ESTIMATING THE RETURN TO COLLEGE SELECTIVITY OVER THE CAREER
USING ADMINISTRATIVE EARNINGS DATA

Stacy Dale
Alan B. Krueger

Estimating the Return to College Selectivity over the Career Using Administrative Earnings
Data
Stacy Dale and Alan B. Krueger
NBER Working Paper No. 17159
June 2011
JEL No. I21,J31

ABSTRACT

We estimate the monetary return to attending a highly selective college using the College and Beyond
(C&B) Survey linked to Detailed Earnings Records from the Social Security Administration (SSA).
This paper extends earlier work by Dale and Krueger (2002) that examined the relationship between
the college that students attended in 1976 and the earnings they self-reported reported in 1995 on the
C&B follow-up survey. In this analysis, we use administrative earnings data to estimate the return
to various measures of college selectivity for a more recent cohort of students: those who entered college
in 1989. We also estimate the return to college selectivity for the 1976 cohort of students, but over
a longer time horizon (from 1983 through 2007) using administrative data.

We find that the return to college selectivity is sizeable for both cohorts in regression models that control
for variables commonly observed by researchers, such as student high school GPA and SAT scores.
However, when we adjust for unobserved student ability by controlling for the average SAT score
of the colleges that students applied to, our estimates of the return to college selectivity fall substantially
and are generally indistinguishable from zero. There were notable exceptions for certain subgroups.
For black and Hispanic students and for students who come from less-educated families (in terms of
their parents' education), the estimates of the return to college selectivity remain large, even in models
that adjust for unobserved student characteristics.

Stacy Dale
Mathematica Policy Research
P.O. Box 2393
Princeton, NJ 08543-2393
sdale@mathematica-mpr.com

Alan B. Krueger
Industrial Relations Section
Firestone Library
Princeton University
Princeton, NJ 08544
and NBER
akrueger@princeton.edu

I. INTRODUCTION

Understanding the labor market return to college quality is crucial for parents deciding where to send their children to college, for colleges selecting students, and for policymakers deciding whether to invest additional resources in higher quality institutions. However, obtaining unbiased estimates of the return to college quality is difficult due to unobserved characteristics that affect both a student's attendance at a highly selective college and their later earnings. In particular, the same characteristics (such as ambition) that lead students to apply to highly selective colleges may also be rewarded in the labor market. Likewise, the attributes that admissions officers are looking for when selecting students for college may be similar to the attributes that employers are seeking when hiring and promoting workers.

Early research attempted to overcome this omitted-variable bias by controlling for observed student characteristics, such as high school grades, standardized test scores, and parental background (see, for example, Monks 2000 and Brewer and Ehrenberg 1996). More recent research has tried to overcome the bias created by unobserved variables through a variety of techniques. Hoekstra (2009) uses a regression discontinuity design that compares the earnings of students who were just above the admissions cutoff for a state university to those that were just below it. He finds that attending the flagship state university results in 20 percent higher earnings five to ten years after graduation for white men, but he does not find an effect on earnings for white women. Lindahl and Regner (2005) use sibling data to illustrate that the returns to college quality might be overstated if family characteristics are not fully adjusted for, as cross-sectional estimates of the return-to-schooling are twice as large as within-family estimates. Black and Smith (2004) use propensity score methods to estimates to estimate the return to college quality for the 1979 cohort of the National Longitudinal Survey of Youth. The magnitude of the effect of college

quality is similar whether they use OLS or propensity score methods, however, when they use propensity score methods, the standard errors are large so the effect of college quality is not statistically significant. Dale and Krueger (2002) hold unobserved student quality constant by controlling for the selectivity of colleges that students apply to and are accepted or rejected by, and find that the return to a variety of college characteristics falls substantially after implementing their selection-correction.

To reconcile the findings across methods and data sources, Long (2008) re-estimated the models used by Dale and Krueger (2002) and Black and Smith (2004); he also estimated an ordinary least squared (OLS) model that controlled for student attributes, family background and neighborhood characteristics, as well as an instrumental variables model that used the quality of colleges located within a certain radius of the student's home as an instrument for the quality of the college that the student attends. In general, he found that a variety of college characteristics have positive effects on educational attainment, even within specifications that adjust for selection bias; however, there is not consistent evidence for a positive relationship between college characteristics and earnings across the selection adjusted models.

Because the percentage of students enrolling in college has increased over time, one might expect that it would be more important for students who entered college recently to distinguish themselves by attending more selective colleges.[1] However, little research has examined recent cohorts, and those studies that do use recent cohorts tend to model earnings early in the career. Long (2008 and 2009), for example, used the 1992 cohort of the National Education Longitudinal Study (NELS) to examine the earnings of students relatively early in their careers, when they were only 26 years old.

---

[1] For example, the percentage of 18 to 24-year olds enrolling in college increased from 26 percent in 1975 to 32 percent in 1990 (Fox and Snyder, 2005).

This paper replicates and extends earlier work that examined the relationship between the college that students attended in 1976 and the earnings they reported in 1995 in the C&B follow-up survey (Dale and Krueger 2002) in important respects. First, we estimate the return several college characteristics that are commonly used as proxies for college quality (college average SAT score, the Barron's index, and net tuition) for a recent cohort of students—those who entered college in 1989. By linking the C&B data to administrative records from the Social Security Administration (SSA), we are able to follow this cohort for 18 years after the students entered (and 14 years after they likely would have graduated from) college. Second, we estimate the return to college characteristics for the 1976 cohort over a long time horizon, from 1983 to 2007. Because we use administrative earnings records from tax data, our earnings measure is presumably more reliable than much of the prior literature, which is generally based on self-reported earnings. The use of administrative earnings data allows us to follow a recent cohort of students over a longer period of time than is possible in many of the longitudinal databases that are typically used to study the returns to college characteristics. For example, the NELS, High School and Beyond, and the National Longitudinal Study of the High School Class of 1972 only follow students for 6 to 10 years after students would have likely graduated from college; while the National Longitudinal Study of Youth (NLSY) follows students for a long period of time, students from the relatively recent cohort (who were age 12 to 16 in 1997) are now too early in their post-collegiate careers to generate accurate estimates of the return to college characteristics.

As in the rest of the literature, we find that the return to each college characteristic is sizeable for both cohorts in cross-sectional least squares regression models that control for variables commonly observed by researchers (such as student characteristics and SAT scores). However, when we adjust for a proxy for unobserved student characteristics – namely, by

controlling for the average SAT score of the colleges that students applied to – our estimates for the return to college characteristics fall substantially and are generally indistinguishable from zero for both the 1976 and 1989 cohort of students. Notable exceptions are for racial and ethnic minorities (black and Hispanic students) and for students whose parents have relatively little education; for these subgroups, our estimates remain large, even in models that adjust for unobserved student characteristics. One possible explanation for this pattern of results is that highly selective colleges provide access to networks for minority students and for students from disadvantaged family backgrounds that are otherwise not available to them.

II. METHODS

The college application process involves a series of choices. First, students choose where to apply to college. Then, colleges decide which students to admit. Finally, students choose which college to attend from among the set of schools to which they were admitted. The difficulty with estimating the labor market return to college quality is that not all of the characteristics that lead students to apply to and attend selective colleges are observed by researchers, and unobserved student characteristics are likely to be positively correlated with both school quality and earnings.

We assume the equation relating earnings to the students' attributes is:

(1) $\ln W_i = \beta_0 + \beta_1 Q_i + \beta_2 X_{1i} + \beta_3 X_{2i} + \varepsilon_i,$

where Q is a measure of the selectivity of the college student i attended, X_1 and X_2 are two sets of characteristics that affect earnings, and ε_i is an idiosyncratic error term that is uncorrelated with the other explanatory variables (1). X_1 includes variables that are observable to researchers, such as grades and SAT scores, while X_2 includes variables that are not observable to researchers, such as

5

student motivation and creativity (that are at least partly revealed to admissions officers through detailed transcript information, essays, interviews, and recommendations). Both X_1 and X_2 affect the set of colleges that students apply to, whether they are admitted, and possibly which school they attend. The parameter β_1 represents the monetary payoff to attending a more selective college.

