College Costs and the Marginal College Graduate

Jonathan Gemus*

June 9, 2010

Abstract

This paper examines the distributional impacts of direct university costs. The primary obstacle in estimating these effects is the endogeneity of schooling costs. To overcome this issue, I use two measures of direct costs that are plausibly exogenous: living close to a university and the elimination of the Social Security Student Benefit Program in the United States. Both sources of cost variation provide results that are very consistent with each other. In most ways, lowering the cost of schooling does not significantly change the pool of university enrollees or graduates. Enrollees and graduates facing higher costs are just as likely to be female, non-white, and have college educated parents as enrollees and graduates facing lower costs. However, lowering the cost of university leads to a marginal graduate who has lower cognitive test scores and, surprisingly, higher family income, while marginal enrollees look similar along these dimensions as well.

*Department of Economics, Uppsala University and Uppsala Center for Labor Studies.
Email: jonathan.gemus@nek.uu.se. I would like to thank Per-Anders Edin, Chris Taber, Sergio Urzúa and seminar participants at Uppsala University and the NoCE Workshop in Bornholm for very helpful comments and suggestions. I would also like to thank Jim Rosenbaum for help accessing High School & Beyond data. All errors are my own.

1 Introduction

One of the most enduring relationships in labor economics is the link between a college education and earnings. Earning a college degree is widely viewed to be the first step in the path to establishing a successful career and a high standard of living. Students, however, vary in the ease with which they can finance a college education. Policy makers and economists have therefore been concerned that, regardless of family background, all qualified students be able to go to college. This concern has led to government aid in the form of grants and low-interest rate loans. Additionally, colleges have considerable leeway in setting the amount of tuition paid by students, and in practice universities have used this ability to set prices to attempt to mitigate the financial burden faced by students who do not have the means to pay for college out of pocket.

The importance of financial aid and tuition policies has given rise to a literature in economics that studies the impact of financial aid on college attainment and a related literature studying the distributional effects of financial aid policies – that is, whether different groups of students respond more or less strongly to financial aid. In this paper, I add to the literature studying the distributional impact of changes in college costs by asking whether changing college costs affects the share of low ability, female, and minority college graduates, as well as the share of students from low-income and less-educated families. While other papers have studied the distributional impacts of aid with respect to family background characteristics, this paper is (to the best of my knowledge) the first to provide evidence with respect to cognitive ability.

This extension is relevant for at least two reasons. First, to the extent that the return to college is driven by the ability distribution of college students (e.g. in signaling models or if peer effects are important), any change in the costs of college could potentially alter the returns to schooling, and it is important for policy makers to have an accurate depiction of any benefits and costs of financial aid policies. Second, ability sorting into different levels of schooling has been one of the primary concerns for economists attempting to estimate the causal effect of education on earnings. This paper provides direct evidence on the extent to which direct schooling costs drive sorting into college based on cognitive ability.

The primary obstacle in estimating the distributional effect of direct schooling costs is that students ultimately have considerable leeway over how much they spend on their education. For example, a student can attend a public university in their home state and pay a relatively low tuition level or attend an out-of-state public university or a private university. These decisions will likely depend on expected returns to education, tastes for education, unobservable costs (both monetary and non-monetary), and so on. Thus it is generally not possible to assign a causal interpretation to the relationship between tuition and college student characteristics.

To overcome this issue, I consider two measures of direct college costs that are plausibly exogenous. First, using the High School & Beyond Sophomore cohort, I proxy direct college costs using distance to nearest four-year university. Versions of this measure of college costs have been widely used.[1] The idea of this measure of direct costs is that it is cheaper to live at home than in a separate residence from one's parents; thus students whose parents live within driving distance of a university have access to cheaper education than students living out of driving range because they have the opportunity to reduce their housing costs. I divide students into two categories: within driving distance and not within driving distance of a university.[2] Regardless of how "within driving distance" is defined, respondents who live close enough to a university to drive to school have pre-college cognitive test scores that are slightly higher than the test scores of students who do not live close enough to commute. However, *among college degree holders*, I find that students living within driving distance of a university have significantly lower pre-college cognitive test scores than students living out of driving range (between 0.11 and 0.18 standard deviations lower depending on the specification), though this difference is concentrated among men. Additionally, college graduates with access to a local college have higher family income than students without access to a local college. Along other dimensions such as parent education, race, and gender, college graduates who live near a college look otherwise similar to college graduates who do not live near a college.

The second measure of direct college costs exploits the elimination of the Social Security Student Benefit Program (SSSBP), a large financial aid program that provided a sizable college grant for

[1] See, for example, Card (1993), Bedard (2001), and Cameron and Taber (2004).

[2] Typically, authors have defined "living near a college" as having a college in the same county as which the respondent resides. High School & Beyond does not contain this measure, but it does provide a variable indicating the distance from the respondent's high school to the nearest 2-year and nearest 4-year college.

children of social security eligible parents who were retired, disabled, or deceased. This policy change was first used in Dynarski (2003) to estimate the effect of aid on college attendance. From 1965 through 1981, the Social Security Administration provided a very generous tuition subsidy to 18-22 year old children of deceased, disabled, or retired Social Security eligible parents. This program was so generous that the average subsidy more than covered average public school tuition and fees, and it almost completely covered average private school tuition. Furthermore, the amount of the benefit was not tied to realized schooling costs but to parents' Social Security benefit levels. After the elimination of this program, there was a drastic drop in tuition aid for individuals who were formerly eligible for this program (Dynarski (2003)). I find that aid eligible college graduates had AFQT scores that were between 0.21 and 0.36 standard deviations below the AFQT scores of aid ineligible college graduates. Once again, I find that aid eligible college graduates tended to come from relatively high income families. Otherwise, aid eligible and ineligible college graduates look very similar with respect to race, gender, and parent education. Finally, I find that marginal aid eligible college enrollees look almost identical to marginal aid ineligible college enrollees.

This evidence adds to a growing literature on the effects of financial aid. One strand of this literature has studied the impact of direct schooling costs on college attendance. In attempting to explain trends in college attendance among African-Americans, Kane (1994) finds that a $1,000 decrease in tuition increases college enrollment by approximately 4%. Dynarski (2003) finds results that are remarkably similar using the elimination of the Social Security Student Benefit Program: She finds that a $1,000 decrease in schooling costs increases enrollment by approximately 4% as well. Using eligibility for the G.I. Bill, Bound and Turner (2002) also find that reducing the cost of schooling increases college attendance.

A smaller group of studies has examined the distributional impacts of direct schooling costs. Kane (1994) finds that blacks respond to changes in costs more than higher income whites but similarly to low-income whites. Turner and Bound (2003) find that, because of segregated colleges, blacks in the South increased college attendance less than whites in response to the GI Bill, though blacks in the North responded similarly to whites. Dynarski (2000) finds that the HOPE scholarship in Georgia had the largest impact on college enrollment for middle-class whites, and

Stanley (2003) finds that veterans from more-educated families responded the most to the Korean War GI BIll. Finally, Cameron and Taber (2004) find little evidence that enrollment rates respond to the presence of a local college differently by family income, minority status, or parent education.

