of Economic Research Project Report Chicago and London: University of Chicago Press, 259–283, Issues in the economics of aging.

BERNHEIM, B. D. AND L. LEVIN (1989): "Social Security and Personal Saving: An Analysis of Expectations," *American Economic Review*, 79, 97–102, rX: 1 (on Mar 24, 2007).

BETTINGER, E. P., B. T. LONG, P. OREOPOULOS, AND L. SANBONMATSU (2012): "The Role of Application Assistance and Information in College Decisions: Results from the H&R Block Fafsa Experiment," *The Quarterly Journal of Economics*, 127, 1205–1242.

CAMERON, S. V. AND C. TABER (2004): "Estimation of educational borrowing constraints using returns to schooling," *Journal of Political Economy*, 112, 132–182.

CARNEIRO, P. AND J. J. HECKMAN (2002): "The Evidence on Credit Constraints in Post-secondary Schooling," *Economic Journal*, 112, 705–734.

COWAN, B. W. (2011): "Forward-thinking teens: The effects of college costs on adolescent risky behavior," *Economics of Education Review*, 30, 813–825.

DOMINA, T., A. CONLEY, AND G. FARKAS (2011): "The link between educational expectations and effort in the college-for-all era," *Sociology of Education*, 84, 93–112.

HOXBY, C. M. AND C. AVERY (2012): "The Missing" One-Offs": The Hidden Supply of High-Achieving, Low Income Students," Tech. rep., National Bureau of Economic Research.

JACOB, B. A. AND T. W. LINKOW (2011): "Educational Expectations and Attainment," *Whither Opportunity?: Rising Inequality, Schools, and Children's Life Chances*, 133.

JERRIM, J. (2014): "The Unrealistic Educational Expectations of High School Pupils: Is America Exceptional?" *The Sociological Quarterly*, 55, 196–231.

KIM, Y., M. SHERRADEN, AND M. CLANCY (2013): "Do Mothers' Educational Expectations Differ by Race and Ethnicity, or Socioeconomic Status?." *Economics of Education Review*, 33, 82–94.

MANSKI, C. F. (1993): *Adolescent Econometricians: How Do Youth Infer the Returns to Schooling?*, A National Bureau of Economic Research Project Report. Chicago and London: University of Chicago Press, 43–57, Studies of supply and demand in higher education, rX: 5 (on Mar 24, 2007).

——— (2004): "Measuring Expectations," *Econometrica*, 72, 1329–1376.

NEAL, D. A. AND W. R. JOHNSON (1996): "The Role of Premarket Factors in Black-White Wage Differences," *Journal of Political Economy*, 104, 869–895, rX: 15 (on Mar 24, 2007).

REYNOLDS, J., M. STEWART, R. MACDONALD, AND L. SISCHO (2006): "Have adolescents become too ambitious? High school seniors' educational and occupational plans, 1976 to 2000," *Social problems*, 53, 186–206.

REYNOLDS, J. R. AND J. PEMBERTON (2001): "Rising College Expectations among Youth in the United States: A Comparison of the 1979 and 1997 NLSY," *Journal of Human Resources*, 36, 703–726.

ROSENBAUM, J. E. (1997): "College-for-all: Do students understand what college demands?" *Social Psychology of Education*, 2, 55–80.

——— (2011): "The Complexities of College for All Beyond Fairy-tale Dreams," *Sociology of Education*, 84, 113–117.

ROUSE, C. E. (2004): "Low-Income Students and College Attendance: An Exploration of Income Expectations*," *Social Science Quarterly*, 85, 1299–1317.

SABATES, R., A. L. HARRIS, AND J. STAFF (2011): "Ambition Gone Awry: The Long-Term Socioeconomic Consequences of Misaligned and Uncertain Ambitions in Adolescence," *Social science quarterly*, 92, 959–977.

STINEBRICKNER, T. AND R. STINEBRICKNER (2012): "Learning about Academic Ability and the College Dropout Decision," *Journal of Labor Economics*, 30, 707–748.

WALKER, J. R. (2001): *Adolescents' Expectations Regarding Birth Outcomes: A Comparison of the NLSY79 and NLSY97 Cohorts*, New York: Russell Sage Foundation, 201–229, Social awakening: Adolescent behavior as adulthood approaches, rX: 1 (on Mar 24, 2007).