In the early literature on the returns to school quality, researchers estimated a wage equation that omitted X_2:

(2) $\ln W_i = \beta'_0 + \beta'_1 Q_i + \beta'_2 X_{1i} + u_i.$

Q_i is typically measured by the average SAT score of the school where the student attended college. Even if students randomly select the college they attend from the set of colleges that admitted them, estimation of (2) will yield biased and inconsistent parameter estimates of β_1 and β_2. If students choose their school randomly from their set of options, the payoff to attending a selective school will be biased upward because students with higher values of the omitted variable, X_2, are more likely to be admitted to and therefore attend highly selective schools. Since the labor market rewards X_2, and Q and X_2 are positively correlated, the coefficient on school quality will be biased upward.

To address the selection problem, we use one of the selection-adjusted models—referred to as the "self-revelation model"-- in Dale and Krueger (2002). This model assumes that students signal their potential ability, motivation and ambition by the choice of schools they apply to. If students with greater unobserved earnings potential are more likely to apply to more selective colleges, the error term in equation (2) could be modeled as a function of the average SAT score (denoted AVG) of the schools to which the student applied: $u_i = t_0 + t_1 AVG_i + v_i.$ If v_i is uncorrelated with the SAT score of the school the student attended, one can solve the selection problem by including AVG in the wage equation. This approach is called the "self-revelation" model because individuals reveal their

unobserved ability by their college application behavior. This model also includes dummy variables indicating the number of schools the students applied to (in addition to the average SAT score of the schools), because the number of applications a student submits may also reveal unobserved student traits, such as their ambition and persistence.

Dale and Krueger (2002) also estimated a matched applicant model that included an unrestricted set of dummy variables indicating groups of students who received the same admissions decisions (i.e., the same combination of acceptances and rejections) from the same set of colleges. The self-revelation model is a special case of the matched applicant model. The matched applicant model and self-revelation model yielded coefficients that were similar in size, but the self-revelation model yielded smaller standard errors. Because of the smaller sample size in the present analysis, we therefore focus on the self-revelation model.

As discussed in more detail Dale and Krueger (2002), a critical assumption of the self-revelation model is that students' enrollment decisions are uncorrelated with the error term of equation (2) and X_2. Our selection correction provides an unbiased estimate of β_1 if students' school enrollment decisions are a function of X_1 or any variable outside the model. However, it is possible that student matriculation decisions are correlated with X_2. For example, past studies have found that students are more likely to matriculate to schools that provide them with more generous financial aid packages (see, e.g., van der Klaauw [1997]). If more selective colleges provide more merit aid, the estimated effect of attending an elite college will be biased upward because relatively more students with higher values of X_2 will matriculate at elite colleges, even conditional on the outcomes of the applications to other colleges. If this is the case, our selection-adjusted estimates of the effect of college quality will be biased upward. However, if less selective colleges provide more generous merit aid (leading students with higher values of X_2 to attend less selective schools), the estimate

7

could be biased downward. More generally, our adjusted estimate would be biased upwards (downwards) if students with high unobserved earnings potential are more (less) likely to attend the more selective schools from the set of schools that admitted them.

Finally, it is possible that the effect of attending a highly selective school varies across individuals (that is, β_1 could have an "i" subscript), and students might sort among selective and less selective colleges based on their potential returns at that college. In such a model, our estimate of the return to attending a selective school can be biased upward or downward, and it would not be appropriate to interpret an estimate of β_1 as a causal effect for the average student. We explore these potential threats to the model below.

III. DATA

A. College and Beyond Data

Our study is based on data from the 1976 and 1989 cohorts of the College and Beyond Survey. The C&B dataset includes information drawn from the applications and transcripts of 34 colleges and universities (including 4 public universities, 4 historically black colleges and universities (HBCUs), 11 liberal arts college, and 15 private universities).[2] Much of the past research using the C&B data (such as Bowen and Bok (1998) and Dale and Krueger (2002)) excluded the 4 HBCUs.[3] In this analysis, we include the 27 schools that agreed to participate in this follow-up study, which included 3 public universities, 10 liberal arts colleges, 12 private universities,

---

[2] The C&B schools that participated in this study included Barnard College, Bryn Mawr College, Columbia University, Duke University, Emory University, Georgetown University, Kenyon College, Miami University of Ohio, Morehouse College, Northwestern University, Oberlin College, Penn State University, Princeton University, Smith College, Stanford University, Swarthmore College, Tufts University, Tulane University, University of Michigan, University of Pennsylvania, Vanderbilt University, Washington University, Wellesley College, Wesleyan University, Williams College, Xavier University, and Yale University.

[3] At the time that Dale and Krueger (2002) was written, the HBCUs were not part of the standard C&B dataset that was provided to researchers.

and 2 HBCUs. The students from these 27 schools represent 81 percent of the students included in the original C&B dataset.

The original C&B Survey, conducted by Mathematica Policy Research (MPR) in 1994-1996, contained questions about earnings, occupation, demographics, education, civic activities and life satisfaction.[4] MPR attempted to survey all students in the 1976 cohort from each of the 34 C&B schools, with the exception of the four public universities, where a sample (of 2000 individuals) was drawn that included all racial and ethnic minorities and athletes, and a random sample of other students. For the 1989 cohort, students from 21 colleges were surveyed.[5] The original 1989 C&B sample included all racial and ethnic minorities and athletes and a random sample of other students. Our regressions are weighted by the inverse of the probability that a student was included in the sample.

Early in the C&B questionnaire respondents were asked, "In rough order of preference, please list the other schools you seriously considered."[6] Respondents were then asked whether they applied to, and were accepted by, each of the schools they listed. Because our analysis relies on individuals' responses to these survey questions, the sample for our primary analysis is restricted to survey respondents.[7] Survey response rates were 80 percent for the 1976 cohort and 84 percent for the 1989 cohort.

---

[4] See Bowen and Bok (1998) for a full description of the schools and variables available in the C&B dataset.

[5] The participating C&B schools included in the 1989 survey were: Bryn Mawr College, Duke University, Georgetown University, Miami University of Ohio, Morehouse College, Oberlin College Penn State University, Princeton University, Stanford University, University of Michigan, University of Pennsylvania, Vanderbilt University, Washington University, Wellesley College, Wesleyan University, Williams College, Xavier University, and Yale University.

[6] Students who responded to the C&B pilot survey were not asked this question and are therefore excluded from our analysis.

[7] We were able to estimate our basic wage equation for the full sample of C&B students (including nonrespondents), and obtained results that were similar to those restricted to survey respondents. For example, if we include all students in the 1976 cohort with non-zero earnings, the coefficient on school SAT score in the 1995 earnings basic regression model was .059 with a standard error of .021; for the sample of survey respondents with non-zero earnings, the coefficient on school SAT score was .061 with a standard error of .019 (not shown).

The C&B Survey data are linked data drawn from individuals' applications to college (such as their SAT scores) as well as data drawn from their transcripts (such as grades in college). The C&B data were also merged to the Higher Education Research Institute's (HERI) Freshman Survey.