Most recently, Oppedisano (2008) (written simultaneously to this paper) uses the introduction of new colleges and expansions of existing universities in the 1990s in Italy to examine, among other things, whether this educational expansion increased the share of students with lower high school grades in college. She finds that the introduction of local colleges increased the share of students with middle-range grades attending university. I believe that the present study offers a contribution that is distinct from Oppedisano (2008) for at least three reasons. Most importantly, a large higher education expansion introducing new colleges not only reduces the costs of schooling for students nearby but also changes the types of schools available to students. It is not clear whether these schools draw different types of students because they differ from existing schools or because they reduce the costs of schooling for students living within commuting distance. It is therefore difficult to interpret the results in Oppedisano (2008) as being solely driven by direct schooling costs.[3] Second, the Italian university system is considerably different from the American university system, and it is therefore relevant to produce evidence from both of these countries. Finally, I focus on cognitive test scores that are comparable across students in different schools rather than high school grades.

The paper proceeds as follows: Section 2 describes the theoretical framework and the empirical strategy, Section 3 describes the data, Section 4 presents results, and Section 5 concludes.

[3]While this paper also uses the presence of a local college as a measure of costs, these colleges were generally well-established and not systematically part of a concerted effort to expand education in the U.S.. While there is some possibility that "local colleges" are very different from schools that people generally do not live near, this seems unlikely both because a majority of students live within commuting distance of a 4-year college and because casual observation suggests that nearly all population centers in the US have major 4-year colleges.

2 Theoretical Framework and Empirical Strategy

2.1 Theoretical Framework

A simple Roy model suggests that changes in direct college costs should change the characteristics students on the margin of college, though the direction of the effect is ambiguous. Suppose that students choose between two levels of schooling, $s = 1$ or $s = 0$. The payoff to each level of schooling is a function of some set or index of characteristics q that influence both lifetime earnings at each level of schooling as well as tastes for that level of education. If students choose $s = 1$ they incur a direct schooling cost, c, and zero schooling costs otherwise. The payoff to $s = 1$ is therefore

$$V_1(q) = W_1(q) - \rho c - t_1(q) \tag{1}$$

and the payoff to choosing $s = 0$ is

$$V_0(q) = W_0(q) - t_0(q), \tag{2}$$

where $W_s(q)$ is the lifetime earnings associated with skill level q and education level s and $t_s(q)$ are tastes associated with skill level q and education level s. ρ reflects the possibility that different students may be differently affected by college costs because of, for example, borrowing costs or splitting the responsibility of paying for college with parents. I assume that $W_s' > 0, t_s' < 0$.[4] For each individual, the school choice problem is given by

$$s = \mathbf{1}[V_1 \geq V_0]. \tag{3}$$

Without further assumptions, this model does not predict the direction in which a change in c affects the distribution of q for people with a college degree. Let q^* be the q of the people on the

[4] One can easily extend this model so that there are two separate indexes, one that is used if a college degree is earned and another that is used if a college degree is not earned. While the notation becomes more burdensome, no additional intuition is gained.

margin of earning a college degree. Then

$$\begin{aligned} V_1 &= V_0 \\ W_1(q^*) - \rho c - t_1(q^*) &= W_0(q^*) - t_0(q^*). \end{aligned} \tag{4}$$

Differentiating with respect to c yields

$$\frac{\partial q^*}{\partial c} = \rho[(W_1' - W_0') - (t_1' - t_0')]^{-1}. \tag{5}$$

The effect of c on the ability of the marginal student is going to depend on the relative sizes of W_1', W_0' , t_1', and t_0'. In general this is ambiguous, but it is clear that, in general, q^* is a function of c, i.e. $q^* = q^*(c)$. Thus, all else equal an increase or decrease in the costs of schooling will change the characteristics of the population with and without a college degree. The direction of this change is an empirical question.

2.2 Empirical Strategy

I motivate the empirical strategy with a hypothetical experiment in which two otherwise similar groups face different college costs.[5]

Two groups of people, A and B, have achieved schooling level $s = 0$ and are assigned two different tuition levels, c^A and c^B, $c^A > c^B$, that they must pay if they wish to continue on to $s = 1$. Let student characteristics once again be denoted q. I assume that q is realized prior to assignment of tuition levels and is fixed by the time of this experiment for each individual student. Suppose that selection into groups A and B is done randomly so that $E(q|A) = E(q|B) = E(q)$. It follows directly from this that $E(q|c^A) = E(q|c^B) = E(q)$ as well so that tuition is unrelated to student characteristics.

As suggested by the Roy model outlined above, the q distribution of group A college graduates is *not* necessarily the same as the q distribution of group B college graduates. These distributions

[5] Additionally, functional form assumptions can be imposed on the wage and taste functions in the Roy model outlined in the previous subsection, and a similar specification can be derived.

would differ if, for example, the slopes of the wage and taste functions in the above model changed with educational status. Thus, while $E(q|c = c^A) = E(q|c = c^B)$, in general $E(q|c = c^A, s = 1) \neq E(q|c = c^B, s = 1)$, and we can get a measure of the degree of selection by estimating

$$\alpha_1 = E(q|c = c^A, s = 1) - E(q|c = c^B, s = 1),$$

i.e. by comparing the mean q of college graduates facing higher tuition to the mean q of college graduates facing lower tuition levels. One could easily obtain an estimate of α_1 by estimating the following regression:

$$q = \alpha_1 c + \alpha_2 c \times (1 - s) + \alpha_3 (1 - s) + \epsilon, \tag{6}$$

where ϵ is an error term. One can think of α_1 in two related ways. Most obviously, it captures the difference in the average q of college graduates facing different schooling costs. If q is cognitive ability, for example, it estimates the difference in the average cognitive ability of college graduates facing high costs and the average cognitive ability of college graduates facing lower costs and thus estimates the extent to which changing college costs changes the average ability of college students. Second, as the model in Section 2.1 makes clear, only marginal students are affected by changes in schooling costs. Thus α_1 provides an estimate of the extent to which the characteristics of the marginal college student changes in response to a change in the price of college.

Thus while this regression has a causal interpretation, it is not causal in a micro sense. Typically when discussing causal estimation, one would think of α_1 as capturing the effect of c on q at the micro level (i.e., "An x unit increase in c_i increases q_i by y units for person i"). α_1 has no such micro level interpretation. Instead, α_1 has a causal interpretation at the macro level. q is fixed at the micro level, so by increasing c we are changing who chooses to earn a college degree. This in turn shifts the distribution of q for the group of people who choose to earn a college degree. Alternatively, one can think of these regressions as capturing how tuition policy changes who is on the margin of graduating college.[6]

[6] This is not the first paper to study how policy changes shift the characteristics of marginal individuals. For example, the literature on the effect of abortion legalization estimates regressions that are similar in spirit. Gruber *et al* (1999) estimate the effect of abortion legalization on the characteristics of children on the margin of being born, and Donohue

Figure 1 provides a graphical example of this estimator. The distribution of q for the whole population is shown. This distribution does not change over the course of our experiment regardless of the college tuition imposed. I assume that the wage and taste functions are such that the college decision problem satisfies a cutoff property in q. Two cutoffs are displayed that show the q of the marginal student given the two different tuition levels: q^B shows the cutoff for group B, and q^A shows the cutoff for group A.[7] In this simplified example, within each group, all students above the cutoff earn a college degree, and all students to the left of the cutoffs do not go to college. The estimate of the selection effect of college costs captures the mean difference between the area to the right of q^A and the area to the right of q^B.