WISWALL, M. AND B. ZAFAR (2011): "Determinants of college major choice: Identification using an information experiment," Tech. rep., Staff Report, Federal Reserve Bank of New York.

ZAFAR, B. (2011): "How do college students form expectations?" *Journal of Labor Economics*, 29, 301–348.

Figure 1: Distribution of 1997 College Expectations, by AFQT Tercile
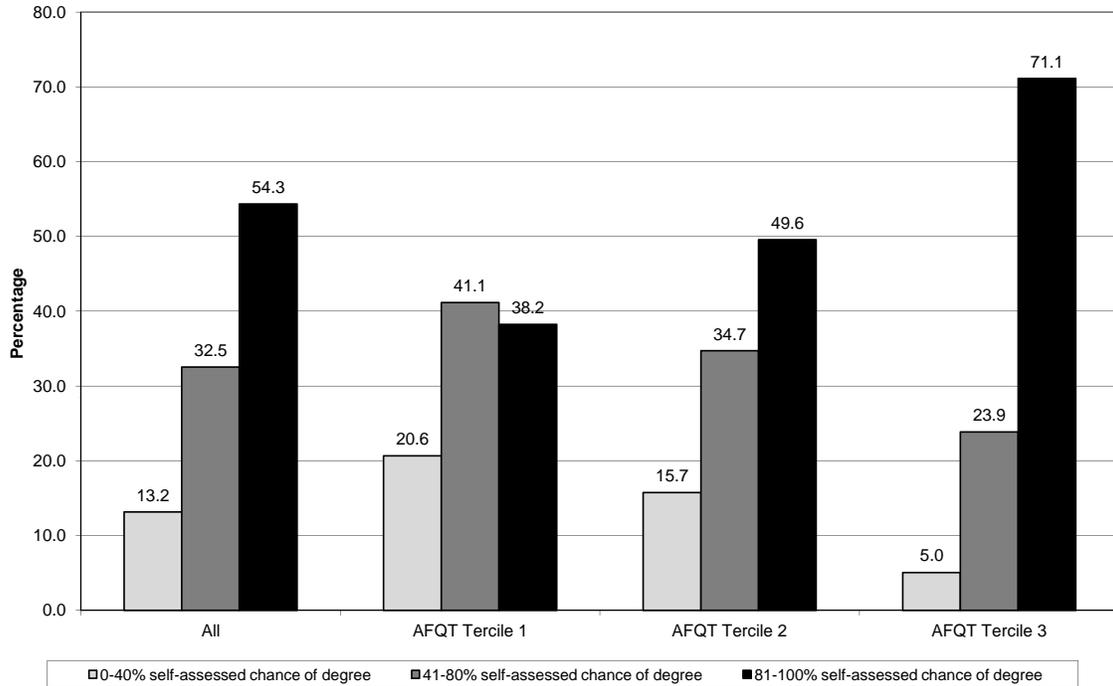
Figure 1: Distribution of 1997 College Expectations, by AFQT Tercile

Figure 2: Percentage of Respondents Who Obtained a 4-year College Degree by Age 30, by 1997 College Expectation and AFQT Score

Figure 2: Percentage of Respondents Who Obtained a 4-year College Degree by Age 30, by 1997 College Expectation and AFQT Score

