Heterogeneous effects of socioeconomic status on postsecondary educational outcomes: Evidence from the Education Longitudinal Study of 2002

Wright, Harrison

University of Missouri

24 January 2025

Heterogeneous Effects of Socioeconomic Status on Postsecondary Educational Outcomes: Evidence from the Education Longitudinal Study of 2002

Harrison Wright*

January 23, 2025

*University of Missouri, Department of Economics, Columbia, MO 65211, United States. Email: hwright@mail.missouri.edu

Abstract

There is extensive economic literature on the returns to higher education, finding substantial benefits. However, access to higher education in the United States is unequal, with students of lower socioeconomic status (SES) attending and completing college at lower rates. This paper studies the effects of SES on these two postsecondary educational outcomes, utilizing data on a cohort of students from the Education Longitudinal Study of 2002 (ELS:2002). Employing both linear probability and probit maximum likelihood models, we analyze the ways SES influences college attendance and completion, while controlling for critical variables such as prior academic performance, high school quality, parental education, and demographic factors to isolate the effects to the financial resources and social capital associated with higher SES. Higher SES is associated with increased probabilities of both attending and completing college, though the marginal effects vary across the range of SES. The average partial effect of a unit increase in the SES variable (which ranges from -2.11 to 1.98) is a 7.3 percentage point increase in the probability of college attendance, and a 7.6 percentage point increase for bachelor's degree completion. There is significant heterogeneity across the range of the independent variable of interest: changes in SES have the highest (lowest) impact on attendance for low (high)-SES students, while the effect on degree completion is strongest for high-SES students and weakest for low-SES students. These results are tested for robustness to model specifications restricted by gender and race. The findings of this study highlight the continuing inequality in college access and degree attainment, with implications for policymakers and universities.

1. Introduction

In the United States, social and economic mobility are revered as hallmarks of a prosperous society. Although there is significant and growing inequality in wealth (Saez & Zucman, 2022), the belief that upward movement is achievable is a critical component of the American Dream. One potential path to moving up the socioeconomic ladder is formal education. In a report on the American middle class in 2023, the Pew Research Center reported that 87% of those with a bachelor's degree or higher lived in middle or upper-income households, compared with 63% of those with only a high school degree (Kochhar, 2024), although this descriptive statistic does not communicate causal effects. Researchers with the National Center for Education Statistics (NCES) found that most high school students in their study had high aspirations for college, with over 70% planning to earn a bachelor's degree (Dynarski, 2015). However, 13 years later, only 14% of the students from the most disadvantaged group had achieved this goal, compared with 60% of the most advantaged students. In another study of high school sophomores, 58% of the low-income students expected to graduate college, but only a quarter of them actually completed (compared with two thirds of the high-income students) (Bjorklund-Young, 2016). There is substantial economic literature on the returns to higher education that utilizes more robust analysis and finds economically meaningful benefits (Oreopoulos and Petronijevic, 2013).[1] Education is vital for mobility, rendering it essential to study college accessibility across economic status to ensure higher education remains an equalizer for low-income families.

2. Related Literature

There are many reasons that students from lower socioeconomic strata may experience these reduced rates of attendance and graduation, including financial constraints, lesser academic ability due to lower resources, or a lack of knowledge and support regarding higher education (Bjorklund-Young, 2016). In some

[1] Their paper is a robust review of the economic literature on the returns to higher education. It is a useful summary of a large portion of the body of work that finds substantial payoff to college attendance.

cases, this is due to the particular colleges that low-SES students attend, which tend to have lower graduation rates among all students (Bjorklund-Young, 2016). There is empirical evidence of significant segregation among colleges by parental income level, even when controlling for academic ability through SAT/ACT scores (Chetty et al, 2020)[2]. Among students with the same standardized test score, those from lower-income families attend less selective colleges, a phenomenon known as "undermatching" (Bowen, Chingos, and Mcpherson, 2009). Cardak and Ryan (2022) studied low-SES students in Australia and found that a variable for high school achievement (a proxy for academic ability) played a significant role in the reduced college outcomes, and controlling for this ability largely eliminated the gap in university attendance for these students of differing socioeconomic strata. This result echoed the findings of previous literature in the United States (Carneiro and Heckman, 2002). Carneiro and Heckman (2002) separate the relationship between family income and college attendance into short-run credit constraints and long-run family effects. Short-run financial constraints tend to be the assumed culprit, as college has become an expensive investment. However, the authors find that the long-run effects of wealthier families play an outsized role, influencing college attendance through better primary and secondary schools, different expectations and taste for schooling, and greater assistance with the child's schoolwork at home. This can be interpreted as a broader credit constraint for the children in lower-income families, as they are unable to purchase changes to their parental environment (Carneiro and Heckman, 2002). These long-run effects serve to further illustrate that socioeconomic status is a complex variable with a broad influence on a number of inputs that can affect a student's educational outcomes.

A meta-analysis of research on SES and student achievement between 1990 and 2000 found that higher SES is strongly associated with higher academic achievement, in part through the quality of the schools these students attended and the social capital of the families (Sirin, 2005). This social capital was found to have an

---

[2] Although it is not the focus of this paper, this undermatching can perpetuate economic inequality and limit the social mobility of students, to the extent that the more "elite" colleges have causal effects on earnings, a question discussed in much economic literature. For more, see Chetty et al (2020).

effect not only on the school attended, but the relationship between the parents and school faculty. The author highlights the fact that in the United States, school financing comes in large part from property taxes within a school district, causing students from neighborhoods with lower-SES families to have under-resourced schools. Hindering the study's ability to compare direct numerical effects for SES on achievement was the difference in methodologies and measurement used across the studies in the meta-analysis.

Format of the paper and contribution to the literature

In this paper, we contribute to the existing literature by studying the effects of students' socioeconomic status (SES) on their collegiate outcomes, both attendance and completion, with a particular focus on the heterogeneous effects across the range of SES. While most economic literature on educational outcomes makes reference to socioeconomic status, only a small number of papers directly address this relationship through empirical data. Among these, there is little attention given to heterogeneity in the effects of SES, even though the average effects fail to capture the complete picture. Section 2 of this paper discusses the dataset employed, a longitudinal survey of high school students nearing college age in the base year of the study. The features of this cohort are detailed through descriptive statistics, and the variables of interest and controls are identified. We control for many correlated variables in order to isolate the effect of SES on the outcomes. Section 3 describes the methodology, beginning with ordinary least squares (OLS) regression. Preliminary results from the OLS estimation are presented, along with potential concerns with the analysis and its assumptions. In section 4, these concerns are addressed through more rigorous analysis, including a discussion of omitted variable bias (OVB), implementation of a probit maximum likelihood regression model, and the addition of polynomials and interaction terms for flexibility of the model. The heterogeneous and nonlinear partial effects of SES are presented. In section 5, there is a discussion of the findings of these estimation methods, as well as possible takeaways for policymakers and higher education leaders. Finally, there is a conclusion, including a discussion of areas for further research. We find statistically and economically significant positive effects of socioeconomic status on both college attendance and bachelor's degree completion.