Regressions similar to Equation 6 are used to estimate the effect of college costs on characteristics of students on the margin of graduating college. This specification is different from the specification typically used in the literature studying the distributional effects of college aid. Previous studies have focused on specifications that estimate heterogeneity in the price elasticity of college:

$$Pr(s = 1|q, c) = Pr(\gamma_1 q + \gamma_2 c + \gamma_3 q \times c \geq \mu|q, c), \tag{7}$$

where the focus is on γ_3.

I focus on estimating Equation 6 rather than Equation 7 primarily because this formulation seems to be closer to the way that policy makers think about the distributional impacts of financial aid. Specifically, when policy makers talk about financial aid, they often do so by specifically referring to hopes that aid will increase the share of particular groups of people enrolling in college.[Citation to be Added] Equation 6 more closely reflects this idea, especially when q is a binary variable like sex or minority status.

A second advantage of the approach used in this paper is that it makes the role of marginal students and heterogeneous price effects more explicit. To see this, consider once again the model in Section 2.1. The effect of c on the marginal student's q is $\rho[W_1'(q^*) - W_0'(q^*) - (t_1'(q^*) - t_0'(q^*))]^{-1}$.

and Levitt (2001) estimate the effect of abortion legalization on crime rates. In both cases, abortion legalization doesn't change the characteristics of the children in the sense that a micro-level treatment would, but it changes the distribution of characteristics observed in the population of living children through selection.

[7] Note that the Roy model does not suggest which cutoff is to the left of the other. Thus it could be that $q^A \geq q^B$ or vice versa, but the intuition of the estimator does not depend on this.

It is clear that this effect is driven by who is on the margin of college and the effect of c on the payoff to $s = 1$ (ρ).

The main challenge in estimating Equation 6 is finding a plausibly exogenous source of variation in college costs. I make use of two sources that yield very similar results: First, I use the presence of a 4-year college within commuting distance of a respondent's high school. Beginning with Card (1993), similar variables have frequently been used as a measure of direct college costs. The motivation for its use stems from the reduction in living expenses that students enjoy if they are able to commute to college while living with their parents, as there is no need to pay for living expenses.

Second, I exploit elimination of the Social Security Student Benefit Program (SSSBP). This policy change was previously used to study the effect of college aid on college attendance and completion in Dynarski (2003). Between 1965 and 1981, 18 to 22 year old children of disabled, deceased, or retired Social Security beneficiaries were eligible to receive monthly payments while enrolled full time in college. In 1981, Congress voted to end this program. The policy change only affected children of disabled, retired, or deceased Social Security beneficiaries, so I compare the difference in characteristics of the affected group of college graduates before and after the policy change to the difference in characteristics of the unaffected group of college graduates before and after the policy change to identify $\tilde{\alpha}_1$.

The SSSBP was a very generous aid program. The average annual payment for students with a deceased parent was \$6,700, which was quite substantial at the time: it was more than enough to cover public school expenses, where average tuition and fees were about \$1,900, and it was nearly enough to cover private school tuition and fees which on average were about \$7,100. At the peak of this program, nearly 12% of college students were receiving benefits from the SSSBP. In 1981, Congress voted to end this policy for all students graduating high school in the spring of 1982 and later. As was documented in Dynarski (2003), the decline in the number of students who were funded by the SSSBP was quite rapid. By the 1984-85 school year, spending on the program had dropped by \$3 billion. This policy thus consituted one of the most rapid and drastic changes in funding for higher education since the GI Bill after World War II.[8]

3 Data

In this paper I make use of two data sets: The sophomore cohort of High School & Beyond (HSB) and the National Longitudinal Survey of Youth 1979 (NLSY). Collected by the U.S. Department of Education, HSB first surveyed two cohorts of students who were sophomores and seniors in high school in 1980. I focus on the Sophomore Cohort, which is followed until 1992. Sampling is conducted by first selecting schools and then sampling students within schools. HSB collected extremely detailed information on students' background and high school and college experiences. It is thus well suited to study the college decision problem. The key variable is distance from the respondent's high school to the nearest 4-year college. Additionally, HSB contains a rich set of background data: family income, parental education, race, gender, number of siblings, distance between a respondent's high school and nearest college, and a cognitive test score at age 16.

I use HSB to estimate the effect of living within commuting distance of a 4-year college on marginal college graduates' characteristics. In order to corroborate this evidence, I turn to the NLSY to estimate the selection effect of college costs using the elimination of the SSSBP. The NLSY is in a number of ways well suited to studying the effects of this policy change. There are five cohorts of respondents who could have been seniors in high school while the survey was being conducted (between 1979 and 1983). Three of these cohorts would have been seniors in high school before the policy change (1979 - 1981) while two would have been seniors in high school after the policy change (1982 and 1983). It is possible to observe the year in which students attended high school as seniors and their educational attainment by age 24. Additionally, NLSY includes a rich set of background characteristics similar to HSB: family income, parental education, race, gender, number of siblings, and a cognitive test score. The cognitive test score is the Armed Forces Qualification Test (AFQT), which consists of the Arithmetic Reasoning, Mathematics Knowledge, Paragraph Comprehension, and Word Knowledge portions of the Armed Services Vocational Aptitude Battery (ASVAB).[9]

[8]The figures in this paragraph were taken from Dynarski (2003), pages 280-281.

[9]This test is used by the U.S. Military to determine whether military applicants are qualified to enlist. Since the

Unfortunately, the NLSY does not indicate whether college students received aid from the SSSBP. Instead, I follow Dynarski (2003) and proxy aid eligibility by an interaction between being a senior before the policy change and having a deceased father. This approach is taken for three reasons. First, at this point in time there were relatively few women who were eligible for Social Security benefits, so having a deceased, retired, or disabled mother is less likely to leave a student eligible for this aid program. Second, while the aid policy may affect a parent's decision to retire or file for disability, it is probably less likely to affect a parent's "decision" to die. Third, from a practical standpoint, it is not easy to identify retired or disabled parents in the NLSY.[10] If anything, to the extent that respondents who are aid eligible are classified as aid ineligible, this should bias estimates of the selection effect towards zero.

Table 1 displays sample statistics for HSB (Column 1) and NLSY (Column 2). In general, the respondents from the two data sets appear similar along most observable dimensions. In HSB, average family income is $28,743 while it is $28,737 in NLSY,[11] slightly under half the sample of HSB is male and slight over half is male in NLSY, about 34% of both the NLSY and HSB samples had fathers who had at least attended some college,[12], while 27% of HSB mothers and 23.4% of NLSY mothers attended college. Respondents from both samples came from sibships that had on average about three children (including the respondent). The one fairly large difference between the samples is that about 23% of the NLSY earned a bachelor's degree by age 24 while only 15% earned a bachelor's by the same age from HSB. 88.9% of HSB respondents attended a high school that was within 50 miles of a 4-year college, while only 2% of respondents from NLSY are considered to be eligible for aid (as proxied by the $Before \times Deceased\ Father$ interaction).[13]

NLSY is a nationally representative sample, the military used the respondents to the NLSY to normalize the test to a scale of 0 to 100.

[10]It is possible to roughly proxy having a retired father by fathers who are over 65 and not working. When this group is added to the "aid eligible" group, the results do not substantially change.