B. Regression Control Variables

Our basic regression model controls for race, sex, high school grade point average, student SAT score, predicted parental income and whether the student was a college athlete; our self-revelation model includes these same variables, and also includes the average SAT score of the schools to which a student applied and the number of applications they submitted. Race, gender, parental education and occupation (used to predict parental income), information on the schools the student applied to, whether the student was an athlete, and student SAT score were drawn from the C&B data. To construct other variables about students' performance in high school and their parents' income, we used data from the HERI freshman survey. Because parental income was missing for many individuals in the sample, but information for other family background information (such as parental education) was generally available, we predicted parental income by first regressing log parental income on mother's and father's education and occupation for the subset of students with available family income data, and then multiplied the coefficients from this regression by the values of the explanatory variables for every student in the sample. When regression control variables for SAT score or high school GPA were missing, we set the variable equal to the mean value for the sample, and also included a dummy variable indicating the data were missing.

C. College Characteristics

Each college's average SAT score and Barron's index of college selectivity (as reported in the 1978 and 1992 Barron's guides) was linked to student's responses to the questions concerning

the schools they applied to.[8] Because there were only one or two colleges in some categories of the Barron's index (particularly for the 1989 cohort), we represent the index with a continuous variable. The schools in our sample ranged from "Competitive" (coded a 2 on our continuous measure) to "Most Competitive" (coded a 5 on our continuous measure).

Net tuition for 1970, 1980, and 1990 was calculated by subtracting the average aid awarded to undergraduates from the sticker price tuition, as reported in the 11th and 12th and 14th editions of *American Universities and Colleges.* The 1976 net tuition was interpolated from the 1970 and 1980 net tuition, assuming an exponential rate of growth.

D. Earnings Measures

The Social Security Administration linked College and Beyond data to SSA's Detailed Earnings Records for the period of 1981 through 2007. The earnings measure for this analysis included the total earnings an individual reported to the Internal Revenue Service, including earnings from self-employment and earnings that were deferred to retirement plans.[9] SSA ran computer programs written by MPR on our behalf so that individual-level earnings data were never viewed by researchers outside SSA. SSA was able to match over 95 percent of the student records we provided. We converted annual earnings for each year to 2007 dollars using the Consumer Price Index (CPI-U). The SSA earnings measure used in our primary analysis is not topcoded; however, to compare to the C&B survey, for one analysis, we deliberately topcode the SSA data to be consistent with the C&B data, as described in more detail below.

In addition to creating outcome measures that captured annual earnings, we also created outcome measures that were the median of an individual's log annual earnings in 2007 dollars over

---

[8] HERI provided the file with average SAT scores for 1978 and Mark Long provided the file for 1992.
[9] Income received from capital gains is excluded.

five year intervals (1983 through 1987, 1998 through 1992, 1993 through 1997, 1998 through 2002, and 2003 through 2007). For example, the dependent variable for the period of 1993 to 1997 was the median (for each individual) of his or her log earnings in the five years from 1993 to 1997. By using medians over five-year intervals, we reduce noise in the earnings measure that would result from brief periods of time that the students may have spent out of the labor market or in non-covered employment.

Finally, the focus of this study is on the productivity of individuals who are employed (and not on whether individuals choose to or are able to work). Because we cannot identify full-time workers or hourly wages in the SSA administrative data, we generally restrict the sample to those earning over $13,822 (in 2007 dollars) during the year, the equivalent of earning the minimum wage for 2,000 hours at the 1982 federal minimum wage value (in 2007 dollars). For those regressions in which the dependent variable is median earnings over a five year interval, individuals were included in the sample if their median earnings over the five year interval exceeded $13,822; individuals were still included in the sample they earned less than $13,822 in a particular year as long as their median earnings exceeded $13,822. Estimates based on a sample that use this restriction are more precise than those based on a sample of all non-zero earners.[10] Also, as shown in Table 3, estimates based on the sample defined by this restriction are closer to estimates drawn from the sample of full-time workers (according to the C&B survey) than estimates drawn from a sample of all non-zero earners.[11]

IV. DESCRIPTIVE STATISTICS FOR SCHOOLS AND STUDENTS

---

[10] Approximately 10 percent of workers in our sample (that is, those with any earnings) in the 1976 cohort and 8 percent those in the 1989 cohort had earnings that were between zero and this minimum wage threshold ($13,822).

[11] Most studies on the return to college quality either restrict the sample to full-time workers (for example, Long 2008) or to non-zero earners (for example, Hoekstra 2009). If we estimate our model using levels instead of logs, and include those with no earnings, we obtain qualitatively similar results. For example, for the 1976 cohort, the return to selectivity was $26,575(7,566) in the basic model and fell to $2,154 (9,884) in the self-revelation model.

A. Characteristics of Colleges and Students in Sample

While the average SAT score for colleges in the C&B dataset ranged from approximately 800 to over 1300, most of the C&B schools were highly selective. The majority of C&B colleges fell into one of the top two Barron's categories (Most Competitive or Highly Competitive; see Appendix Table 1), and had an average SAT score of greater than 1175. The high selectivity of the colleges within the C&B database make it particularly well-suited for this analysis, because the majority of students that attend selective colleges submit multiple applications, which is necessary for our identification strategy. In contrast, many students who attend less selective colleges submit only one application, because many less selective colleges accept all students who apply. For example, according to data from the National Longitudinal Study of the High School Class of 1972, only 46 percent of students who attended college applied to more than one school.

The regression sample includes students who entered (but did not necessarily graduate from) one of the C&B schools. The student and school characteristics among those included in the regression sample are shown in Table 1. Because the schools included in the database were highly selective, the students were in the sample had high academic qualifications. The students in the 1976 cohort had an average SAT scores of 1160, and an average high school grade point average of 3.6. (Note that for ease of interpretation, in our tables and regression analysis, we divide our measures of school SAT score and student SAT score by 100). Similarly, for the 1989 cohort, the average student SAT score was over 1200, and the average GPA was 3.6. The percentage of students that were racial and ethnic minorities was higher for the 1989 cohort than for the 1976 cohort. Of the 1976 cohort, 6 percent of students were black and 1 percent were Hispanic, whereas 8 percent in the 1989 cohort were black and 3 percent were Hispanic. Finally, earnings for the sample were high, as the average of each individual's median earnings over the 2003 to 2007 period was $164,009 for the

1976 cohort. Average annual earnings in 2007 were $183,411 for the 1979 cohort and $139,698 for the 1989 cohort.

B. Application and Matriculation Patterns

Table 2 provides descriptive statistics about the application behavior of the students who entered one of the C&B schools in our study in 1976 or 1989. Nearly two-thirds of the 1976 cohort and 71 percent of the 1989 cohort submitted at least one additional application (in addition to the school they attended). For both cohorts, of those students submitting at least one additional application, over half applied to a school with a higher average SAT score than that of the college they attended, and nearly 90 percent of students were accepted to at least one additional school. Of those accepted to more than one school, about 35 percent were accepted to a more selective school than the one they ended up attending. The data for black and Hispanic students (shown in columns 2 and 4) are similar, though blacks and Hispanics were somewhat more likely than students in the full sample to be accepted to at least one additional school, and to be accepted to a more selective school than the one they attended.

While we could not explore whether students' unobserved ability is related to the school they attend, we were able to examine how students' observed characteristics are related to their choice of school, among those who were admitted to similar schools. Appendix Table 2 shows the relationship between student characteristics and the average SAT score of a school they choose to attend (conditional on the average SAT score of the most selective school to which they were accepted). For 1976, the coefficient on student SAT score is positive and statistically significant (for both the full sample and the black and Hispanic subgroup). While the relationship between high school GPA and the selectivity of the college attended is close to zero for the full sample, it is

positive for black and Hispanic students. These results suggest that students in the 1976 cohort with better academic credentials tended to matriculate to more selective schools, controlling for the average SAT score of the most selective school they were accepted at.

For the 1989 cohort, there was not a consistent pattern between student characteristics and students' choice of schools. While the relationship between SAT score and school selectivity was positive and statistically significant (within both the full sample and the black and Hispanic subgroup), the relationship between high school GPA and the selectivity of the college attended was negative and statistically significant.