21
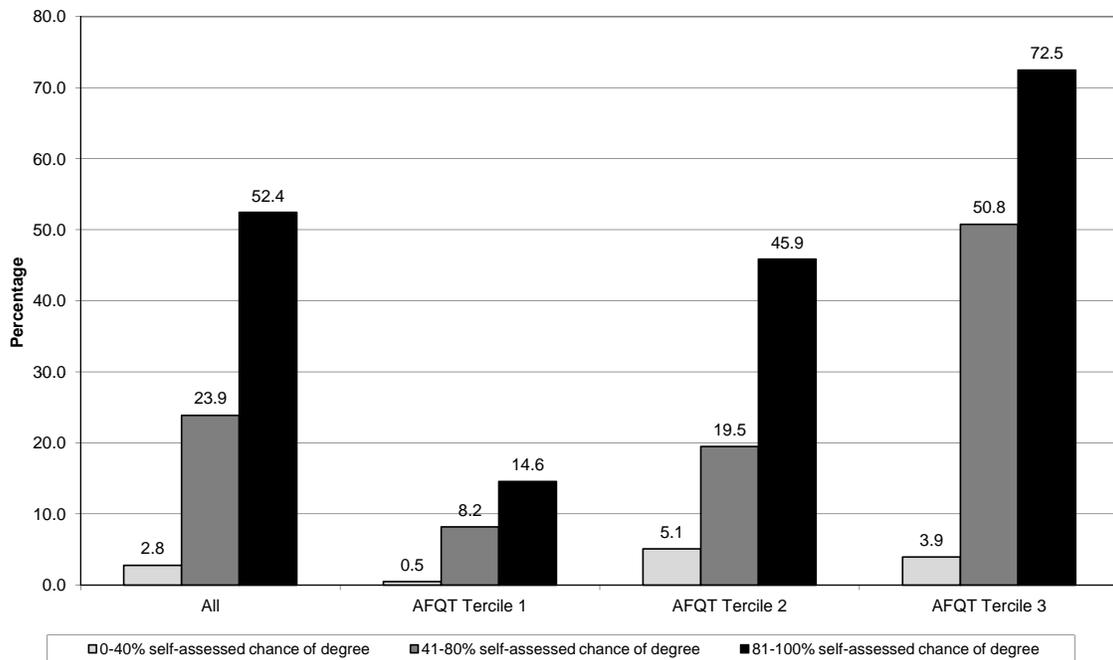
Table 1: Summary statistics for variables used in regressions

	Mean	Std. Dev.
Obtained a college degree by age 30 (CD)	0.37	0.48
Self-assessed percent chance of obtaining a college degree by age 30 in 1997 at ages 15-18 (CE)	0.76	0.30
Difference between CE and CD	0.39	0.46
Parental income ($1000 in 1996)	54.48	43.77
AFQT percentile	52.07	29.05
Resides in South	0.34	0.47
Resides with both biological parents	0.56	0.50
Resides in urban location	0.68	0.47
Number of siblings	2.05	1.48
Mother's education: high-school diploma	0.38	0.49
Mother's education: at least some college	0.45	0.50
Black	0.14	0.34
Hispanic	0.12	0.32
Female	0.50	0.50
Sixteen years old as of 12/31/96	0.48	0.50

Notes: N=1,550. Data are weighted by 1997 sampling weights.

Table 2: Results from Regressions of College Degree Attainment on Teenage College Expectations and Other Variables

	Full Sample		Low AFQT Tercile		Middle AFQT Tercile		High AFQT Tercile	
	No controls	W/ controls	No controls	W/ controls	No controls	W/ controls	No controls	W/ controls
Self-assessed percent chance of college degree by age 30	0.598***	0.301***	0.190***	0.138***	0.508***	0.369***	0.873***	0.682***
	(0.030)	(0.032)	(0.032)	(0.032)	(0.061)	(0.065)	(0.081)	(0.090)
Lower-middle income quartile		-0.013		0.007		-0.026		0.014
		(0.027)		(0.029)		(0.062)		(0.080)
Upper-middle income quartile		0.021		0.090**		-0.035		0.014
		(0.031)		(0.040)		(0.065)		(0.079)
High income quartile		0.098***		0.127**		0.030		0.082
		(0.036)		(0.056)		(0.070)		(0.081)
Middle AFQT tercile		0.134***		---		---		---
		(0.025)		---		---		---
High AFQT tercile		0.372***		---		---		---
		(0.030)		---		---		---
Resides in South		-0.011		0.022		-0.023		-0.060
		(0.022)		(0.025)		(0.045)		(0.044)
Resides with both biological parents		0.127***		0.023		0.136***		0.234***
		(0.023)		(0.029)		(0.043)		(0.049)
Resides in urban location		0.006		0.008		-0.026		0.017
		(0.023)		(0.028)		(0.045)		(0.042)
Number of siblings		-0.014**		-0.015**		-0.012		-0.011
		(0.006)		(0.006)		(0.013)		(0.016)
Mother's education: high-school diploma		-0.016		-0.008		-0.013		0.064
		(0.026)		(0.027)		(0.055)		(0.091)
Mother's education: at least some college		0.090***		0.042		0.058		0.218**
		(0.030)		(0.037)		(0.056)		(0.088)
Black		0.090***		0.063**		0.181***		0.059
		(0.027)		(0.031)		(0.059)		(0.068)
Hispanic		-0.031		0.071**		-0.109**		-0.093
		(0.029)		(0.036)		(0.051)		(0.074)
Female		0.067***		0.047*		0.096**		0.038
		(0.020)		(0.025)		(0.039)		(0.040)
Sixteen years old as of 12/31/96		-0.027		-0.024		-0.035		-0.017
		(0.020)		(0.026)		(0.039)		(0.039)
Constant	-0.116***	-0.183***	-0.025	-0.076	-0.070	-0.043	-0.124*	-0.297**
	(0.019)	(0.043)	(0.016)	(0.048)	(0.043)	(0.086)	(0.071)	(0.126)
Observations	1,550	1,550	584	584	478	478	488	488
R-squared	0.142	0.339	0.043	0.090	0.106	0.182	0.143	0.260
P-value associated with F-test that coefficient on expectations is "1" and constant is "0"	0.0000		0.0000		0.0000		0.0000	