[11]In HSB, the family income variable provides ranges of incomes rather than the actual amount. I take the midpoint of each interval as the income level. NLSY provides the actual family income. Family income is in 1982 dollars in both the NLSY and HSB and is measured in the senior year of all respondents.

[12]Parental education in HSB is measured as a categorical variable, while in the NLSY it is measured as years of schooling. Any father or mother who at least attended "less than two years of college" in HSB is considered to have at least attended some college. In the NLSY, any parent listed as having 13 years or more of education is considered to have attended some college.

[13]More students were in reality eligible for aid since I categorize those who had retired or disabled fathers and social security eligible mothers as aid ineligible. At the peak of the SSSBP, about 12% of college enrollees were receiving aid

Although neither source of cost variation has similar numbers of students facing expensive and less-expensive schooling, one attractive feature of the combination of these two sources of variation is that one source has a large fraction of students who can access relatively cheap schooling while the other source has a small minority that is eligible for cheaper schooling. It is therefore reassuring that the results from each source are generally in accordance. The final two rows of Table 1 show that about 70% of the NLSY sample were seniors before the elimination of the SSSBP, and about 4% had deceased fathers. Details about the construction of the data are contained in the Data Appendix.

4 Results

4.1 Local Colleges

As discussed above, the main difficulty in estimating the distributional effects of college costs is that students can decide how much they would like to pay for college. The first measure of college costs overcomes this issue by using geographical variation in college locations: People whose parents live closer to universities have the option of living at home and commuting to school, thus reducing the amount spent on housing costs. In High School & Beyond, the measure of "local college" takes the form of distance from the respondents high school to the nearest four-year college.[14]

In order to capture the idea that someone who can commute to school has access to a cheaper college education, I create a dummy variable called *Local College* that is equal to 1 if a respondent lives within commuting distance of a 4-year university. For the main analysis, I consider anyone living within 50 miles ($\approx$ 80 kilometers) of a university to live within commuting distance. Since this is somewhat arbitrary, and in all likelihood the definition of a commutable distance varies by person, I consider cutoffs at 40, 60, and 70 miles in the Appendix. The results do not substantially change with these different definitions.

---

under this program. See Dynarski (2003) for details.

[14] Unfortunately, in High School & Beyond one cannot distinguish between 4-year private and 4-year public universities. To the extent that living near a private but not a public university does not provide access to cheaper schooling, some people classified as living near a 4-year university will in fact face more expensive schooling. This should bias estimates towards zero, so any effect that is statistically different from zero should be viewed as a lower bound (in absolute value).

Table 2 presents evidence on the correlation between having access to a local college and a number of background characteristics for all students regardless of educational attainment. In general, there are not substantial correlations between local college and these characteristics. In Column 1, a $10,000 increase in family income is associated with a 0.6% higher likelihood of living within 50 miles of a 4-year college. While statistically significant from zero at the 5% level, this number is extremely small. Similarly, a standard deviation increase in cognitive test scores is associated with a 0.8% higher likelihood of having access to a local college, and this estimate is once again statistically significant at the 5% level but very small. Having a mother that attended college appears to be unrelated to the local college variable, but having a father with at least some college increases the likelihood of living within 50 miles of a 4-year college by 1.7%. This relationship is slightly larger, but it goes away once family income is conditioned on in Column 7. Race and gender are unrelated to local college, and the relationship between number of siblings and local college is significantly different from zero but once again very small. In the final column, local college is regressed on all variables, and once again all of the estimated coefficients are very small. Only the family income and number of siblings coefficients are significantly different from zero at standard levels.

Table 2 shows that a number of key observable characteristics are not strongly correlated with college costs in the whole sample. The question now arises whether direct schooling costs affect the characteristics of students who choose to graduate from college. Table 3 presents results from estimating Equation 6, where c is proxied by a dummy for local college, q is several different observable characteristics (family income, non-white, number of siblings, having college educated parents, and a cognitive test score), and $1 - s = 1$ if the respondent did not earn a bachelor's degree by age 24 and zero if the respondent did earn one by 24. The coefficient of interest is on *Local College*. In most cases, college graduates who attended high school near a 4-year college look very similar to college graduates who did not: The differences along race, gender, number of siblings, and mother's education are small and statistically insignificant. However, college graduates with access to local colleges came from families that earned on average $6,840 more than college graduates without local colleges. After conditioning on background characteristics (not including family income), this difference drops to $5,970 and is still substantial and significantly different from zero. This result suggests that decreasing the costs of schooling actually *increases* the share of high income college graduates. In Column (7), college graduates living near a 4-year college are 8.3% more likely to have a father with some college (significant at the 10% level), though after controlling for background characteristics this relationship becomes much smaller and statistically no different from zero. Finally, even though higher ability students are slightly more likely to live near 4-year colleges, *college graduates* who went to high school near a 4-year college have cognitive test scores that are between 0.109 and 0.184 standard deviations *below* the test scores of college graduates who went to high school far from a 4-year college. The number at the lower end of the range is statistically insignificant at the 10% level, but the larger number is significant at the 5% level.

These results suggest that the pool of college graduates who went to high school in areas that faced low costs of schooling is on average lower ability than the pool of college graduates who went to high school in areas facing relatively high costs of schooling. Thus low schooling costs appear to increase the share of (relatively) low ability college graduates. This result is not driven by the availability of different school qualities for people with and without access to a local college. For example, one might be concerned that "local colleges" are schools that appeal to relatively worse students. On the contrary, I find that students living near 4-year colleges tend to attend higher quality schools: college graduates who lived within 50 miles of a 4-year college attended schools whose average SAT scores were 50.078 points higher (S.E. = 15.191) than students living more than 50 miles from a 4-year college.

The results in Table 3 do not vary substantially by gender except along one important dimension: cognitive test scores. Tables 4 and 5 show that, along nearly all observable characteristics, the effect of direct college costs on the marginal male college graduate is very close to the the effect of direct college costs on the characteristics of the marginal female college graduate. The one notable exception is for cognitive test scores. The finding that lower college costs tend to draw in a lower ability pool of college graduates appears to be concentrated on men. Table 4 displays evidence that male college graduates who went to high school near a 4-year college had test scores that were between 0.304 and 0.420 standard deviations below the test scores of male college graduates without access to a local 4-year college. Both are significant at the 1% level. In contrast, the difference for female college graduates was 0.097 standard deviations with no controls and 0.057 with controls. Neither is significant at the 10% level.

Taken as a whole, the evidence suggests the following: having access to a local college is not strongly correlated a set of key observable characteristics. If we take this as a measure of direct schooling costs, the evidence also suggests that along many dimensions such as parent education, race, gender, and number of siblings, reducing college costs does not substantially alter the composition of the pool of college graduates. Along two important dimensions, however, reducing college costs does appear to change the way that college graduates look: college graduates facing lower costs tend to be higher income and to have lower cognitive test scores. In the latter case, the effect appears to be concentrated among male college graduates.