2. Data and descriptive statistics

The data come from the Education Longitudinal Study of 2002 (ELS:2002), which is published by the National Center for Education Statistics (NCES) under the U.S. Department of Education (National Center for Education Statistics, 2005).[3] The ELS:2002 is a longitudinal survey of students who were in the 10th grade in 2002. This cohort of students was then followed throughout the rest of their high school and postsecondary years and beyond, with robust surveys of the students, as well as their parents and teachers. Data collection rounds were completed in 2004, 2006, and 2012. The information provided helps researchers study a wide variety of questions on student trajectories and labor market outcomes. This dataset was selected due to the extensive information available regarding individual and school characteristics, providing a useful set of variables to pull from for a student's socioeconomic status, educational ability, and family background.

The ELS:2002 data was not gathered by the NCES in a simple random sample, but rather in a stratified, two-stage sample. In the first stage the school was selected, and in the second stage students were randomly sampled from that school. Since the dataset is not generated from a simple random sample, survey weights are calculated and provided by the NCES[4]. For the purposes of this study, however, we will use the raw data. As a result, there may be overrepresentation or underrepresentation of certain groups, but we do not expect this to have a significant impact on results.

For our study, the primary dependent (outcome) variables of interest are college enrollment and college completion. Binary variables were created for these outcomes using the ELS:2002's "date of first known

[3] An overview of the ELS:2002 study can be found here: https://nces.ed.gov/surveys/els2002/. To access the data requires using the NCES DataLab: https://nces.ed.gov/datalab/
[4] For further details on the survey design: https://nces.ed.gov/training/datauser/ELS_04/assets/ELS_04_transcript.pdf

postsecondary attendance" and "date when first known bachelor's degree was earned" (with a similar process for associate degrees). Our research question seeks to evaluate how socioeconomic status affects these outcomes. A composite of five equally weighted and standardized components is used to make up our primary independent variable "socioeconomic status", provided in the ELS:2002 as the variable BYSES2. The components are father's/guardian's education and income, mother's/guardian's education and income, and family income. For the occupation component, a prestige value was assigned using the 1961 Duncan Index (Duncan, 1961).

Next, we identify ideal controls to include in order to isolate the impacts of socioeconomic status on our outcomes. This helps prevent confounders and omitted variable bias (OVB), which will be discussed in further detail in section 3. The selected controls include high school GPA, high school quality, race, sex, region, and binary variables for whether health or military enlistment affected enrollment. High school GPA helps control for prior academic ability, which could have a significant impact on college enrollment and completion. We create an index variable for high school quality by using z-scores to standardize several inputs, which can be found in Table 1. Finally, we include a control for the amount of schooling parents desire for their 10th grader to complete, which could have an impact on the parental support for continuing education, monetary and otherwise.

Table 1: Components of variable "School Quality Index"

Source of data: Education Longitudinal Study of 2002

Variable	***ELS:2002 Variable Label***
Percent of 10th graders in school that receive free lunch	BY10FLP
Percent of students in dropout prevention program	BYA12E
Percent of full-time teachers that are certified	BYA24A
Percent of student body taking Advanced Placement courses	F1A22F
Percent of 2003 graduates that went to 4-year colleges	F1A19A

Summary statistics are shown in Table 2. The total number of observations is 16,197, representing individual students. Although not every variable contains a value for every observation, each variable in Table 2 has at least 14,796 observations, providing sufficient data for our analysis. In our sample, 72% of students attended college (either associate or bachelor's programs). 33% of students in the sample completed a bachelor's degree, which is approximately consistent with estimates of the total U.S. population over 25 in 2012, with roughly 31% having completed at least a bachelor's degree (National Center for Education Statistics, 2021). Socioeconomic status (SES) ranges from -2.11 to 1.98, a range of 4.09. The highest percentage of students in our dataset lived in the South, representing 36% of the sample. This outsized sample could affect postsecondary outcomes (to the extent that they differ by region) and will thus be controlled for as a factor variable in our regression specification. The same factor variable treatment will be applied to the variable "race".

Table 2: Summary Statistics

Source of data: Education Longitudinal Study of 2002

Variable	***Mean***	***Std. Dev.***	***Min***	***Max***
Binary Dependent Variables				
Attended college	0.71754	0.45021	0	1
Completed associate	0.09489	0.29308	0	1
Completed bachelor's	0.33000	0.47023	0	1
Completed assoc. and/or bach.	0.39143	0.48809	0	1
Independent Variables				
Socioeconomic status	0.04443	0.75031	-2.11	1.98
Socioeconomic status (quartiles)	2.57432	1.13189	1	4
Control Variables				
High school GPA*	3.91187	1.54270	0	6
Race	5.48731	1.92379	1	7
Sex	1.50208	0.50001	1	2
Region	2.58980	1.00899	1	4

High school urbanicity	1.84324	0.70435	1	3
School quality index	-0.00054	0.59781	-2.93	2.02
School amt. parents want 10th grader to complete**	5.38486	1.27478	1	7
Did not continue education due to:				
Health reasons	0.01111	0.10483	0	1
Military enlistment	0.00241	0.04901	0	1

**High school GPA is not a continuous variable. It is measured as a discrete variable from 1-6, with each response representing a GPA range. Thus, the mean should not be interpreted as the mean GPA.*

***Responses 1-7 represent: less than HS, HS or GED, attend/complete 2 year college, attend 4 year, graduate 4 year, master's or equivalent, PhD/MD/other advanced degree*

3. Preliminary analysis and OLS estimation

Now that we have identified the variables to include in our analysis and viewed summary statistics, we shift to the regression model specification. To help inform this selection and identify relationships, we first run a correlation of all variables of interest with each of the dependent variables. While the correlation is a helpful indicator, it is only one method of identifying which variables we include as regressors or controls. The correlation results are shown below in Table 3.

Table 3: Correlation between outcome and dependent variables

Source of data: Education Longitudinal Study of 2002	***Binary outcome variables***	
Variable	***Bachelor's degree completion***	***Bachelor's degree attendance***
Socioeconomic status	0.3782	0.2848
Socioeconomic status (in quartiles)	0.3515	0.2814
High school GPA	0.5177	0.3991
School quality index	0.2231	0.1472
Health interrupted school	-0.0663	-0.1061
Military enlistment interrupted school	-0.0277	0.0080
How much school parents want student to complete	0.2400	0.2369

This table shows the correlation between our two dependent variables of interest (bachelor's degree attendance and bachelor's degree completion) and other selected variables.

The correlations in Table 3 represent the strength and direction of the linear association between the two variables. Since the two outcome variables are binary, the coefficients can be interpreted specifically as the correlation between the probability of attending or completing college and the respective independent variable. We see positive relationships between many of the independent variables and the outcomes. This is unsurprising given our expectation that higher socioeconomic status positively impacts college attendance and completion. The strong relationship between high school GPA and college completion is also expected, as it is a measure of academic ability, and helps establish that this is a valid control variable. If this were excluded, we would expect to see upwardly-biased estimates of the causal effect of socioeconomic status on the outcomes of interest.

We also see positive correlations between the school quality index and both college outcomes. This would indicate that higher-quality schools (whether through better resources, better teacher quality, or simply the makeup of the students) send more students to college and put out more students that graduate college. We cannot make the causal inference that these high schools better prepare students for college outcomes, although this may be the case, as endogeneity could impact this relationship. Finally, there are negative correlations between college outcomes and health or military interruptions to school, though the associations are weak.