Notes: Robust standard errors in parentheses. *** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 3: Results from Regressions of College Degree Attainment, Teenage College Expectations, and the Difference Between Them on Individual Characteristics

	Dependent Variable		
	Obtained College Degree by Age 30 (CD)	Self-assessed Percent Chance of College Degree at Age 30 (CE)	Difference between CE and CD
Lower-middle income quartile	-0.007	0.018	0.025
	(0.027)	(0.024)	(0.033)
Upper-middle income quartile	0.032	0.039	0.006
	(0.032)	(0.025)	(0.036)
High income quartile	0.124***	0.087***	-0.037
	(0.037)	(0.027)	(0.041)
Middle AFQT tercile	0.153***	0.062***	-0.091***
	(0.025)	(0.019)	(0.029)
High AFQT tercile	0.421***	0.165***	-0.256***
	(0.030)	(0.019)	(0.032)
Resides in South	0.005	0.053***	0.048**
	(0.022)	(0.015)	(0.024)
Resides with both biological parents	0.137***	0.034**	-0.103***
	(0.023)	(0.016)	(0.025)
Resides in urban location	0.013	0.025	0.012
	(0.024)	(0.016)	(0.025)
Number of siblings	-0.019***	-0.016***	0.003
	(0.006)	(0.005)	(0.007)
Mother's education: high-school diploma	-0.012	0.014	0.026
	(0.026)	(0.023)	(0.032)
Mother's education: at least some college	0.120***	0.098***	-0.022
	(0.030)	(0.024)	(0.034)
Black	0.117***	0.089***	-0.028
	(0.027)	(0.020)	(0.031)
Hispanic	-0.022	0.027	0.049
	(0.030)	(0.021)	(0.032)
Female	0.096***	0.097***	0.000
	(0.020)	(0.014)	(0.022)
	-0.038*	-0.035**	0.003
Sixteen years old as of 12/31/96	(0.020)	(0.014)	(0.022)
Constant	-0.028	0.513***	0.541***
	(0.041)	(0.035)	(0.049)
Observations	1,550	1,550	1,550
R-squared	0.309	0.191	0.112
P-value associated with F-test that coefficients on income quartiles are all "0"	0.0006	0.0015	0.3004

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

Table 4: Results from Regressions of College Degree Attainment, Teenage College Expectations, and the Difference Between Them on Individual Characteristics, by AFQT Tercile