4.2 Elimination of the Social Security Student Benefit Program

To test the robustness of the results in the previous section using a very different source of variation in college costs, I present evidence from the elimination of the Social Security Student Benefit Program. As with the last section, I begin by presenting evidence that this measure of college costs is plausibly exogenous by estimating the correlation between it and observable background characteristics. Aid eligibility is proxied by an interaction between a dummy equal to 1 if the respondent was a senior in high school before the policy change and a dummy equal to one if the respondent's father died before the respondent's 18th birthday. If aid eligibility is truly exogenous, the interaction between deceased father and "before the policy change" dummies should not be related to observable characteristics conditional on cohort and deceased father dummies. To test this, I estimate the following equations:

$$q = \alpha_1 Before_i \times Deceased\ Father_i + \alpha_2 Before_i + \alpha_3 Deceased\ Father_i + \mu_i, \quad (8)$$

where q is once again some pre-college characteristic or set of characteristics, $Before$ is an indicator variable equal to 1 if the respondent was a senior in high school before the policy elimination and zero otherwise, $Deceased\ Father$ is an indicator variable equal to 1 if the respondent was no older than 18 when his father died, and μ is an error term. In practice, q is AFQT score, family income, number of siblings, and dummy variables for non-white, sex, and having a father/mother who attended college.

Table 6 presents results. In all cases, the relationship between aid eligibility and each variable is small and insignificant. When each variable is used on its own as a regressor, the coefficients are all close to zero and statistically no different from zero; when local college is regressed on all background variables, once again none are significantly different from zero. The evidence is therefore consistent with the notion that aid eligibility is exogenous conditional on $Before$ and $Deceased\ Father$ effects.

I now turn to estimating the effect of aid eligibility on the background characteristics of marginal college graduates. To do this, I estimate the following equation:

$$q_i = \beta_1 c_i + X_i' \beta_2 + \epsilon_i, \tag{9}$$

where $c_i = Before_i \times Deceased\ Father_i$ and X_i is a vector that consists of $Before_i$, $Deceased\ Father_i$, $1 - s_i$, a full set of interactions between these variables, and an interaction between $1 - s_i$ and c_i. $1 - s_i = 1$ if the respondent did not earn a bachelor's degree by age 24. The coefficient of interest is β_1. Unlike in Section 4.1, I cannot separate the results by sex because there are not enough college graduates with deceased fathers to do this.

Table 7 presents results from estimating Equation 9. The coefficient of interest is on $Before \times Deceased\ Father$. As was the case with local colleges, the marginal aid eligible college graduate had significantly higher income than the marginal aid ineligible college graduate by about \$10,500 to \$11,000. On most other dimensions, aid eligible marginal graduates look similar to aid-ineligible marginal graduates – they are about as likely to be white (Columns (3) and (4)), have college educated parents (Columns (9) - (12)), and have similar numbers of siblings (Columns (7) and

(8)). The marginal aid eligible college graduate is 32 - 33% more likely to be male than the marginal aid-ineligible graduate, but this difference is not statistically significant at the 10% level. Finally, the marginal aid-eligible college graduate has a lower AFQT score than the marginal aid-ineligible graduate (Columns (13) - (16)). The average AFQT score is between -0.21 and -0.36 standard deviations lower for aid eligible college graduates compared to aid ineligible graduates. While these differences are not significantly different from zero at conventional levels, if one drops the respondents whose missing AFQT scores were replaced by imputed values, the coefficient becomes larger (in absolute value) and statistically significant (between -0.23 and -0.48 standard deviations with the latter significant at the 5% level). Finally, Columns (15) and (16) present evidence that the marginal aid eligible college graduate is between 21% and 31% more likely to have an AFQT score below the median than the marginal aid ineligible college graduate. These estimates are significant at the 5% and 1% levels, respectively.

Table 8 extends this analysis and considers the students on the margin of enrolling in college. In contrast to the marginal college graduates, I find no differences between the average characteristics of aid-eligible and -ineligible college enrollees. No coefficient is statistically different from zero, and in nearly all cases the coefficients are very close to zero. Thus direct schooling costs appear to change some characteristics of marginal college graduates but not enrollees.

5 Conclusion

In this paper I have presented evidence that direct college costs play a large role in determining non-random selection into college. To generate plausibly exogenous variation in college costs, I use presence of a nearby 4-year college and the elimination of the Social Security Student Benefit Program. The results from both sources of variation are very consistent with each other. Both suggest that changing the costs of schooling does not change the likelihood that the marginal college graduate is female, non-white, or has college educated parents. Changing schooling costs does appear to change the marginal graduate in two important ways: lowering the costs of schooling leads to a marginal graduate who comes from a higher income family and a marginal graduate who has cognitive test scores that are between one-fifth and one-half a standard deviation lower. However, I find some evidence using local colleges as a measure of costs that the cognitive test score results are concentrated among men.

6 Data Appendix

6.1 High School and Beyond

6.1.1 Variables

- Parent education: fy55 (Father's education), fy56 (Mother's Education)
- Family Income: fy111. Income variable gives 8 ranges of income; the midpoint of each is taken as the value. The last category is any income level greater than $50,000; I take 75,000 as the value for individuals in that category.
- Number of Siblings: fy106
- Highest Degree by 1992 (aged 28): y4205
- Cognitive test: bytest
- Sex: sex
- White: race
- Birth Date: birthyr, birthmo
- Higher education degree: tdeg1...7
- Year earned degree: tdyear1...7
- Bachelor Degree by age 24: tdyear1...7, birthyr, tdeg1...7
- School ID: schid or schoolid
- ID For parent file: sparid
- ID: id
- Distance to nearest 4-year college: colldist

6.1.2 Creation of the Data Set

Missing Data

In most cases, missing values for explanatory variables were imputed. For continuous variables, this involved replacing the missing value with the average of the variable for non-missing cases. This included the measure of family income as well as number of siblings. A dummy indicating whether the case was missing this variable was created and included in regressions. For dummy variables such as whether the respondent's mother and father attended college, an additional category was created for the missing variable, and in the regressions the omitted variable in this set of dummies was the "education observed, did not attend college" category. Observations missing race, sex, the cognitive test score, individual/school IDs, educational attainment data, or a FICE code were dropped. Observations missing individual and school IDs are also dropped.

Creation of key variables

- Cognitive test score: after dropping anyone with a missing id or school id, the cognitive test score is converted to a mean 0 standard deviation 1 variable (for non-missing cases).
- Local College: equals 1 if respondent's high school was within 50 miles of a 4-year college
- Bachelor's degree by age 24: Age at bachelor's degree receipt is calculated by age at degree receipt minus birth year for individuals who earned a bachelor's degree. If the respondent was 24 or younger, this variable takes on the value 1 and 0 otherwise.

6.2 NLSY

Data appendix to be added shortly.

References

Angrist, J. D., and A. B. Krueger (1999): "Empirical Strategies in Labor Economics," in *Handbook of Labor Economics*, ed. by O. Ashenfelter, and D. Card, vol. 3 of *Handbook of Labor Economics*, chap. 23, pp. 1277–1366. Elsevier.

Angrist, J. D., and J.-S. Pischke (2008): *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Autor, D. H., and M. G. Duggan (2003): "The Rise In The Disability Rolls And The Decline In Unemployment," *The Quarterly Journal of Economics*, 118(1), 157–205.

Bedard, K. (2001): "Human Capital versus Signaling Models: University Access and High School Dropouts," *Journal of Political Economy*, 109(4), 749–775.

Betts, J. R., and D. Morrell (1999): "The Determinants of Undergraduate Grade Point Average: The Relative Importance of Family Background, High School Resources, and Peer Group Effects," *Journal of Human Resources*.

Bound, J., and S. Turner (2002): "Going to War and Going to College: Did World War II and the G.I. Bill Increase Educational Attainment for Returning Veterans?," *Journal of Labor Economics*, 20(4), 784–815.