Taken as a whole, these results do not suggest that students with greater potential chose to attend less selective schools; in fact, if anything, results from the 1976 cohort suggest that students with greater potential attend more selective schools. If, among students who were admitted to similar schools, more ambitious students choose to attend more selective schools, then even our selection-adjusted estimates of the effect of college selectivity will be biased upwards.

V. RESULTS

A. Comparison of Earnings Using C&B Survey and SSA Administrative Data

We begin by comparing earnings drawn from the C&B survey to those drawn from SSA administrative data. The C&B survey asked individuals to report their earnings in categories; we assigned those individuals with earnings over 200,000 a topcode of $245,662. (This topcode was set to be equal to the mean log earnings for graduates age 36 to 38 who earned more than $200,000 per year, according to data from the 1990 census (in 1995 dollars)). If we recode the SSA data so that those earning over 200,000 have this same topcode, the correlation for the 1976 cohort

15

between SSA earnings (in 1995) and C&B earnings during the same year is 0.90.[12] This is similar to estimates of the reliability of self-reported earnings data in Angrist and Krueger (1999).

To compare results from this analysis to the results reported in Dale and Krueger (2002), we first estimated a regression where the log of C&B earnings is the outcome measure, but restricted the sample to students in the merged C&B and SSA sample (that is, they matriculated at one of the C&B schools participating in this study, reported that they were working full-time during all of 1995 on the C&B survey, and matched to the SSA data). The coefficient on school SAT score/100 in the basic model using this sample restriction is .068 (.014) (see Table 3, column 3), indicating that attending a school with a 100-point higher SAT score is associated with approximately 7 percent higher earnings later in a student's career. This estimate is similar (though slightly less than) the .076 (.016) estimate for the C&B sample reported in Dale and Krueger (2002) (shown here in column 1).[13] In both samples, the return becomes indistinguishable from zero in the self-revelation model (shown in columns 2 and 4).

Next, we use earnings drawn from the SSA data. In column 5, we use the same sample of full-time workers, but use SSA earnings that were topcoded in the same way that earnings in the C&B survey were topcoded. In column 7, we use SSA earnings and use the same sample of full-time workers, but do not topcode the data. In column 9, we use the median of [log(1993 earnings) through log (1997 earnings)] (in 2007 dollars) as our outcome measure, and restrict the sample to those with non-zero earnings. In column 11, we restrict the sample to those with annual earnings that were greater than a minimum wage threshold (defined as $13,822 in 2007 dollars). In each model, the estimates for the coefficient on school SAT score drawn from our basic model

---

[12] When SSA earnings are not topcoded, the correlation between SSA earnings and C&B earnings falls to 0.67.

[13] Note that the Dale and Krueger (2002) estimate is based on students from 30 C&B schools (all of the C&B schools except for the HBCUs), whereas the estimate in column 3 includes the 27 C&B schools participating in this study.

range from .048 to .061, and are similar to (but somewhat less than) the estimate using earnings from the C&B survey as the outcome measure.

Columns 6, 8, 10, and 12 show results from the self-revelation model for each of these samples. The return to school SAT score in each of these selection-adjusted models is negative and indistinguishable from zero.

In summary, for the 1976 cohort, across a variety of sample restrictions and across both sources of earnings data (C&B survey data and SSA administrative data), the return to school SAT score has a large and positive effect on earnings when we do not adjust for unobserved student characteristics. However, when we adjust for unobserved student characteristics in the self-revelation model, the return falls substantially, becoming indistinguishable from zero.

B. Alternative Selection Controls

We also re-estimated the series of models from Dale and Krueger (2002) that use a variety of selection controls in place of the average SAT score that the student applied to. For example, in one model, we controlled for the highest SAT score of the schools a student was accepted by but did not attend.   In another model, we controlled for the average SAT score of the colleges that rejected the student. Consistent with Dale and Krueger (2002), in each of these models, the return to the school SAT score of the school that the student actually attended was less than the return to the colleges they applied to but (did not attend). For example, for the 1976 cohort, the coefficient on average school SAT score among the schools a student was rejected by was .079 (.022), and from this same model the coefficient on the average SAT score for the school the student attended was .035 (.028).   A full set of these results is available upon request.

The finding that the average SAT score of schools that rejected a student is more than twice as strong a predictor of the student's subsequent earnings as the average SAT score of the school the student attended provides powerful evidence that conventional estimates of the payoff to attending highly selective colleges is biased upward, at least for this sample.

C. Estimated Effect of College Characteristics over the Lifecycle for the 1976 Cohort

To assess the return to school characteristics over the course of student's career for the 1976 cohort, we estimate regressions where the outcome measure was the median of log(annual earnings) for each individual (in 2007 dollars) over a five year interval (1983 through 1987, 1998 through 1992, 1993 through 1997, 1998 through 2002, and 2003 through 2007).   Results are shown in Table 4.

In our basic model with a standard set of regression controls, the return to college SAT score increases over the course of a student's career, from indistinguishable from zero for the earliest period (1983 to 1987, about 3 to 7 years after students likely would have graduated), to over seven percent in for the period of 2003 to 2007 (23 to 27 years after college graduation). However, in our self-revelation models, the estimates are not significantly different from zero for any time period. (To save space, we only report parameter estimates for school characteristics in these tables. In Appendix Table 3, we report a full set of parameter estimates for selected models).

We also estimated regressions separately by gender. In the basic model, the return to college SAT score for men was about 5 percent in 1988 to 1992, and increased over time, reaching a high of nearly 9 percent for the period of 1998 through 2002. For women, the return was consistently lower than the return for men, ranging from 3 percent (in 1988 to 1992) to nearly 6 percent (in 2003 to 2007). The lower return for women does not appear to be due solely to the fact

18

that we cannot identify which women were working full time in SSA's administrative data, as the return for women (5 percent) was smaller than the return for men (7 percent) in the C&B survey when we limited the sample to those who reported working full-time. For both men and women, the coefficient was zero (and sometimes even negative) in the self-revelation model.[14] To increase sample size and improve the precision of our estimates, we focus on results based on the pooled sample of men and women together throughout the rest of the paper.

We estimated these same regressions for two other college characteristics, the Barron's Index and the log of net tuition. The results are summarized in Table 5. In our basic model, the estimated impact of these school characteristics increased over the course of the career, with the coefficient on log tuition reaching a high of 0.14 and the Barron's Index reaching 0.08 at in the last five year interval (last set of rows, Table 5). However, in the self-revelation model the estimates fall substantially and are statistically insignificant at the 0.10 level.[15]

These results are partly a contrast to Dale and Krueger (2002), in that the earlier analysis of self-reported earnings data showed a statistically significant relationship between earnings and the log of net tuition in the self-revelation model, as the coefficient on net tuition was of .058 (.018).

---

[14] This lower return to college selectivity for women is consistent with other literature. Results from Hoekstra (2009), Black and Smith (2004) and Long (2008) all suggest that the effect of college selectivity on earnings is lower for women than for men. Also, while the coefficients for school SAT in the self-revelation model was negative and significant for women in some years (1993 to 1997 and 1998 to 2002), the pattern of results across all of the models we estimated (which included, for example, different measures of college quality measures and different minimum wage thresholds) did not suggest that the return for women was significantly less than zero. For example, the coefficients for the Barron's index for women for women was .051 (.011) in the basic model and .010 (.022) in the self-revelation model in 1993 to 1997; similarly in 1998 through 1992 the coefficient was .050 (.008) in the basic model and -.004 (.027) in the self-revelation model.

[15] We probed the sensitivity of the estimates by including dummy variables for categories (such as "Most Competitive") for the gradations of the Barron's Index. The estimates for the most selective categories were sizeable and significant compared with the base group of the least selective schools in the basic model but were small and statistically insignificant in the self-revelation model.