After observing the simple correlation between variables, we now turn to the regression specifications. In Tables 4 and 5, coefficient estimates from the OLS regression are provided for each of the outcome variables, with separate columns representing regressions with different controls included. The first column corresponds to the simple linear regression, with the outcome regressed only on the socioeconomic status (SES) variable. The estimate here represents the change in the probability of the individual attending or completing college for a one unit increase in the SES composite variable (which ranges from -2.11 to 1.98). The coefficient would indicate that a unit increase in the SES variable is associated with a 23.8% increase in the probability of the individual completing college, on average, and a 17.3% increase in the probability of the individual attending college. The respective 95% confidence intervals are [0.2291, 0.2467] and [0.1641, 0.1811], and both p values are less than .001, indicating that we have a high degree of confidence in the statistical significance of these estimates. However, we expect that these are biased estimates, with excluded variables captured by the error term having a significant impact on the outcomes. Omitted Variable Bias (OVB) is present when an excluded variable is both correlated with an existing independent variable and has an effect on our outcome, which we predict is the case for the variables added to the model in tables 4 and 5. We expect that the beta coefficient on SES in the simple linear regression has positive (or upward) bias, since the excluded variables high school GPA and school quality in particular should have a positive effect on both college attendance and college completion, and are likely correlated with SES. Families of higher socioeconomic status may have access to better quality schools, and are potentially able to offer more support for their students academically, which could in turn lead to higher high school GPA. This correlation would cause the effect of the omitted variables to be partially captured by the SES coefficient, overstating its estimated effect. The assumption of OVB is supported by the diminishing SES coefficient estimates as high school GPA, school quality, and other controls are added to the model. Additional controls include sex and factor variables for race, region, and urbanicity (urban, rural, or suburban). Finally, we include a vector of two variables for interruptions to schooling: health interruptions and military enlistment. We confirm that there is heteroskedasticity in the sample after conducting a White (1980) test and finding that $p<0.0001$. We utilize heteroskedasticity-robust standard errors in the regression models to address potential bias in the standard error estimates. This helps ensure the validity of our hypothesis testing for more accurate inference when the variance of the error term is non-constant (exhibiting heteroskedasticity).

The effect of socioeconomic status on the outcomes shrinks as additional controls are included and the model becomes more tightly fitted to the data (shown by the increasing R-squared value). This suggests that the initial effect was biased upward, and that excluded variables are both correlated with socioeconomic status and have an effect on the outcomes of interest. The results are still highly statistically significant, with p<.001. For the regression on college completion, we can reject the null hypothesis that the true coefficient is 0 for many of the explanatory and control variables, including socioeconomic status, which is this study's primary research focus. The variables high school GPA, school quality, parent's desired education level for the student as of 10th grade, and interruptions from health or military enlistment all have highly statistically significant effects in the dataset, and many are economically meaningful as well. As expected, there were negative effects on college completion for students who had an interruption to schooling related to health or the military. The effect of high school quality (using our index described in table 1) is positive, with a coefficient of 0.0736 (95% CI [.0616, .0855]), indicating that there is a 7.4% increase in the probability of college completion for a unit increase in the school quality index. However, it is important to note that this is not our primary explanatory variable, and the controls were specifically designed for parsing out the effect of SES.

Being female was associated with a 1.4% greater chance of completing college in the sample (95% CI [.0003, .0266]). While statistically significant at the 5% level, this is a relatively small economic impact. However, this result is consistent with other economic literature regarding the gender gap in college attendance and completion, including at the time of this dataset (Jacob, 2002). Among the race variables in the regression, while all of the groups had a positive effect relative to the reference group (American Indian or Alaska Native), only one of the 6 groups had a statistically significant effect: being in the Asian, Hawaiian, or Pacific Islander group had a positive effect of 8.6% on the probability of completing college when compared with the reference. This result is statistically significant with p<.01 and the confidence interval not containing 0. Among the other controls, urbanicity and region both had statistically meaningful impacts as well. Being from the northeast has a positive effect on college completion, with the midwest, south, and west regions all returning negative coefficients compared with the northeast reference. On a local level, being from a rural or suburban area also has negative effects, suggesting that students from urban areas are more likely to complete college.

Table 4: Preliminary regression results for bachelor's degree completion

Source of data: Education Longitudinal Study of 2002

Variables	*Bachelor's degree completion*	*Bachelor's degree completion*	*Bachelor's degree completion*
Socioeconomic status composite	0.2379	0.139	0.1141
	(0.0045)	(0.0050)	(0.0052)
High school GPA		0.1359	0.1304
		(0.0022)	(0.0023)
School quality index			0.0736
			(0.0061)
Health interrupted school			-0.114
			(0.0189)
Military enlistment interrupted school			-0.1697
			(0.0425)
How much school parents want 10th grade student to complete			0.0272
			(0.0024)
Observations	15,244	13,981	13,934
R-squared	0.1434	0.3293	0.3415
Controls incl.	No	Yes*	Yes*

The regression estimates are provided as additional controls are added. The second line under each variable provides the standard errors for each estimate.
**Additional controls not provided in the table include race dummies, sex, region, and urbanicity.*

When regressing with college attendance as the outcome, the results are qualitatively similar, with directionally identical findings across the independent variables. The magnitude of most coefficients is smaller, including socioeconomic status, indicating that college attendance may not be as strongly impacted by socioeconomic status as college completion. This could be a function of credit and financial constraints faced by college students, which is a significant cause of dropout (Desjardins et al, 2006), especially as the share of the financial costs of higher education has continued to shift to students (State Higher Education Executive Officers Association, 2023).

Although most of the effects of the independent variables were smaller for college attendance, we find a greater effect of parents' desired education level for their student. This could suggest that parental pressure is more effective at leading students to enroll in college than enabling them to see it through to graduation.

Table 5: Preliminary regression results for bachelor's degree attendance

Source of data: Education Longitudinal Study of 2002

Variables	***Bachelor's degree attendance***	***Bachelor's degree attendance***	***Bachelor's degree attendance***
Socioeconomic status composite	0.1726	0.1028	0.0813
	(0.0043)	(0.0049)	(0.0051)
High school GPA		0.0979	0.0894
		(0.0025)	(0.0026)
School quality index			0.0345
			(0.0063)
Health interrupted school			-0.3317
			(0.0391)
Military enlistment interrupted school			0.1397
			(0.0665)
How much school parents want 10th grade student to complete			0.0387
			(0.0032)
Observations	15,244	13,981	13,934
R-squared	0.084	0.1966	0.2145
Controls incl.	No	Yes*	Yes*

The regression estimates are provided as additional controls are added. The second line under each variable provides the standard errors for each estimate.
**Additional controls not provided in the table include race, sex, region, and urbanicity.*

While statistically and economically significant results are clear from these regression models, it is important to note potential concerns with the analysis as well. Equation (1) includes controls for certain hand-selected time-invariant variables, but the selection of variables could result in omitted variable bias if important characteristics are excluded, violating the exogeneity assumption for OLS regression. For example, we expect that parental education has a positive impact on college attendance and completion, and is also highly correlated with socioeconomic status and potentially school quality as well. This would indicate upward bias on the coefficient on socioeconomic status in our model. Blanden et al. (2022) analyze panel data across multiple countries and find a significant effect of parental education on students attending a university by age 20 and graduating by age 25, providing evidence for these assumptions. Peer effects, to the extent the student's contemporaries may have differing collegiate aspirations, could also play a significant role. If a student's peers attend college at a high rate and this variable is excluded, we would again expect upward bias on the SES coefficient.