Cameron, S. V., and C. Taber (2004): "Estimation of Educational Borrowing Constraints Using Returns to Schooling," *Journal of Political Economy*, 112(1), 132–182.

Card, D. (1993): "Using Geographic Variation in College Proximity to Estimate the Return to Schooling," NBER Working Papers 4483, National Bureau of Economic Research, Inc.

Carneiro, P., and J. J. Heckman (2002): "The Evidence on Credit Constraints in Post–secondary Schooling," *Economic Journal*, 112(482), 705–734.

Conley, T., and C. Taber (2009): "Inference with "Difference in Differences" with a Small Number of Policy Changes," *Review of Economics and Statistics, forthcoming.*

DeGroot, M. (1970): *Optimal Statistical Decisions*. John Wiley & Sons, Inc., Hoboken, NJ.

Donohue, J. J., and S. D. Levitt (2001): "The Impact Of Legalized Abortion On Crime," *The Quarterly Journal of Economics*, 116(2), 379–420.

Dynarski, S. (2000): "Hope and for Whom? Financial Aid for the Middle Class and Its Impact on College Attendance," *The National Tax Journal*, 53(3), 629–662.

Dynarski, S. M. (2003): "Does Aid Matter? Measuring the Effect of Student Aid on College Attendance and Completion," *American Economic Review*, 93(1), 279–288.

Griliches, Z. (1977): "Estimating the Returns to Schooling: Some Econometric Problems," *Econometrica*, 45(1), 1–22.

Griliches, Z., and W. M. Mason (1972): "Education, Income, and Ability," *The Journal of Political Economy*, 80(3), S74–S103.

Gruber, J., P. Levine, and D. Staiger (1999): "Abortion Legalization And Child Living Circumstances: Who Is The "Marginal Child"?," *The Quarterly Journal of Economics*, 114(1), 263–291.

Hearn, J. C. (1991): "Academic and Nonacademic Influences on the College Destinations of 1980 High School Graduates," *Sociology of Education*, 64(3).

Imbens, G. W., and J. D. Angrist (1994): "Identification and Estimation of Local Average Treatment Effects," *Econometrica*, 62(2), 467–75.

Kane, T. J. (1994): "College Entry by Blacks since 1970: The Role of College Costs, Family Background, and the Returns to Education," *Journal of Political Economy*, 102(5), 878–911.

Lang, K. (1993): "Ability Bias, Discount Rate Bias and the Return to Education," mimeo, Boston University.

Meyer, B. D. (1995): "Natural and Quasi-experiments in Economics," *Journal of Business & Economic Statistics*, 13(2), 151–61.

Oppedisano, V. (2008): "The (Adverse) Effects of Expanding Higher Education: Evidence from Italy," mimeo, University of Torino.

Spence, A. M. (1973): "Job Market Signaling," *The Quarterly Journal of Economics*, 87(3), 355–74.

Stanley, M. (2003): "College Education And The Midcentury Gi Bills," *The Quarterly Journal of Economics*, 118(2), 671–708.

Taber, C. R. (2001): "The Rising College Premium in the Eighties: Return to College or Return to Unobserved Ability?," *Review of Economic Studies*, 68(3), 665–91.

Turner, S., and J. Bound (2003): "Closing the Gap or Widening the Divide: The Effects of the G.I. Bill and World War II on the Educational Outcomes of Black Americans," *The Journal of Economic History*, 63(01), 145–177.

Willis, R. J., and S. Rosen (1979): "Education and Self-Selection," *Journal of Political Economy*, 87(5), S7–36.

Figure 1: Selection Estimator

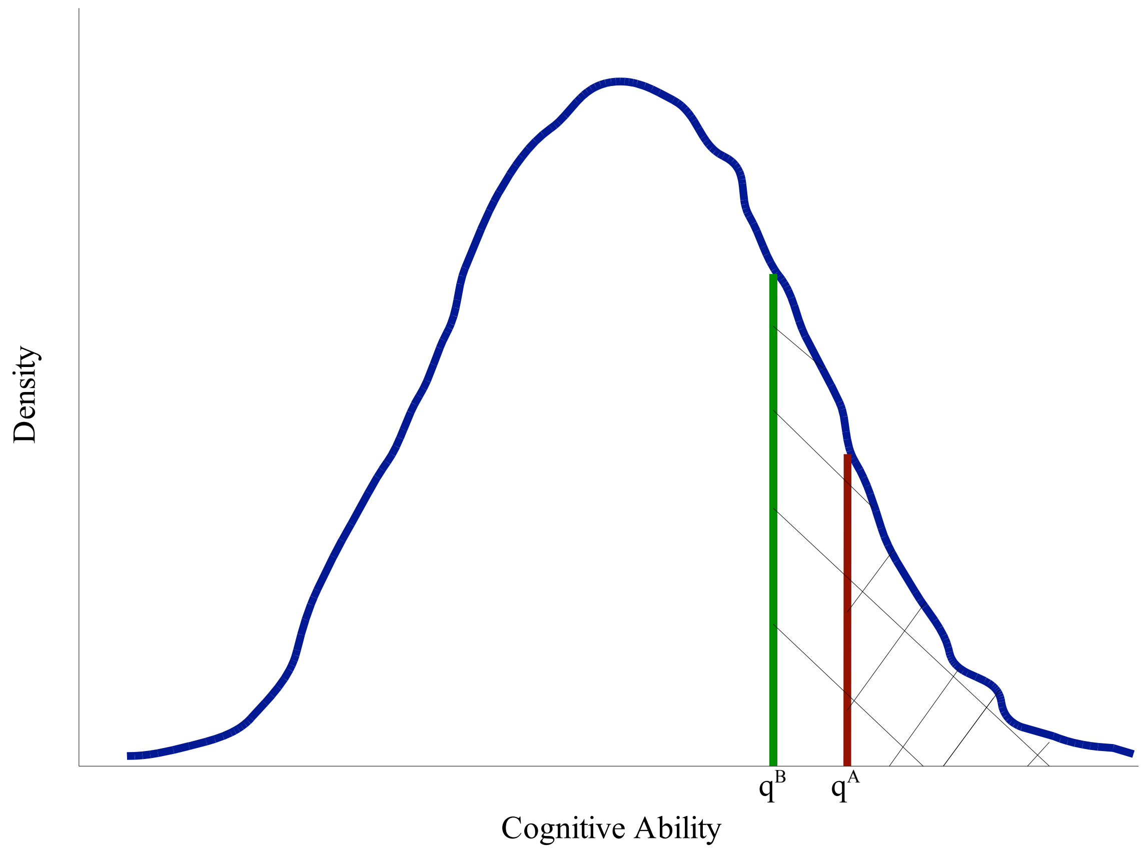

Table 1: Sample Statistics

	HS&B (1)	NLSY (2)
Family Income	$28,743	$28,737
Male	0.498	0.517
White	0.726	0.803
Father Attended College	0.337	0.337
Mother Attended College	0.270	0.234
Siblings	2.924	3.086
Bachelor's Degree by Age 24	0.143	0.228
Local College	0.889	
Before x Deceased Father		0.027
Before		0.695
Deceased Father		0.041

Notes: Family Income (in 1982 dollars) is the respondent's family income in her senior year of high school. Local college is a dummy variable equal to 1 if the respondent attended high school within 50 miles of a 4-year college. Eligibility for the Social Security Student Benefit Program is proxied by the interaction between Before and Deceased Father. Before equals 1 if the respondent was a senior in high school before the policy change, and Deceased Father equals 1 if the respondent was 18 or younger at the time of the father's death.