To attempt to reconcile these results with Dale and Krueger (2002), we re-estimated the effect of net tuition on self-reported earnings for full-time workers from the C&B survey in 1995 using the subset of students from the schools participating in this study, and found that the coefficient (adjusted for clustering) on log (net tuition) from the self-revelation model was somewhat smaller: .041 (.038), and not statistically significant. When we estimated the same regression for the same sample, but used SSA's administrative earnings data in 1995 (instead of self-reported earnings data from the C&B survey), the coefficient (and standard error) on net tuition was even smaller: .033 (.046). Moreover, over the full study period (1983 to 2007) the coefficient on net tuition was generally between 0 and .02 (and never greater than .033) in the self-revelation model based on earnings drawn from SSA administrative data as the outcome measure. Thus, the effect of net tuition based on the single year of self-reported earnings reported in Dale and Krueger (2002) appears to been atypically high relative to the series of estimates we were able to generate using SSA's administrative data, though the large standard errors make it difficult to draw inferences.

C. Estimated Effects of College Characteristics for the 1989 Cohort

Unlike the 1976 cohort where we have data for most of the student's career, we only have a limited number of post-college years for the 1989 cohort. As shown for the 1976 cohort, there is no return to college characteristics in the early part of a student's career, possibly because many graduates from highly selective colleges attend graduate school, and thus forego work experience early in their careers. Therefore, for the 1989 cohort, we focus on the most recent year with earnings data available, 2007, when the students, were, on average, 35 years old. Although the 1989 cohort is too young to assess changes in the return to school selectivity over the student's career, it does allow us to assess whether estimates for the return to school selectivity are similar across cohorts at one point in the lifecycle.

20

In 2007, the coefficient for school SAT score/100 was .056 with a standard error of .014 (or .031 if we adjust for clustering among students who attended the same schools) in the basic model (Table 6). Consistent with the results for the 1976 cohort, the coefficient was indistinguishable from zero (-.008 with a standard error of .019) in the self-revelation model. When we estimated these models by gender, the results are also similar to those of the 1976 cohort: the coefficient for women (.032) was lower than the coefficient for men (.067) in the basic model, and in the self-revelation model, estimates for both men and women are indistinguishable from zero (not shown).

The results for the Barron's index were consistent with the results for school SAT score. Specifically, the return to the Barron's index was nearly 7 percent in the basic model, but was close to zero in the self-revelation model. For net tuition, our estimates from both models were negative and had large standard errors.

D. Estimated Effect of College Characteristics for Racial and Ethnic Minorities

Because some past studies have found that the return to college selectivity varies by race (Behrman et al. 1996, Long 2009, and Loury and Garman 1995), we also examined results separately for racial and ethnic minorities. To increase the sample size, we pooled blacks and Hispanics together, both of which often receive preferential treatment in the college admissions process (Bowen and Bok, 1998). For the 1976 cohort, the return to each proxy for school quality increased over the course of the students' career, and the magnitude of the coefficients did not fall substantially in the self-revelation model. The coefficient was most pronounced for the Barron's index, where going to a school in a higher Barron's category translated to 6.7 percent higher earnings (for the period of 2003 to 2007) in the basic model and 6.2 percent higher earnings in the

21

self-revelation model (not shown). However, the standard errors were large, so the estimate for the Barron's index in the self-revelation model was not statistically significant at the .10 level.[16]

Estimates for the 1989 cohort are shown in Table 7. Parameter estimates for each proxy for school quality are large in the basic model, ranging from 6.3 for the Barron's Index to 17.3 percent for the log of net tuition (Table 7). In the self-revelation model, these estimates remained large, ranging from 4.9 for the Barrons Index to 13.8 for the log of net tuition. While the standard errors are also large, some of the estimates are significantly greater than zero. For example, the coefficient on school SAT score/100 was .076 with a standard error of .032 (or .042 after accounting for clustering of students within schools).

Because the historically black colleges and universities in this sample had lower average SAT scores (and lower Barron's indices and net tuition) than the rest of the institutions in the C&B database, we investigated whether the reason that there was a return to school selectivity for minority students (but not for all students) was due to the greater range in school selectivity observed for minority students.[17] Specifically, we re-estimated the regressions but excluded the HBCUs from the sample. For the 1976 cohort, the return to the Barron's Index fell from 6.2 percent to 1.6 percent (and indistinguishable from zero) when we imposed this sample restriction. However, for the 1989 cohort, the estimates for minority students remained large when we excluded the HBCUs, implying returns of 12 percent for attending a school with 100 point higher SAT score and 14 percent for attending a school in a higher Barron's category, even in the self-revelation model.

E. Estimated Effect of School Average SAT Score by Parental Education

---

[16] Results for the 1976 cohort are available on request.

[17] See Fryer and Greenstone (2007) for estimates of effect of HBCUs on earnings (and other outcome measures).

Finally, we explored whether the effect of college selectivity varied by average years of parental education.[18] The interaction term for school-average SAT and years of parental education was negative for both cohorts, implying a higher payoff to attending a more selective school for students from more disadvantaged family backgrounds (Table 8). For example, in the self-revelation model for the 1989 cohort, our results suggest that attending a college with a 200-point higher SAT score would lead to 5.2 percent higher earnings in 2007 for those with average parental education of 12 years (equivalent to graduating from high school); however, for those whose parents averaged 16 years of education (approximately equivalent to college graduates), there was virtually no return to attending a more selective college. Similar to Dale and Krueger (2002), we also found a negative interaction between predicted parental income and school average SAT score, though for most years, the interaction term was not statistically significant.

VI. CONCLUSION

Consistent with the past literature, we find a positive and significant effect of the return to college selectivity during a student's prime working years in regression models that do not adjust for unobserved student quality for cohorts that entered college in 1976 and 1989 using administrative earnings data from the SSA's Detailed Earnings Records. Based on these same regression specifications, we also find that the return to selectivity increases over the course of a student's career. However, after we adjust for unobserved student characteristics, the return to college selectivity falls dramatically. For the 1976 cohort, the return to school-SAT score for the

---

[18] Average years of parental education was equal to the average of the mother's and father's education; if data were missing for one parent, the average was set equal to the years of education for the parent for whom data were available. The 13 students in the 1989 cohort and 22 students in the 1976 cohort that were missing education data for both parents were excluded from these regressions.

full sample is always indistinguishable from zero when we control for the average SAT score of the colleges that students applied to in order to control for omitted student variables. Similarly, the returns to other college characteristics (the Barron's Index and net tuition) are substantial in the basic model that controls for commonly observed student characteristics but small and never statistically distinguishable from zero in the self-revelation model, which (partially) controls for unobserved student variables.

There were noteworthy exceptions for subgroups. First, for the 1989 cohort, the estimates indicate a positive return to attending a more selective school for black and Hispanic students, which is robust in the selection-adjusted model. Second, our results suggest that students from disadvantaged family backgrounds (in terms of educational attainment) experience a higher return to attending a selective college than those from more advantaged family backgrounds. For example, for the 1989 cohort, our estimates from the selection-adjusted model imply a positive return to attending a more selective college for students who had parents with an average of less than 16 years of schooling; however, the return to attending a more selective college was zero (or even negative) for students whose parents averaged 16 or more years of education. One possible explanation for this pattern is that while most students who apply to selective colleges may be able to rely on their families and friends to provide job-networking opportunities, networking opportunities that become available from attending a selective college may be particularly valuable for black and Hispanic students, and for students from less educated families.