4. Advanced analysis

We have highlighted several potential concerns with relying solely on the OLS regression model utilized previously. Even within the OLS specification itself, there could be omitted variable bias. Since OLS assumes a linear relationship, predicted results can fall outside the realm of reality for the binary outcomes (attending and completing college) that we are studying. In the ideal model, these predicted outcomes would remain bounded within the interval [0, 1], as they represent probabilities. Additionally, there could be non-linear relationships between the independent variables and the outcomes.

In this section, we address these concerns in multiple ways. First, we consider the effect on our coefficient estimates of including a previously-highlighted omitted variable, parental education. As discussed in the previous section, the White (1980) test confirmed the presence of heteroskedasticity, so we utilize heteroskedasticity-robust standard errors to ensure the validity of those estimates. Due to the limitations of OLS for our outcomes, we consider a maximum likelihood estimation procedure, specifically probit regression. Finally, in order to further address non-linearity across different values of the independent variables, we test different model specifications including interaction terms and second order polynomials.

We test for the correlation of parental education with the outcome variables and socioeconomic status and find significant correlation, suggesting there could be biased estimates and endogeneity. Tables 6 and 7 present the results of the original regression specification alongside the model including parental education. Table 6 provides the estimates for the degree completion outcome, where the effect of socioeconomic status diminishes from .1141 to .0946 when including the additional control; Table 7 (displaying the college attendance outcome) shows a diminished estimate as well, with the SES coefficient falling from .0813 to .0666. Both of these results are still highly statistically significant, and the individual estimates on parental education are significant at the 5% level as well.

Next, we transition to a maximum likelihood estimation model, utilizing probit regression. The probit model is based on a latent (or unobserved) version of the outcome variable, which we call Y^*. Since the model is concerned with the unobserved outcome, we are interested in the distribution of the error term. The probit model assumes a standard normal distribution with mean 0 and variance 1. This standard normal distribution gives the model an s-shape, allowing the probabilities to differ at varying levels of the independent variables. The probit regression specification provides important benefits. The relationship between the dependent and independent variables is now estimated to be nonlinear, which could allow for a better fit of the model to the data, specifically capturing diminishing marginal effects. The probit model is also bounded between 0 and 1 in its probability estimates, ideal for our binary outcome variables which cannot fall outside of this probability range in reality. These improvements over OLS for binary response variables should result in a more precise estimator and smaller standard errors. The probit model estimates the latent Y^*, so we cannot directly interpret the magnitude of the coefficient estimates, as $\hat{\boldsymbol{\beta}}$ represents the estimated average change in Y^* for a unit change in the independent variable, not Y. Instead, we take the average partial effect across the individuals, providing coefficients comparable to the linear probability model (LPM). Tables 6 and 7 present the probit model estimates alongside the LPM.

Table 6: Probit maximum likelihood estimation results compared with LPM for bachelor's degree completion

Source of data: Education Longitudinal Study of 2002

Variables	*Linear probability (OLS)*	*Linear probability (OLS)*	*Maximum Likelihood (Probit)*
Socioeconomic status composite	0.1141	0.0946	0.0761
	(0.0052)	(0.0080)	(0.0077)
High school GPA	0.1304	0.1303	0.1328
	(0.0023)	(0.0023)	(0.0022)
School quality index	0.0736	0.0732	0.0637
	(0.0061)	(0.0061)	(0.0062)
Health interrupted school	-0.114	-0.1132	-0.2811
	(0.0189)	(0.0190)	(0.0701)
Military enlistment interrupted school	-0.1697	-0.1716	-0.2091
	(0.0425)	(0.0429)	(0.0901)
How much school parents want 10th grade student to complete	0.0272	0.0269	0.0314
	(0.0024)	(0.0024)	(0.0029)
Parental education		0.0086	0.0086397
		(0.0026)	(0.0026)
Observations	13,934	13,934	13,934
R-squared	0.3415	0.3419	-
Controls incl.	Yes	Yes	Yes

The first column contains the original OLS estimate with all controls added. The second column adds parental education as an additional control, and the third column provides the estimates from the probit estimation.
**Additional controls not provided in the table include race, sex, region, and urbanicity.*

For the bachelor's degree completion outcome, the effect of socioeconomic status is smaller under the probit model than OLS, with an average marginal effect of 7.6 percentage points (pp) on the probability of completing college (95% CI [.0609, .0912]), suggesting that OLS may have overestimated the true effect. The probit model's shape flattens the effect of the independent variables at their extremes, and the linear relationship estimated by OLS may have biased the estimates upward. OLS assumes that socioeconomic status has a constant effect on degree completion, while probit's flexibility allows for the possibility of varying partial effects across the distribution. In Table 7, where we consider college attendance, the probit estimate is higher than OLS, with an average marginal effect of 7.4 pp (95% CI [.0584, .0901]).

Table 7: Probit maximum likelihood estimation results compared with LPM for bachelor's degree attendance

Source of data: Education Longitudinal Study of 2002			
Variables	***Linear probability (OLS)***	***Linear probability (OLS)***	***Maximum Likelihood (Probit)***
Socioeconomic status composite	0.0813	0.0666	0.0742
	(0.0051)	(0.0082)	(0.0081)
High school GPA	0.0894	0.0893	0.0799
	(0.0026)	(0.0026)	(0.0022)
School quality index	0.0345	0.0343	0.0399
	(0.0063)	(0.0063)	(0.0063)
Health interrupted school	-0.3317	-0.3311	-0.2478
	(0.0391)	(0.0391)	(0.0341)
Military enlistment interrupted school	0.1397	0.1383	0.1205
	(0.0665)	(0.0666)	(0.0703)
How much school parents want 10th grade student to complete	0.0387	0.0384	0.0327
	(0.0032)	(0.0032)	(0.0027)
Parental education		0.0065	0.0056
		(0.0029)	(0.0027)
Observations	13,934	13,934	13,934
R-squared	0.2145	0.2148	-
Controls incl.	Yes	Yes	Yes

The first column contains the original OLS estimate with all controls added. The second column adds parental education as an additional control, and the third column provides the estimates from the probit estimation.
**Additional controls not provided in the table include race, sex, region, and urbanicity.*

Standard errors for the effect of socioeconomic status are smaller for both outcomes under maximum likelihood estimation, which indicates greater efficiency of the estimator. Provided that the conditional distribution of the outcome variable is correctly specified, we expect to see efficiency and consistency in the maximum likelihood estimator. The probit regression assumes a normal distribution for the latent variable.