Table 2: Relationship Between Local College and Background Characteristics

		N = 13,252					
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Family Income	0.006						0.005
	[0.002]**						[0.002]**
Cognitive Test Score		0.008					0.004
		[0.004]*					[0.004]
Dad Attended College			0.017				0.009
			[0.008]*				[0.007]
Mom Attended College			0.004				-0.002
			[0.008]				[0.008]
White				0.001			-0.007
				[0.010]			[0.009]
Male					-0.001		-0.002
					[0.004]		[0.004]
Number of Siblings						-0.005	-0.004
						[0.002]*	[0.002]*
R-squared	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Notes: ** Significant at 1% level, * significant at 5% level, + significant at 10% level. Standard errors are clustered by high school. Local college is defined as having a 4-year college within 50 miles of the respondent's high school. Family income is from 1982 and in $10,000. Missing data is replaced with the average for non-missing observations, and a dummy indicating whether the observation is missing is included in the regression. Results are weighted by base year weights.

Table 3: Relationship Between Local College and Background Characteristics of Marginal College Graduates

	N = 13,564					
	Famly Income		Non-White		Siblings	
	(1)	(2)	(3)	(4)	(5)	(6)
Local College	0.684	0.597	0.002	0.026	0.075	0.117
	[0.204]**	[0.179]**	[0.033]	[0.032]	[0.107]	[0.110]
Local College x No Bachelor's Degree	-0.397	-0.402	-0.040	-0.049	-0.303	-0.305
	[0.202]*	[0.179]*	[0.031]	[0.029]+	[0.131]*	[0.131]*
No Bachelor's Degree	-0.463	0.014	0.227	0.147	0.841	0.588
	[0.189]*	[0.167]	[0.028]**	[0.027]**	[0.122]**	[0.124]**
R-squared	0.04	0.15	0.02	0.09	0.02	0.05
	Dad Attended College		Mom Attended College		Cognitive Test	
	(7)	(8)	(9)	(10)	(11)	(12)
Local College	0.083	0.033	0.052	0.012	-0.109	-0.184
	[0.049]+	[0.043]	[0.054]	[0.052]	[0.086]	[0.085]*
Local College x No Bachelor's Degree	-0.024	-0.001	-0.006	0.013	0.190	0.197
	[0.050]	[0.046]	[0.057]	[0.055]	[0.091]*	[0.090]*
No Bachelor's Degree	-0.295	-0.241	-0.233	-0.194	-1.258	-0.899
	[0.048]**	[0.044]**	[0.054]**	[0.052]**	[0.086]**	[0.085]**
R-squared	0.07	0.15	0.04	0.1	0.16	0.33
Background Controls++	No	Yes	No	Yes	No	Yes

Notes: ** Significant at 1% level, * significant at 5% level, + significant at 10% level. Standard errors are clustered by high school. Local college is defined as having a 4-year college within 50 miles of the respondent's high school. Results are from regressions of the stated dependent variable on presence of local 4-year college, an indicator for not having a bachelor's degree by age 24 and it's interaction with local college, family income in 1982, dummies equal to one if each parent attended college, a dummy for white, a dummy for male, and number of siblings. Missing data is replaced with the average for non-missing observations, and a dummy indicating whether the observation is missing is included in the regression. Results are weighted by base year weights.
++ Background controls include all background characteristics except for the dependent variable and all missing value indicators.

Table 4: Relationship Between Local College and Background Characteristics of Marginal College Graduates - Men

	N = 6,674					
	Famly Income		Non-White		Siblings	
	(1)	(2)	(3)	(4)	(5)	(6)
Local College	0.731	0.533	0.018	0.044	0.097	0.159
	[0.308]*	[0.289]+	[0.041]	[0.040]	[0.151]	[0.156]
Local College x No Bachelor's Degree	-0.459	-0.338	-0.027	-0.043	-0.270	-0.309
	[0.299]	[0.282]	[0.046]	[0.044]	[0.181]	[0.182]+
No Bachelor's Degree	-0.336	0.022	0.217	0.149	0.774	0.561
	[0.280]	[0.262]	[0.043]**	[0.041]**	[0.168]**	[0.171]**
R-squared	0.03	0.14	0.02	0.09	0.01	0.05
	Dad Attended College		Mom Attended College		Cognitive Test	
	(7)	(8)	(9)	(10)	(11)	(12)
Local College	0.158	0.106	0.082	0.042	-0.304	-0.420
	[0.069]*	[0.065]	[0.068]	[0.067]	[0.114]**	[0.101]**
Local College x No Bachelor's Degree	-0.107	-0.080	-0.037	-0.016	0.384	0.458
	[0.072]	[0.067]	[0.071]	[0.069]	[0.122]**	[0.112]**
No Bachelor's Degree	-0.197	-0.152	-0.167	-0.135	-1.437	-1.136
	[0.068]**	[0.064]*	[0.067]*	[0.065]*	[0.115]**	[0.105]**
R-squared	0.06	0.14	0.03	0.09	0.15	0.33
Background Controls++	No	Yes	No	Yes	No	Yes

Notes: ** Significant at 1% level, * significant at 5% level, + significant at 10% level. Standard errors are clustered by high school. Local college is defined as having a 4-year college within 50 miles of the respondent's high school. Results are from regressions of the stated dependent variable on presence of local 4-year college, an indicator for not having a bachelor's degree by age 24 and it's interaction with local college, family income in 1982, dummies equal to one if each parent attended college, a dummy for white, a dummy for male, and number of siblings. Missing data is replaced with the average for non-missing observations, and a dummy indicating whether the observation is missing is included in the regression. Results are weighted by base year weights.
++ Background controls include all background characteristics except for the dependent variable and all missing value indicators

Table 5: Relationship Between Local College and Background Characteristics of Marginal College Graduates - Women

	N = 6,890					
	Famly Income		Non-White		Siblings	
	(1)	(2)	(3)	(4)	(5)	(6)
Local College	0.640	0.638	-0.014	0.004	0.056	0.075
	[0.253]*	[0.235]**	[0.041]	[0.041]	[0.147]	[0.148]
Local College x No Bachelor's Degree	-0.344	-0.443	-0.053	-0.051	-0.336	-0.302
	[0.267]	[0.250]+	[0.045]	[0.044]	[0.171]*	[0.170]+
No Bachelor's Degree	-0.588	-0.009	0.237	0.140	0.906	0.610
	[0.252]*	[0.237]	[0.043]**	[0.043]**	[0.157]**	[0.160]**
R-squared	0.05	0.16	0.03	0.11	0.02	0.05
	Dad Attended College		Mom Attended College		Cognitive Test	
	(7)	(8)	(9)	(10)	(11)	(12)
Local College	0.005	-0.045	0.015	-0.023	0.097	0.057
	[0.069]	[0.066]	[0.073]	[0.070]	[0.109]	[0.114]
Local College x No Bachelor's Degree	0.061	0.081	0.031	0.045	-0.017	-0.067
	[0.073]	[0.071]	[0.077]	[0.073]	[0.122]	[0.120]
No Bachelor's Degree	-0.395	-0.333	-0.301	-0.253	-1.070	-0.655
	[0.070]**	[0.068]**	[0.073]**	[0.070]**	[0.116]**	[0.116]**
R-squared	0.09	0.17	0.06	0.11	0.18	0.34
Background Controls++	No	Yes	No	Yes	No	Yes