Contrary to expectations, our estimates do not suggest that the return to college selectivity (within the set of C&B schools) increased for students who entered college more recently, as estimates for the 1976 and 1989 cohort are similar when we compare the returns for each cohort at a similar stage relative to college-entry (approximately 18 to 19 years after the students entered

24

college). Specifically, for the 1976 and 1989 cohorts, attending a college with a 100-point higher SAT score lead to students receive about 6 percent higher earnings (in 1995 and 2007 respectively) according to results from the basic model; for both cohorts, this return was close to zero in our selection-adjusted model.

Several caveats should be borne in mind in interpreting our findings. First, the analysis does not pertain to a nationally representative sample of schools, as the sample is derived from the 27 colleges and universities in the C&B dataset, the majority of which are very selective. However, estimates of the return to school selectivity based on the C&B dataset were similar to -- indeed, slightly higher than -- those based on a nationally representative dataset, the National Longitudinal Study (NLS) of the High School Class of 1972 (see Dale and Krueger, 2002). In addition, Dale and Krueger (2002) found an insignificant payoff to attending more selective schools when they used the NLS to estimate the self-revelation model.

Second, the estimates from the selection-adjusted models are imprecise, especially for the 1989 cohort. Thus, even though the point-estimates for the return to school quality are close to zero, the upper-bound of the 95 percent confidence intervals for these estimates are sometimes sizeable. Nonetheless, our results do imply that estimates that do not adjust for unobserved student characteristics are biased upward. Indeed, the finding that the average SAT score of the highest ranked school that *rejected* a student is a much stronger predictor of that student's subsequent earnings than the average SAT score of the school the student actually attended should give pause to those who interpret conventional regression-based estimates of the effect of college characteristics as causal effects of the colleges themselves.

Finally, it is possible that our estimates are affected by students sorting into the colleges they attended from their set of options based on their unobserved earnings potential. About 35 percent of the students in each cohort in our sample did not attend the most selective school to which they were admitted.[19] Our analysis indicates that students who were more likely to attend the most selective school to which they were admitted had observable characteristics that are associated with higher earnings potential. If unobserved characteristics bear a similar relationship to college choice, then our already-small estimates of the payoff from attending a selective college would be biased upward. It is also possible that the benefit in terms of future earnings from attending a selective college varies across students, and that students sort into college based on their perceived benefit. For example, students who expect a lucrative career because they intend to earn an M.B.A. after college may sort into less selective undergraduate colleges. If students sort on the basis of their idiosyncratic return from attending a selective college, then equation (1) cannot be given a causal interpretation. However, if this is the case, then our results suggest that the typical student does not unambiguously benefit from attending the most selective college to which he or she was admitted. Rather, our results would suggest that students need to think carefully about the fit between their abilities and interests, the attributes of the school they attend, and their career aspirations.

---

[19] Hoxby (2009) mistakenly reports that only 10 percent of students in the C&B sample used in Dale and Krueger (2002) did not attend the most selective college to which they were admitted. However, in actuality, 38 percent of the students in the C&B sample used in Dale and Krueger (2002) did not attend the most selective college to which they were admitted (similar to the results for the sample used in this paper).

REFERENCES

American Council on Education, *American Universities and Colleges,* 11th edition (Hawthorn, NY: Walter de Gruyter, 1973).

American Council on Education, *American Universities and Colleges,* 12th edition (Hawthorn, NY: Walter de Gruyter, 1983).

American Colleges and Universities. American Council on Education, *American Universities and Colleges*, 14th edition. Walter de Gruyter, Hawthorn, NY, 1992.

Angrist, Joshua and Alan Krueger. (1999). "Empirical Strategies in Labor Economics," *Handbook of Labor Economics*, vol. 3A, edited by Orley Ashenfelter and David Card, Amsterdam: North Holland, pp. 1277-1366.

Barron's Profiles of American Colleges, 19th Edition. Barron's Educational Series, Inc., Hauppauge, NY,1992.

Behrman, J. R., Rosenzweig, M. R., & Taubman, P. (1996).College choice and wages: estimates using data on female twins. *Review of Economics and Statistics*, *78*(4), 672–685.

Black, Dan A. and Jeffrey A. Smith, "How Robust is the Evidence on the Effects of College Quality? Evidence From Matching." *Journal of Econometrics*, 121, pp. 99 – 124, 2004.

Bowen, William G. and Derek Bok. *The Shape of the River: Long-Term Consequences of Considering Race in College and University Admissions.* Princeton University Press, Princeton, NJ, 1998.

Brewer, D., & Ehrenberg, R. (1996). Does it pay to attend an elite private college? Evidence from the senior class of 1980. *Research in Labor Economics,* 15, 239-272.

Dale, Stacy Berg and Alan B. Krueger, "Estimating the Payoff to Attending a More Selective College: an Application of Selection on Observables and Unobservables. *Quarterly Journal of Economics*, 117(4), pp. 1491–1528, November 2002.

Fox, M.A., Connolly, B.A., and Snyder, T.D. (2005).*Youth Indicators 2005: Trends in the Well-Being of American Youth,* (NCES 2005–050). U.S. Department of Education, National Center for Education Statistics. Washington, DC: U.S. Government Printing Office.

Fryer, Roland G., and Michael Greenstone. 2010. "The Changing Consequences of Attending Historically Black Colleges and Universities." *American Economic Journal: Applied Economics*, 2(1): 116–48.

Hoekstra, Mark. 2009. "The Effect of Attending the Flagship State University on Earnings: A Discontinuity-Based Approach." *The Review of Economics and Statistics,* Volume 91 (4).

Hoxby, Caroline. The Changing Selectivity of American Colleges. *Journal of Economic Perspectives—Volume 23, Number 4—Fall 2009—Pages 95–118*

Katz, Lawrence F. and Kevin M. Murphy. 1992. "Changes in Relative Wages, 1963-87: Supply and Demand Factors." *Quarterly Journal of Economics* 107 (February): 35-78.

Lena Lindahl, Hakan Regner. (2005) College Choice and Subsequent Earnings: Results Using Swedish Sibling Data*. *Scandinavian Journal of Economics* 107:3, 437-457 Online publication date: 1-Oct-2005.

Long, Mark. "College Quality and Early Adult Outcomes." *Economics of Education Review*, 27(5),pp. 588-602, October 2008.

Long, Mark. "Changes in the Returns to Education and College Quality. *Economics of Education Review* (2009). In press.

Loury, L. D., & Garman, D. (1995). College selectivity and earnings. *Journal of Labor Economics*, *13*(2), 289–308.

Monks, James. 2000. "The Returns to Individual and College Characteristics Evidence from the National Longitudinal Survey of Youth." *Economics of Education Review,* 19(3): 279–89.

Saavedra, Juan. 2008. "The Returns to College Quality: A Regression Discontinuity Analysis." Unpublished paper, Harvard University.