Next, we address additional non-linearity in the effects of the independent variables through polynomials and interaction terms, which account for more complex relationships between the explanatory variables and the probability of attending or completing college. For example, socioeconomic status may have higher effects at the middle of the distribution, where a change in resources could be the difference between a family being able to afford college or not. At the right tail of the distribution, where a family already has significant wealth, the effect of additional financial means may have a lessened impact. We include quadratics of the variables for SES, high school GPA, school quality index, and parental education. We also include interaction terms between SES and both high school GPA and parental education. Interaction terms allow the partial effect of each interacted variable to depend on the value of the other. The partial effect of socioeconomic status on the outcomes is now non-constant, and depends on the two variables that are interacted with SES as well. For the probit model, the interpretation with polynomials is now more complex. Ai and Norton (2003) note that many applied researchers misinterpret the coefficient on interaction terms due to this complication in nonlinear models. We consider the following probit model for a binary response variable, adapted from Greene (2010) and including a quadratic and interaction term:

$$P(y = 1) = E(y|x_1, x_2, z) = \Phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_{12} x_1 x_2 + \delta z) \tag{1}$$

where $\Phi(\beta_1 x_1 + \beta_2 x_2 + \beta_3 x_1^2 + \beta_{12} x_1 x_2 + \delta z)$, or $\Phi(X\beta)$, is the standard normal cumulative distribution function (CDF) of the linear index of explanatory variables, and z is the vector of additional controls. For the marginal effect of x_1 on the outcome y, we take the partial derivative with respect to x_1:

$$\frac{\partial E(y|x_1, x_2, z)}{\partial x_1} = \Phi'(X\beta) \times (\beta_1 + 2\beta_3 x_1 + \beta_{12} x_2) \tag{2}$$

Now, we can clearly see that the partial effect of x_1 on the outcome y is dependent on the value of x_1 (introducing nonlinearity through the quadratic term), as well as $\Phi'(X\beta)$, which is the derivative of the standard normal CDF. This derivative gives the steepness of the CDF curve at $X\beta$, and scales the strength of the effect of $X\beta$ on the probability that y is equal to 1. Additionally, the partial derivative of the interaction term $\beta_{12} x_1 x_2$ means that the effect also depends on the value of x_2, making it a context-specific effect. As discussed previously, the function $\phi(X\beta)$ links the linear index to the CDF, which ensures that the probability of y is bounded between 0 and 1. Together, these components explain the complex and nonlinear relationship between x_1 and y.

The results from the probit model specification including these polynomials and interactions are included in Table 8.

Table 8: Probit maximum likelihood estimation including quadratics and interaction terms

Variables	***Complete bachelor's degree (probit)***		***Attend college (probit)***	
	Coeffic. estimates	Average partial effects	Coeffic. estimates	Average partial effects
Socioeconomic status composite	0.6478	0.0762	0.1592	0.0734
	(0.1687)	(0.0081)	(0.1435)	(0.0084)
Socioeconomic status quadratic	0.1191		-0.0711	
	(0.0528)		(0.0501)	
High school GPA	0.4344	0.1345	0.3143	0.0796
	(0.0729)	(0.0025)	(0.0373)	(0.0022)
High school GPA quadratic	0.0149		-0.0017	
	(0.0085)		(0.0052)	
SES x HS GPA interaction	-0.0516		-0.0414	
	(0.0170)		(0.0139)	
School quality index	0.0170	0.0688	0.1681	0.0423
	(0.0255)	(0.0062)	(0.0263)	(0.0064)
School quality index quadratic	0.0745		0.0289	
	(0.0296)		(0.0255)	
How much school parents want 10th grade student to complete	0.1288	0.0310	0.1251	0.0324
	(0.0124)	(0.0029)	(0.0105)	(0.0027)
Parental education	0.0787	0.0068	0.2208	0.0079
	(0.0640)	(0.0029)	(0.0577)	(0.0028)
Parental education quadratic	-0.0045		-0.0224	
	(0.0069)		(0.0063)	
SES x parental education	-0.0298		0.0613	
	(0.0310)		(0.0278)	
Observations	13,934	13,934	13,934	13,934
Controls incl.	Yes	Yes	Yes	Yes
Quadratics and interactions incl.	Yes	Yes	Yes	Yes

This table presents the results of the probit estimation when the model includes quadratic terms for socioeconomic status, HS GPA, school quality index, and parental education. Additionally, interaction terms between socioeconomic status and both GPA and parental education are included. The partial effects of the interaction terms and quadratics are not presented separately, as their effects are reflected in the overall effect of SES and the other variables.

The probit coefficient estimates are provided alongside the average partial effects. The sign of each coefficient tells us whether the effect is increasing or diminishing, but the nominal values of the coefficients cannot be interpreted without looking at the marginal effects. The estimated average marginal effect of socioeconomic status on bachelor's degree completion in the probit model is now a 7.6 pp increase in probability (95% CI [.0604, .0921]), with heterogeneity across the range of the continuous socioeconomic status variable. The effect on bachelor's degree attendance is an increase of 7.3 pp (95% CI [.0569, .0900]). These results are close to the original probit estimates without polynomials or interaction terms. For both binary outcomes, the standard errors in the probit model are small, producing a tighter confidence interval and more efficient estimates than the linear probability model. Given the nonlinearity of the probit model, the heterogeneous effects of socioeconomic status are best displayed graphically. Figures 1 and 2 show the plot of the average marginal effects of socioeconomic status on the probability of the outcome variables across the explanatory variable's range.

Figure 1: partial effects of SES on pr(compl. coll = 1)

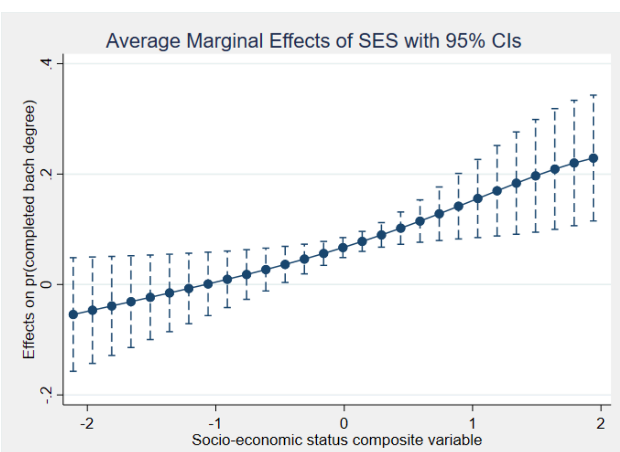
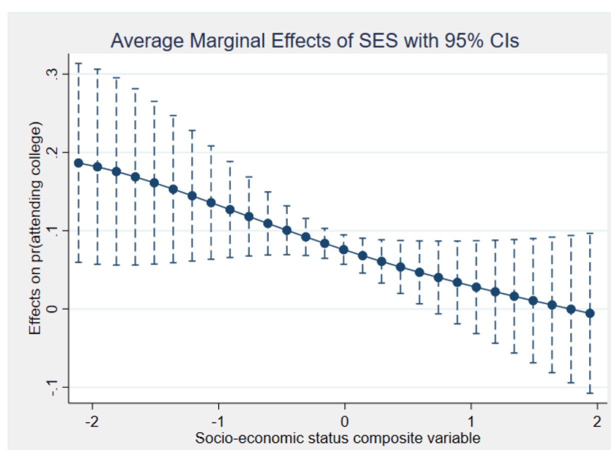


Figure 2: partial effects of SES on pr(attend coll = 1)



Source of data: Education Longitudinal Study of 2002

These figures highlight the nonlinearity of the effects of SES on the educational outcomes. The confidence intervals for each partial effect are displayed by the dashed lines. For completing college, the effect of SES increases at higher values. For attending college, the effect of SES decreases with higher values.

For degree completion, the effect of SES increases along the range of the variable, although the confidence intervals for the lower values of SES include 0 and lack statistical significance. The strength of the marginal impact of SES on college attendance diminishes at higher levels, with confidence intervals crossing 0 toward the right tail of the margin plot. The implications of these heterogeneous effects are considered in greater detail in the discussion in section 5.

In addition to the within-group variation across the range of the explanatory variable, there could be heterogeneous effects across different subgroups within our cohort. To test for this, the probit models are run again, this time specifying the average partial effects to be calculated by group. We first test for differences across gender, and plot the average marginal effects in Figures 3 and 4.