	Low AFQT Tercile			Middle AFQT Tercile			High AFQT Tercile		
	Dependent Variable			Dependent Variable			Dependent Variable		
	CD	CE	CE - CD	CD	CE	CE - CD	CD	CE	CE - CD
Lower-middle income quartile	0.007	-0.005	-0.012	-0.008	0.047	0.056	0.035	0.031	-0.004
	(0.029)	(0.033)	(0.041)	(0.062)	(0.047)	(0.073)	(0.091)	(0.048)	(0.080)
Upper-middle income quartile	0.092**	0.017	-0.075	-0.015	0.053	0.069	0.054	0.059	0.005
	(0.040)	(0.039)	(0.052)	(0.066)	(0.049)	(0.075)	(0.090)	(0.048)	(0.078)
High income quartile	0.146**	0.132***	-0.013	0.066	0.099*	0.033	0.143	0.090*	-0.053
	(0.056)	(0.045)	(0.068)	(0.071)	(0.053)	(0.080)	(0.092)	(0.050)	(0.080)
Resides in South	0.028	0.046	0.018	0.012	0.096***	0.084*	-0.038	0.033*	0.071
	(0.025)	(0.029)	(0.036)	(0.045)	(0.026)	(0.048)	(0.045)	(0.020)	(0.044)
Resides with both biological parents	0.029	0.042	0.013	0.156***	0.054*	-0.102**	0.251***	0.024	-0.227***
	(0.029)	(0.029)	(0.038)	(0.044)	(0.028)	(0.046)	(0.052)	(0.024)	(0.049)
Resides in urban location	0.017	0.061*	0.044	-0.020	0.015	0.036	0.017	-0.001	-0.018
	(0.028)	(0.034)	(0.041)	(0.047)	(0.029)	(0.048)	(0.045)	(0.021)	(0.043)
Number of siblings	-0.016***	-0.012*	0.004	-0.021	-0.023***	-0.003	-0.022	-0.016**	0.006
	(0.006)	(0.007)	(0.009)	(0.013)	(0.008)	(0.014)	(0.017)	(0.007)	(0.017)
Mother's education: high-school diploma	-0.008	0.001	0.009	0.011	0.065	0.054	0.001	-0.091**	-0.093
	(0.027)	(0.033)	(0.040)	(0.055)	(0.043)	(0.065)	(0.092)	(0.045)	(0.093)
Mother's education: at least some college	0.059	0.120***	0.061	0.107*	0.133***	0.026	0.212**	-0.009	-0.221**
	(0.038)	(0.038)	(0.048)	(0.056)	(0.044)	(0.064)	(0.089)	(0.042)	(0.091)
Black	0.078**	0.107***	0.029	0.220***	0.106***	-0.114*	0.085	0.039	-0.046
	(0.031)	(0.034)	(0.043)	(0.059)	(0.032)	(0.061)	(0.070)	(0.031)	(0.070)
Hispanic	0.077**	0.037	-0.039	-0.096*	0.034	0.130**	-0.099	-0.008	0.091
	(0.037)	(0.036)	(0.047)	(0.052)	(0.039)	(0.057)	(0.078)	(0.030)	(0.073)
Female	0.060**	0.088***	0.029	0.143***	0.127***	-0.017	0.095**	0.083***	-0.012
	(0.025)	(0.026)	(0.034)	(0.040)	(0.025)	(0.042)	(0.041)	(0.018)	(0.040)
	-0.035	-0.082***	-0.047	-0.031	0.011	0.042	-0.029	-0.017	0.011
Sixteen years old as of 12/31/96	(0.025)	(0.026)	(0.033)	(0.040)	(0.026)	(0.042)	(0.041)	(0.018)	(0.040)
Constant	-0.007	0.502***	0.509***	0.134	0.480***	0.346***	0.252**	0.804***	0.553***
	(0.045)	(0.056)	(0.067)	(0.082)	(0.062)	(0.094)	(0.110)	(0.058)	(0.108)
Observations	584	584	584	478	478	478	488	488	488
R-squared	0.070	0.103	0.020	0.136	0.178	0.041	0.184	0.138	0.120
P-value associated with F-test that coefficients on income quartiles are all "0"	0.0104	0.0078	0.4877	0.4738	0.2556	0.7726	0.1733	0.1318	0.6603

Notes: Robust standard errors in parentheses. *** p<0.01, ** p<0.05, * p<0.1
CD = Obtained College Degree by Age 30; CE = Self-assessed Percent Chance of College Degree at Age 30; CE - CD = Difference between CE and CD.