TABLE 1

DESCRIPTIVE STATISTICS

	1976 Cohort				1989 Cohort			
	Full Sample		Black and Hispanic		Full Sample		Black and Hispanic	
	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation	Mean	Standard Deviation
Dependent Variables: Earnings Measures, 2007 Dollars								
Log (2007 Earnings)	11.53	1.07	11.29	0.81	11.41	1.21	11.21	0.73
2007 Annual Earnings	183,411	711,803	119,861	242,161	139,698	359,255	98,472	107,852
Median of [Log(1993 Earnings) through Log(1997 Earnings)]	11.30	0.847	10.78	1.02				
Median of [Log(2003 Earnings) through Log(2007 Earnings)]	11.48	1.022	11.00	1.19				
Median of (1993 to 1997 Earnings)	106,638	148,615	70,760	77,178				
Median of (2003 to 2007 Earnings)	164,009	489,647	102,714	211,876				
Regression Control Variables								
Average School SAT Score/100	11.58	1.22	10.95	2.00	12.0	1.4	11.54	1.45
Average SAT Score of Schools Applied to/100	11.40	1.20	10.73	1.83	11.9	1.4	11.33	1.33
Average Student SAT/100	11.61	1.89	9.46	2.15	12.1	2.6	10.24	2.20
Student SAT is missing	0.05	0.24	0.07	0.26	0.00	0.09	0.02	0.16
Log(Net Tuition)	7.66	0.55	7.69	0.47	8.95	0.73	8.97	0.41
Barrons Index	3.34	1.20	3.32	1.19	4.19	1.13	3.74	1.21
Female	0.43	0.56	0.53	0.51	0.45	0.72	0.51	0.50
Black	0.06	0.27	0.88	0.33	0.08	0.40	0.74	0.44
Hispanic	0.01	0.10	0.12	0.33	0.03	0.24	0.26	0.44
Asian	0.02	0.16	0.00	0.00	0.08	0.40	0.00	0.00
Other Race	0.04	0.22	0.00	0.00	0.00	0.10	0.00	0.00
High School GPA, 4-Point Scale	3.57	0.36	3.37	0.44	3.62	0.36	3.49	0.36
High School GPA Missing	0.37	0.54	0.40	0.50	0.61	0.71	0.59	0.49
Predicted Parental Income	9.98	0.39	9.70	0.42	11.05	0.54	10.82	0.45
Student Athlete	0.07	0.29	0.06	0.23	0.08	0.39	0.06	0.23
Student Submitted								
0 additional applications	0.37	0.48	0.38	0.49	0.29	0.45	0.31	0.46
1 additional application	0.22	0.42	0.24	0.44	0.20	0.58	0.21	0.41
2 additional applications	0.21	0.41	0.21	0.41	0.22	0.60	0.22	0.42
3 additional applications	0.15	0.36	0.14	0.35	0.23	0.61	0.21	0.41
4 additional applications	0.04	0.22	0.03	0.17	0.07	0.38	0.05	0.21
Sample Size (Unweighted)	**12,075**		**1,167**		**6,479**		**1,508**	

Source: Data from the College and Beyond Survey and Detailed Earnings Records from the Social Security Administration.

Notes: Means are weighted to make the sample representative of all students in the C&B institutions. Means for the regression control variables for the 1976 cohort reflect the sample used in the 1993 - 1997 earnings regression (reported in Table 3). Means for the 1989 cohort reflect the sample used in the 2007 earnings regression (as reported in Table 6). Earnings means exclude those with annual earnings below the minimum wage threshold (equivalent to $13,822 in 2007 dollars).

DRAFT

TABLE 2

COLLEGE APPLICATION PATTERNS AMONG STUDENTS ATTENDING COLLEGE AND BEYOND SCHOOLS

	1976 Cohort		1989 Cohort	
	Full Sample (Percent)	Blacks and Hispanics (Percent)	Full Sample (Percent)	Blacks and Hispanics (Percent)
Percent of Students Submitting:				
At least one additional application	64.3	64.1	71.3	70.1
Among Those Submitting at Least One Additional Application:				
Applied to a school with a higher average SAT Score than the school attended	53.7	49.0	55.0	48.4
Percent accepted to at least one additional school	87.9	94.0	88.3	92.5
Among Those Accepted to at Least One Additional School: Percent accepted to a school with a higher average SAT score than the one they attended	35.0	40.2	36.1	40.3
Sample Size (Unweighted)	**17,223**	**1,411**	**8,830**	**2,016**

Source: College and Beyond Survey.

Notes: Sample includes all survey respondents from C&B schools participating in this study. Means are weighted to make the sample representative of all students attending the C&B schools participating in this study.

DRAFT

TABLE 3

COMPARISON OF PARAMETER ESTIMATES FOR THE RETURN TO COLLEGE QUALITY ON EARNINGS USING C&B AND SSA DATA, 1976 COHORT

	C&B Sample[a]		Merged C&B and SSA Sample[b]									
	Log(1995 C&B Earnings)		Log(1995 C&B Earnings)		Log(1995 SSA Earnings (topcoded))		Log(1995 SSA Earnings (not topcoded))		Median of [Log(1993 Earnings) through Log(1997 Earnings)], SSA Data		Median of [Log(1993 Earnings) through Log(1997 Earnings)], SSA Data	
	1	2	3	4	5	6	7	8	9	10	11	12
	Basic	Self–Revelation	Basic	Self-Revelation	Basic	Self–Revelation	Basic	Self-Revelation	Basic	Self-Revelation	Basic	Self-Revelation
Parameter Estimate for School SAT/100	0.076	-0.001	0.068	-0.007	0.048	-0.021	0.058	-0.015	0.059	-0.025	0.061	-0.023
	(.008)	(.012)	(.007)	(.012)	(.009)	(.014)	(.009)	(.015)	(.008)	(.012)	(.007)	(.012)
	{.016}	{.018}	{.014}	{.018}	{.016}	{.018}	{.017}	{.016}	{.012}	{.013}	{.013}	{.014}
Sample Size	14,238		10,886		10,886		10,886		11,932		12,075	
Sample Restriction	Full Time Workers (according to responses on C&B survey)		Full Time Workers (according to responses on C&B survey)		Full Time Workers (according to responses on C&B survey)		Full Time Workers (according to responses on C&B survey)		Workers with non-zero median earnings (according to SSA data)		Workers who earned over minimum wage threshold (according to SSA data)	

Source: College and Beyond Survey and SSA's Detailed Earnings Records.

Notes: Estimates are drawn from weighted least squares regressions that control for age, race, student SAT score, a dummy for whether student SAT score was missing, high school grade point average, a dummy for whether high school grade point average is missing, predicted parental income and student athlete. Weights were used to make the sample representative of the population of students at the C&B schools. Each cell corresponds to a different regression. Standard errors are in brackets and are robust to correlated errors among students who attended the same institutions. Minimum wage threshold is equal to $13,822 in 2007 dollars. The top earnings category for the C&B data (greater than $200,000) was assigned a value of $242,662 (see text for details). For comparison purposes, we topcoded the SSA data in the same way as the C&B data were topcoded in columns 5 and 6.

[a]Sample includes survey respondents from the 30 C&B institutions analyzed in Dale and Krueger (2002), and only includes those with non-zero earnings; shaded box on table shows other sample restriction.

[b]Sample includes survey respondents from the 27 C&B institutions participating in this study that were matched to SSA data. Sample only includes those with non-zero earnings; shaded boxes on table show additional sample restrictions for specific regressions.

SSA=Social Security Administration

C&B=College and Beyond

TABLE 4

EFFECT OF SCHOOL SAT SCORE/100 ON LOG(EARNINGS), 1976 COHORT OF MEN AND WOMEN

	Men		Women		Men and Women Pooled	
	Basic	Self-Revelation	Basic	Self-Revelation	Basic	Self-Revelation
Effect on Median of [Log(1983 Earnings) through (1987 Earnings)]						
Parameter Estimate for School SAT/100	0.011	-0.004	-0.007	-0.037	0.004	-0.017
	(.007)	(.011)	(.007)	(.012)	(.005)	(.008)
	{.012}	{.019}	{.012}	{.023}	{.011}	{.014}
N	6,294	6,294	5,690	5,690	11,984	11,984
Effect on Median of [Log(1988 Earnings) through (1992 Earnings)]						
Parameter Estimate for School SAT/100	0.054	-0.001	0.031	-0.034	0.045	-0.014
	(.009)	(.013)	(.009)	(.015)	(.006)	(.010)
	{.014}	{.016}	{.014}	{.019}	{.012}	{.013}
N	6,911	6,911	6,294	6,294	12,407	12,407
Effect on Median of [Log(1993 Earnings) through Log(1997 Earnings)]						
Parameter Estimate for School SAT/100	0.080	0.001	0.034	-0.059	0.061	-0.023
	(.010)	(.016)	(.010)	(.018)	(.007)	(.012)
	{.013}	{.015}	{.012}	{.016}	{.013}	{.014}
N	6,896	6,896	5,179	5,179	12,075	12,075
Effect on Median of [Log(1998 Earnings) through Log(2002 Earnings)]						
Parameter Estimate for School SAT/100	0.087	0.002	0.042	-0.069	0.070	-0.024
	(.012)	(.018)	(.012)	(.020)	(.008)	(.013)
	{.015}	{.022}	{.010}	{.016}	{.012}	{.019}
N	6,869	6,869	5,195	5,195	12,064	12,064
Effect on Median of [Log(2003 Earnings) through Log(2007 Earnings)]						
Parameter Estimate for School SAT/100	0.083	0.006	0.057	-0.035	0.074	-0.008
	(.013)	(.020)	(.012)	(.021)	(.009)	(.014)
	{.015}	{.027}	{.013}	{.018}	{.014}	{.018}
N	6,650	6,650	5,244	5,244	11,894	11,894

Source: College and Beyond Survey and Detailed Earnings Records from the Social Security Administration.