Figure 3: partial effects of SES on pr(compl. coll = 1)

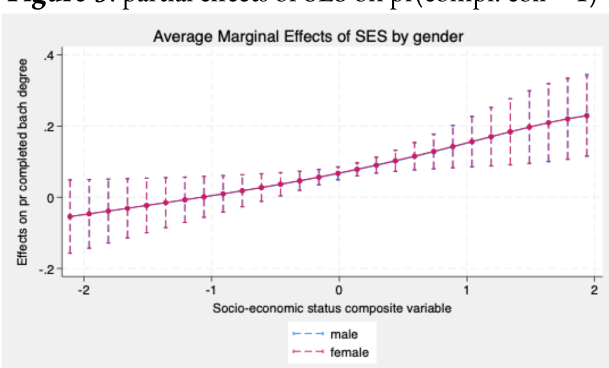
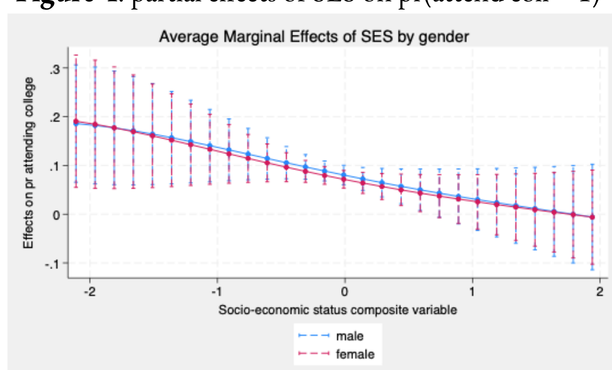


Figure 4: partial effects of SES on pr(attend coll = 1)



Source of data: Education Longitudinal Study of 2002
These figures overlay the effects of SES on the educational outcomes for the male and female subgroups. The distribution of the effects are nearly identical, particularly for the bachelor's degree completion outcome variable.

In Figure 3, the effect of SES on bachelor's degree completion is nearly identical, with overlapping effects for the male and female groups. Figure 4 displays largely the same result. Figures A.1 and A.2 in the appendix display the additional average partial effects graphs for race, with the black and white subgroup effects displaying visually identical overlap as well. In addition to these visualizations, we can run a test of these nested model specifications to ascertain if there are statistically significant differences in the effects. A likelihood ratio (LR) test (Greene, 2018) can be set up for this purpose. The estimates from the full (unrestricted) model are calculated, followed by the model with the constraint that the coefficient estimates for "male" and "female" are identical. The LR test then tests this null hypothesis. For the gender-restricted model, the associated chi-squared statistic is 1.97 ($p = 0.1599$), indicating that we have insufficient evidence to reject the null hypothesis of identical effects for the male and female subgroups. For the LR test for the black and white subgroups, the associated chi-squared is 0.71 ($p = 0.3997$), so we cannot reject the null hypothesis for this group either. Contextualizing these results in terms of our research question, we see that although there is heterogeneity in the effects of socioeconomic status across the range of the variable, there is not statistically significant heterogeneity in the effect of SES across gender or race. These results may warrant further analysis of other cohorts.

We have now introduced multiple specifications of our model and provided their resulting estimates, as well as evaluated heterogeneity within the total sample across SES and across groups divided by gender and race. In order to evaluate which model specification is appropriate for our analysis, we can run functional form misspecification tests. Following Ramsey (1969), we use the Regression Equation Specification Error Test (RESET), specifically adapted for the probit model by Papke and Wooldridge (1996). This test estimates the linear prediction $X\hat{\beta}$ from the probit model, then adds higher-order iterations of $X\hat{\beta}$ and regresses the model again including these terms. Finally, the joint statistical significance of those higher-order predictions is tested. If the result is statistically significant, the model may be misspecified and require changes for a better fit.

When running this test for the original probit model, including only linear terms and no interactions, the joint test is highly statistically significant with $p < .0001$, confirming the need for adjustments to this model.

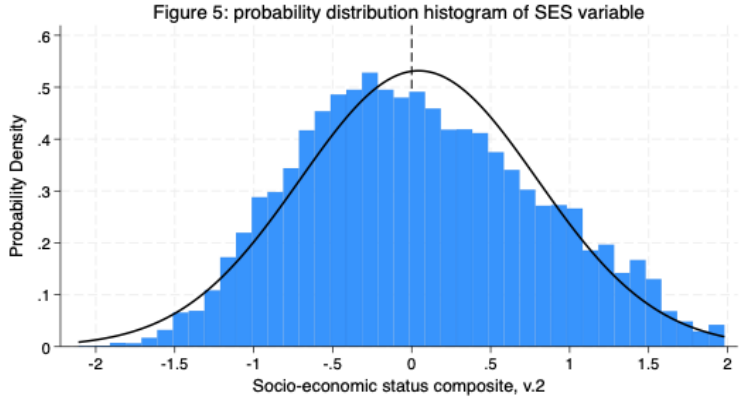
However, we repeatedly run this test while adding the previously referenced quadratic terms and interactions, and continue to produce significant test statistics. This could indicate that there is still room for improvement of the model's fit, and additional nonlinear terms not considered here may be important. Additionally, since the probit maximum likelihood estimation model assumes a normal distribution, this test's significance could be due to non-normality of the latent model's error term. Traditional OLS procedures for verifying distributional normality, including visual tests such as histograms and Q-Q plots, are not possible since the predicted $X\hat{\beta}$ assumed normality in making the predictions. These potential misspecifications could be further evaluated in future research.

5. Discussion

Our study found statistically significant effects of socioeconomic status on both the probability of attending and completing college that were robust to multiple controls and specifications. It is helpful to put the numbers in context to interpret their economic significance. The effect was estimated to be an average increase of 7.6 percentage points (pp) on the likelihood of completing college for a 1 unit increase in socioeconomic status, and an increase of 7.3 pp for college attendance. Socioeconomic status is a complex and multi-faceted variable, but controlling for academic ability, parental education, school quality, and other key independent variables discussed in section 2 allows us to hone in on specific mechanisms through which SES affects tertiary educational outcomes. Specifically, our estimated effects better isolate the direct influence of financial resources, affecting the student's ability to afford tuition and other costs without taking on excessive debt, as well as access to social and cultural capital, including networks and knowledge that facilitate understanding of the college landscape.

The SES composite variable ranges from -2.11 to 1.98, and distribution of the SES variable is approximately normal. The distribution of the socioeconomic status variable is shown in Figure 5.

Figure 5: probability distribution histogram of SES variable



The distribution of the SES independent variable is approximately normal. The normal distribution is overlaid here.

Since SES is close to normally distributed, the majority of students are clustered toward the mean and have the greatest influence on the average partial effects, although the heterogeneity is more pronounced in the tails of the distribution.

To add numerical context to these average effects, the mean attendance and completion outcomes by SES quartile are shown in Table 9 below, and in the appendix in Figure A.3. For example, the average number of students in the lowest quartile that complete a bachelor's degree is 13.8%, compared with 57.5% for the highest quartile. Additionally, the differential between attendance and completion is lowest in the highest SES strata, indicating that persistence to degree attainment for college attendees is lower among the lower SES groups.