Notes: ** Significant at 1% level, * significant at 5% level, + significant at 10% level. Standard errors are clustered by high school. Local college is defined as having a 4-year college within 50 miles of the respondent's high school. Results are from regressions of the stated dependent variable on presence of local 4-year college, an indicator for not having a bachelor's degree by age 24 and it's interaction with local college, family income in 1982, dummies equal to one if each parent attended college, a dummy for white, a dummy for male, and number of siblings. Missing data is replaced with the average for non-missing observations, and a dummy indicating whether the observation is missing is included in the regression. Results are weighted by base year weights.
++Background controls include all background characteristics except for the dependent variable and all missing value indicators

Table 6: Relationship between Aid Eligibility and Background Characteristics

			N = 4,184				
	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Family Income	-0.0003						-0.0005
	[0.0008]						[0.0008]
AFQT		0.0011					0.0012
		[0.0018]					[0.0020]
Dad Attended College			-0.0003				0.0001
			[0.0036]				[0.0036]
Mom Attended College			-0.0026				-0.0030
			[0.0043]				[0.0042]
White				-0.0011			-0.0066
				[0.0039]			[0.0046]
Male					0.0003		0.0002
					[0.0033]		[0.0033]
Number of Siblings						-0.0016	-0.0016
						[0.0010]	[0.0011]
R-squared	0.69	0.69	0.69	0.69	0.69	0.69	0.69

Notes: ** significant at the 1% level, * significant at the 5% level, + significant at the 10% level. Robust standard errors in parentheses. Dependent variable is the interaction between the Deceased Father and Before dummies. AFQT is converted to a mean zero standard deviation one variable. Family income is from the respondent's senior year in high school and is divided by $10,000. Missing data is replaced with the average for non-missing observations in each Before x Deceased father cell, and a dummy indicating whether the observation is missing is included in the regression. Results are weighted by base year weights.

Table 7: Relationship between Aid Eligibility and Background Characteristics of Marginal College Graduate

	N = 4,105							
	Family Income		Non-White		Male		Siblings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Before x Deceased Father	1.103	1.053	0.040	0.067	0.333	0.320	-0.495	-0.026
	[0.662]+	[0.547]+	[0.116]	[0.112]	[0.245]	[0.249]	[0.676]	[0.635]
Before	0.191	0.206	0.001	0.001	0.010	0.008	0.142	0.174
	[0.178]	[0.166]	[0.018]	[0.019]	[0.049]	[0.049]	[0.147]	[0.148]
Deceased Father	-2.061	-1.680	0.037	-0.080	-0.218	-0.184	0.896	0.278
	[0.325]**	[0.304]**	[0.096]	[0.094]	[0.205]	[0.210]	[0.548]	[0.516]
R-squared	0.09	0.23	0.03	0.27	0	0.01	0.05	0.13
	Dad Attended College		Mom Attended College		AFQT		AFQT < Median	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Before x Deceased Father	-0.094	-0.164	0.095	0.106	-0.212	-0.361	0.207	0.310
	[0.256]	[0.281]	[0.250]	[0.260]	[0.288]	[0.251]	[0.099]*	[0.107]**
Before	-0.018	-0.026	0.005	0.008	0.003	-0.007	-0.008	-0.001
	[0.046]	[0.041]	[0.049]	[0.044]	[0.054]	[0.051]	[0.021]	[0.021]
Deceased Father	-0.174	-0.003	-0.216	-0.106	-0.045	0.300	-0.067	-0.260
	[0.220]	[0.262]	[0.211]	[0.235]	[0.202]	[0.168]+	[0.019]**	[0.060]**
R-squared	0.11	0.31	0.09	0.26	0.2	0.38	0.12	0.27
Background Controls++	No	Yes	No	Yes	No	Yes	No	Yes

Notes: ** significant at the 1% level, * significant at the 5% level, + significant at the 10% level. Robust standard errors in parentheses. All regressions include a full set of interactions between Before, Deceased Father, and No Bachelor's Degree dummy variables as well as the level effects. AFQT is converted to a mean zero standard deviation one variable. Family income is from the respondent's senior year in high school and is divided by $10,000. Missing data is replaced with the average for non-missing observations in each Before x Deceased father cell, and a dummy indicating whether the observation is missing is included in the regression. Results are weighted by base year weights.

++ Background controls include all background characteristics except for the dependent variable and all missing value indicators.

Table 8: Relationship between Aid Eligibility and Background Characteristics of Marginal College Enrollee

	N = 4,184							
	Family Income		Non-White		Male		Siblings	
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Before x Deceased Father	0.324	0.297	-0.071	-0.048	-0.006	-0.030	0.116	0.376
	[0.398]	[0.283]	[0.120]	[0.090]	[0.158]	[0.160]	[0.552]	[0.596]
Before	0.118	0.084	0.000	0.002	-0.022	-0.028	0.138	0.182
	[0.106]	[0.097]	[0.016]	[0.015]	[0.030]	[0.030]	[0.097]	[0.096]+
Deceased Father	-1.499	-1.178	0.146	0.011	-0.040	0.014	0.947	0.521
	[0.319]**	[0.204]**	[0.109]	[0.081]	[0.140]	[0.144]	[0.432]*	[0.505]
R-squared	0.08	0.22	0.01	0.3	0	0.01	0.05	0.13
Background Controls	No	Yes	No	Yes	No	Yes	No	Yes
	Dad Attended College		Mom Attended College		AFQT		AFQT < Median	
	(9)	(10)	(11)	(12)	(13)	(14)	(15)	(16)
Before x Deceased Father	-0.029	0.013	-0.164	-0.166	0.066	-0.012	-0.071	-0.017
	[0.151]	[0.143]	[0.158]	[0.148]	[0.237]	[0.151]	[0.114]	[0.080]
Before	0.020	0.000	0.044	0.035	0.022	0.000	0.007	0.019
	[0.030]	[0.026]	[0.029]	[0.026]	[0.045]	[0.039]	[0.022]	[0.020]
Deceased Father	-0.126	-0.044	-0.013	0.076	-0.170	0.118	0.073	-0.077
	[0.135]	[0.129]	[0.145]	[0.138]	[0.213]	[0.128]	[0.102]	[0.067]
R-squared	0.14	0.32	0.11	0.27	0.19	0.39	0.12	0.28
Background Controls	No	Yes	No	Yes	No	Yes	No	Yes

Notes: ** significant at the 1% level, * significant at the 5% level, + significant at the 10% level. Robust standard errors in parentheses. All regressions include a full set of interactions between Before, Deceased Father, and Did Not Attend College dummy variables as well as the level effects. AFQT is converted to a mean zero standard deviation one variable. Family income is from the respondent's senior year in high school and is divided by $10,000. Missing data is replaced with the average for non-missing observations in each Before x Deceased father cell, and a dummy indicating whether the observation is missing is included in the regression. Results are weighted by base year weights.

++ Background controls include all background characteristics except for the dependent variable and all missing value indicators.