Table 9: Mean college attendance and completion by SES quartile

Source of data: Education Longitudinal Study of 2002

Variable	***Observations***	***Mean***	***Std. Dev.***
First Quartile			
Attended college	3,600	0.53861	0.49858
Completed bachelor's degree	3,600	0.13806	0.34501
Second Quartile			
Attended college	3,590	0.67354	0.46898
Completed bachelor's degree	3,590	0.21838	0.41321
Third Quartile			
Attended college	3,753	0.77218	0.41948
Completed bachelor's degree	3,753	0.35412	0.47831
Fourth Quartile			
Attended college	4,301	0.88166	0.32305
Completed bachelor's degree	4,301	0.57545	0.49433

This table provides the simple averages for the attendance and completion outcomes by SES quartile. These results are useful for contextualizing the results from our regression analysis, but are not intended to be interpreted as effects of SES.

Given the controls in our model, the 7.6 pp average partial effect would mean that, if the effects exhibited linearity, for two students at the 25th and 75th SES percentiles, even if they are the same across a number of variables, the student at the 75th percentile would have a 15 pp greater chance of both attending and completing college.

However, we find non-linear, heterogenous effects of SES, as seen in Figures 1 and 2. The strength of the marginal effect of an increase in SES exhibits nonlinearity. For college completion, the effect grows with higher SES. For changes in SES at the bottom of the range, the average effect is small and we fail to reject the null hypothesis that the effect could be zero. At higher values of SES, the effect of a unit change is substantial, approaching an increase of 20 pp. For college attendance, the variability in the marginal effect moves in the opposite direction: at low SES, the magnitude is high, and approaches 0 at high SES. This provides an interesting glimpse into college access. The average marginal effect curve for college completion is possibly upward-sloping due to the reduced financial constraints that a student faces with greater family resources, as well as the safety net and potential guidance through the college experience, while at low SES, even additional resources are not enough to overcome these hurdles. The diminishing marginal effect seen on attendance could be due to the fact that at high levels of SES, the vast majority of students attend college, so further increase at that end of the spectrum has little effect. Taking the two results together, a potentially discouraging picture of college access and outcomes is painted, as students from low SES backgrounds experience a significant increase in the likelihood of attendance with changes to SES, but do not see the same effects on completion. This could be due to financial burdens that cause dropout or a lack of preparedness and support for college.

Now, we return to the motivation for our study, and shed light on some reasons why this difference in access is important. Although it is outside the scope of this study, the effect of education on earnings is significant, and holds for marginal students as well. In a robust review of the economic literature, Oreopoulos and Petronijevic (2013) found that across most studies, the causal effect on earnings of attending college was 7-15 percent per year in school. A recent study leverages admission requirement cutoffs, including SAT and ACT scores, and applies a regression discontinuity design to analyze marginal students admitted to universities in Texas. Although these students were found to be of lower academic ability and had less family advantages than the average college student, the effect of completing a degree on their own earnings (the internal rate of return) was 19-23 percent 8 years after their time of application (Mountjoy, 2024). Moreover, Mountjoy finds that the additional tax revenue generated from their earnings premiums surpasses the costs to taxpayers (as the study focuses on public universities) after 25 years for these marginal students. This last component is essential when considering implications for policy. College educated adults, on average, pay more in taxes and disproportionately fund social insurance programs such as Medicare and Social Security, while being less likely to incur costs by drawing from these programs or being incarcerated (Carroll & Erkut, 2009).

From the body of research, it is clear that substantial benefits are realized by both individuals and society as young people attend and complete college, even when the students are marginal. Programs designed to increase rates of college attendance can maximize social welfare as well as increase utility for the marginal student. Our study finds that students from wealthier families attend and complete college at substantially higher rates. Grants aimed at closing the financial gap for these students, including Federal Pell Grants and Federal Supplemental Education Opportunity Grants administered by the Department of Education, have been highly effective at improving enrollment and outcomes among lower-resourced students (Denning et al, 2019). Even for these students who rely on greater financial aid assistance, the increase in their estimated earnings offsets government expenditures within 10 years (Denning et al, 2019). Policymakers could explore the merits of expansion to these programs, especially as the share of college costs covered by the grants diminishes over time. Senate bill S.4595, The Pell Grant Preservation and Expansion Act of 2024, was recently proposed with the aim of maintaining the purchasing power of the grant funds for students (S.4595, 2024). Increased investment in the community college to four-year degree pipeline could additionally help students with lower resources. To the extent that students from lower SES backgrounds have worse higher education outcomes due to lack of knowledge, support, or other non-financial constraints, targeted approaches could prove effective. For high school students, empirical research has shown that assistance with FAFSA and other aspects of the application process can help increase attendance among these students (Bettinger et al., 2009). For students enrolled in college, interventions including advisory support, mental health resources, and work-study opportunities could additionally prevent lower SES students from slipping through the cracks.

6. Conclusion

The research is clear: higher education provides substantial economic benefits for students, and a pathway out of poverty for those from lower socioeconomic backgrounds. Ensuring that these students have access to this societal ladder is an important consideration for policymakers and institutions seeking to keep social mobility alive in the United States. Our study contributes to the body of literature by using probit regression to show substantial differences in college attendance and completion across the socioeconomic spectrum. The results are statistically significant, and control for various factors, including prior academic ability, parental education levels and expectations for their children, and high school quality. The regression models allow for heteroskedasticity and nonlinearity, capturing the complex relationship between SES and college outcomes and the heterogeneity among students.

Limiting our ability to interpret these results as fully causal is the possibility of endogeneity. If any of the explanatory variables fail the assumption of exogeneity and the error term captures critical variables for our analysis, we could have confounders that affect our results. An identification strategy utilizing quasi-experimental design (since socioeconomic status cannot be randomly assigned) such as instrumental variables could address this potential endogeneity issue further, and is a recommended area for further research. An instrumental variable z should have a strong correlation with the treatment x variable (instrument relevance) and be uncorrelated with the error term (instrument exogeneity), so it only has an indirect effect on the outcome. If an instrumental variable can be found that satisfies these conditions, the regression estimates can better be interpreted as causal as the endogeneity problem is addressed.

Appendix

Figure A.1: partial effects of SES on pr(compl. coll = 1)

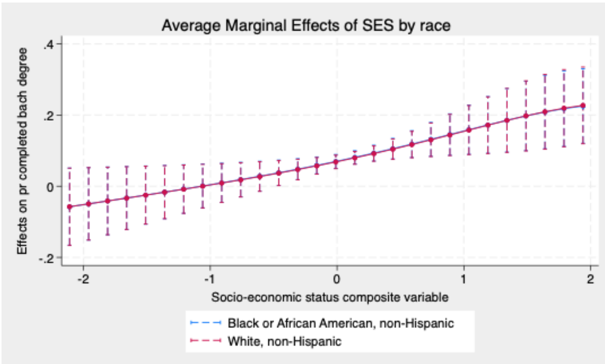
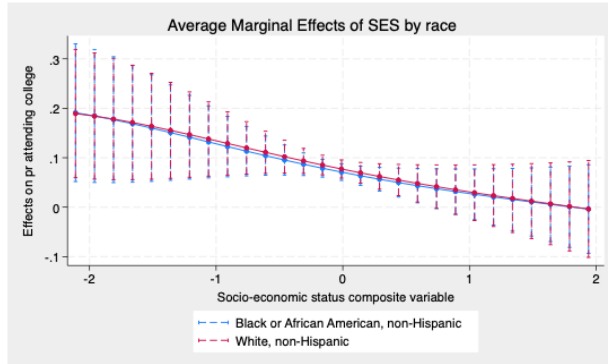
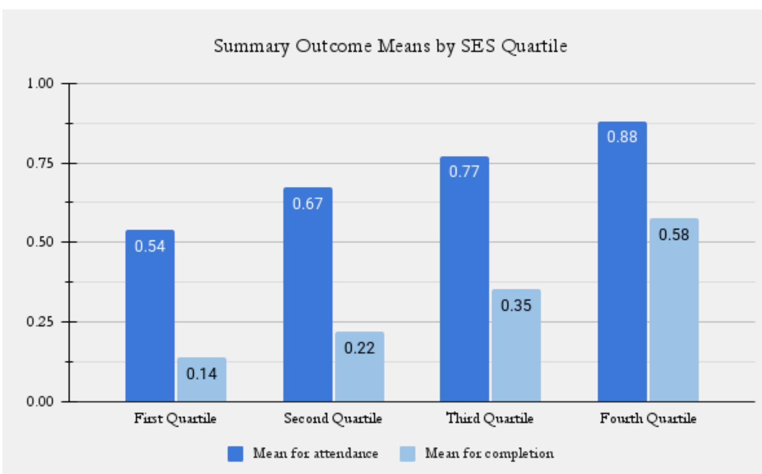


Figure A.2: partial effects of SES on pr(attend coll = 1)



Source of data: Education Longitudinal Study of 2002
These figures overlay the effects of SES on the educational outcomes for the Black or African American (non-Hispanic) and White (non-Hispanic) subgroups. The distribution of the effects are once again nearly identical, particularly for the bachelor's degree completion outcome variable.

Figure A.3: Mean college attendance and completion by Socioeconomic status quartile



This bar graph provides a simple descriptive view of the average levels of college attendance and completion of students from each SES quartile. Higher SES status is associated with higher rates of college attendance and completion. This graph does not provide any causal insights or drivers of college outcomes for higher SES students. *Source of data: Education Longitudinal Study of 2002.*

References

Ai, C., & Norton, E. (2003). Interaction terms in logit and probit models. *Economics Letters*, 80(1), 123-129. https://doi.org/https://doi.org/10.1016/S0165-1765(03)00032-6

Averett, S., & Burton, M. (1996). College attendance and the college wage premium: differences by gender. *Economics of Education Review*, *15*(1), 37-49. https://doi.org/https://doi.org/10.1016/0272-7757(95)00027-5

Bettinger, E., Long, B., Oreopoulos, P., & Sanbonmatsu, L. (2009). The role of simplification and information in college decisions: results from the H&R Block FAFSA experiment. National Bureau of Economic Research. https://www.nber.org/system/files/working_papers/w15361/w15361.pdf

Bjorklund-Young, A. (2016). Family income and the college completion gap. Institute for Education Policy, Johns Hopkins University.

Blanden, J., Doepke, M., & Stuhler, J. (2022). Educational inequality. National Bureau of Economic Research Working Paper Series, No. 29979. https://doi.org/10.3386/w29979

Bowen, W., Chingos, Matthew., McPherson, M. (2009). *Crossing the Finish Line: Completing College at America's Public Universities*. Princeton University Press, 2009.

Carneiro, P., & Heckman, J. (2002). The evidence on credit constraints in post-secondary schooling. The Economic Journal, 112(482), 705-734.

Carroll, S., & Erkut, E. (2009). How taxpayers benefit when students attain higher levels of education. RAND Corporation. https://doi.org/10.7249/RB9461

Chetty, R. Friedman, J., Saez, E., Turner, N., & Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the United States. *The Quarterly Journal of Economics*, Volume 135, Issue 3, August 2020, Pages 1567–1633, https://doi.org/10.1093/qje/qjaa005

Denning, J., Marx, B., & Turner, L. (2019). ProPelled: the effects of grants on graduation, earnings, and welfare. *American Economic Journal: Applied Economics*, 11 (3): 193–224.

DesJardins, S., Ahlburg, D., & McCall, B. (2006). The effects of interrupted enrollment on graduation from college: racial, income, and ability differences. *Economics of Education Review*, 25(6), 575-590.

Duncan, O. (1961). A social economic index for all occupations. Free Press. In A.J. Reiss Jr. (Ed.), Occupations and Social Status, Free Press.

Dynarski, S. (2015, June 2). For the poor, the graduation gap is even wider than the enrollment gap. *The New York Times*. https://www.nytimes.com/2015/06/02/upshot/for-the-poor-the-graduation-gap-is-even-wider-than-the-enrollment-gap.html

Greene, W. (2010). Testing hypotheses about interaction terms in nonlinear models. *Economics Letters*, 107(2), 291-296. https://doi.org/https://doi.org/10.1016/j.econlet.2010.02.014

Greene, W. (2018). Econometric Analysis. 8th ed. New York: Pearson.

Gutman, L., & Schoon, I. (2013). The impact of non-cognitive skills on outcomes for young people: a literature review. Education Endowment Foundation, University of London. Retrieved from https://discovery.ucl.ac.uk/id/eprint/10125763/.

Jacob, B. (2002). Where the boys aren't: non-cognitive skills, returns to school and the gender gap in higher education. *Economics of Education Review*, 21(6), 589-598. https://doi.org/https://doi.org/10.1016/S0272-7757(01)00051-6

Kochhar, R. (2024). The state of the American middle class: Who is in it and key trends from 1970 to 2023. Pew Research Center.

Mountjoy, J. (2024). Marginal returns to public universities. National Bureau of Economic Research Working Paper Series, No. 32296. https://doi.org/10.3386/w32296

National Center for Education Statistics, United States Department of Education (2005). *Education Longitudinal Study (ELS), 2002*. https://doi.org/10.3886/ICPSR04275.v1

National Center for Education, United States Department of Education (2021). *Digest of Education Statistics: 2021*. https://nces.ed.gov/programs/digest/d15/tables/dt15_104.10.asp

Oreopoulos, P., & Petronijevic, U. (2013). Making college worth it: a review of the returns to higher education. The Future of Children, 23. https://doi.org/10.2307/23409488

Papke, L., & Wooldridge, J. (1996). Econometric methods for fractional response variables with an application to 401 (K) plan participation rates. *Journal of Applied Econometrics*, 11(6), 619-632.

Pell Grant Preservation and Expansion Act of 2024, S.4595, 118th Congress (2024). https://www.congress.gov/bill/118th-congress/senate-bill/4595/text

Ramsey, J. (1969). Tests for Specification Errors in Classical Linear Least-Squares Regression Analysis. Journal of the Royal Statistical Society. Series B (Methodological), 31(2), 350-371.

Saez, E., & Zucman, G. (2016). Wealth inequality in the United States since 1913: Evidence from capitalized income tax data . *The Quarterly Journal of Economics*, 131(2), 519-578. https://doi.org/10.1093/qje/qjw004

Sirin, S. (2005). Socioeconomic status and academic achievement: a meta-analytic review of research. *Review of Educational Research*, 75(3), 417-453.

State Higher Education Executive Officers Association (SHEEO) (2024). State higher education finance: FY 2023. SHEEO. https://shef.sheeo.org/wp-content/uploads/2024/04/SHEEO_SHEF_FY23_Report.pdf

White, H. (1980). A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, *48*(4), 817–838. https://doi.org/10.2307/1912934