Notes: Parameter estimates drawn from a weighted least squares regression. Each cell represents a separate regression. Both the basic and self-revelation models control for race, gender, predicted parental income, student's SAT score, a dummy indicating if student SAT score was missing, student's high school grade point average, a dummy indicating if high school grade point average was missing, and whether the student was a college athlete; the self-revelation model also controls for the average SAT score of the schools to which the student applied and dummies for the number of applications the student submitted. Two sets of standard errors are reported, one in parentheses and one in brackets. Standard errors in brackets are robust to correlated errors among students who attended the same institution.   Individuals are excluded if the median of annual earnings over the five year interval was less than $13,822 in 2007 dollars.

DRAFT

TABLE 5

EFFECT OF COLLEGE CHARACTERISTICS ON EARNINGS, 1976 COHORT OF MEN AND WOMEN

	College Characteristic: Log (Net Tuition)		College Characteristic: Barron's Index	
	Basic	Self -Revelation	Basic	Self -Revelation
Effect on Log (Median of 1983 to 1987 Annual Earnings)				
Parameter Estimate for	0.014	-0.007	0.010	0.001
School Quality Measure	(.010)	(.013)	(.005)	(.013)
N=11,984	{.024}	{.027}	{.012}	{.015}
Effect on Log (Median of 1988 to 1992 Annual Earnings)				
Parameter Estimate for	0.092	0.012	0.055	0.020
School Quality Measure	(.012)	(.016)	(.006)	(.017)
N=12,407	{.028}	{.028}	{.011}	{.015}
Effect on Log (Median of 1993 to 1997 Annual Earnings)				
Parameter Estimate for	0.124	0.013	0.071	0.017
School Quality Measure	(.015)	(.019)	(.007)	(.010)
N=12,075	{.030}	{.038}	{.009}	{.015}
Effect on Log (Median of 1998 to 2002 Annual Earnings)				
Parameter Estimate for	0.140	0.017	0.077	0.014
School Quality Measure	(.012)	(.017)	(.008)	(.012)
N=12,064	{.026}	{.034}	{.008}	{.019}
Effect on Log (Median of 2003 to 2007 Annual Earnings)				
Parameter Estimate for	0.143	0.026	0.080	0.023
School Quality Measure	(.018)	(.023)	(.009)	(.012)
N=11,894	{.032}	{.039}	{.010}	{.017}

Source: C&B Survey and Detailed Earnings Records from the Social Security Administration.

Notes: Parameter estimates drawn from a weighted least squares regression. Each cell represents a different regression. Both the basic and self-revelation models control for race, sex, predicted parental income, student's SAT score, a dummy indicating if student SAT score was missing, student's high school grade point average, a dummy indicating if high school grade point average was missing, and whether the student was a college athlete; the self-revelation model also controls for the average SAT score of the schools to which the student applied and dummies for the number of applications the student submitted. Weights were used to make the sample representative of students at C&B schools. Two sets of standard errors are reported, one in parentheses and in brackets. Standard errors in brackets are robust to correlated errors among students who attended the same institution. The Barron's measure is coded as a continuous measure. Within our sample, this measure ranged from 2 to 5, where "2" represented "Competitive Colleges" and "5" represented "Most Competitive Colleges." Individuals are excluded if the median of annual earnings over the five year interval was less than $13,822 in 2007 dollars.

DRAFT

TABLE 6

EFFECT OF COLLEGE CHARACTERISTICS ON 2007 EARNINGS, 1989 COHORT OF MEN AND WOMEN

| | College Characteristic | | | | | |
| | School SAT Score/100 | | Log (Net Tuition) | | Barron's Index | |
	Basic	Self - Revelation	Basic	Self - Revelation	Basic	Self - Revelation
Parameter Estimate for	0.056	-0.008	-0.011	-0.108	0.069	-0.002
Effect of Quality Measure	(.014)	(.019)	(.025)	(.028)	(.017)	(.022)
on Log(2007 Earnings)	{.031}	{.034}	{.062}	{.070}	{.038}	{.042}
Sample Size	**6,479**	**6,479**	**6,479**	**6,479**	**6,479**	**6,479**

Source: C&B data and Detailed Earnings Records from the Social Security Administration.

Notes: Parameter estimates drawn from a weighted least squares regression. Each cell represents a different regression. Both the basic and self-revelation models control for race, sex, predicted parental income, student's SAT score, a dummy indicating if student SAT score was missing, student's high school grade point average, a dummy indicating if high school grade point average was missing, and whether the student was a college athlete; the self-revelation model also controls for the average SAT score of the schools to which the student applied and dummies for the number of applications the student submitted. Weights were used to make the sample representative of students at C&B schools. Two sets of standard errors are reported, one in parentheses and in brackets. Standard errors in brackets are robust to correlated errors among students who attended the same institution. The Barron's measure is coded as a continuous measure. Within our sample, this measure ranged from 2 to 5, where "2" represented "Competitive Colleges" and "5" represented "Most Competitive Colleges." Individuals were excluded if their annual earnings was less than $13,822 in 2007 dollars.

DRAFT

TABLE 7

EFFECT OF SCHOOL CHARACTERISTICS ON LOG (2007 EARNINGS)
(Black and Hispanic Students Only, 1989 Cohort)

Dependent Variable	School SAT Score/100		Log(Net Tuition)		Barron's Index	
	Basic	Self - Revelation	Basic	Self - Revelation	Basic	Self - Revelation
All Black and Hispanic Students						
Parameter Estimate for	0.067	0.076	0.173	0.138	0.063	0.049
Effect of Quality Measure	(.019)	(.032)	(.056)	(.071)	(.022)	(.036)
on Log(2007 Earnings)	{.028}	{.042}	{.076}	{ .092}	{.033}	{.046}
Sample Size	**1,508**		**1,508**		**1,508**	
All Black and Hispanic Students, Excluding HBCUs						
Parameter Estimate for	0.122	0.120	0.187	0.116	0.158	0.143
Effect of Quality Measure	(.030)	(.042)	(.064)	(.079)	(.040)	(.053)
on Log(2007 Earnings)	{.035}	{.056}	{.081}	{.101}	{.038}	{.051}
Sample Size	**995**		**995**		**995**	

Source: College and Beyond Survey and Detailed Earnings Records from the Social Security Administration.

Notes: Parameter estimates drawn from weighted least squares regression models. Weights were used to make the sample representative of the population of students at C&B schools. Both the basic and self-revelation models control for race, sex, predicted parental income, student SAT score, a dummy variable indicating if student SAT score was missing, student high school grade point average, a dummy variable indicating if high school grade point average was missing, and whether the student was a college athlete; the self-revelation model also controls for the average SAT score of the schools to which the student applied and dummies for the number of applications the student submitted. Two sets of standard errors are reported, one in parentheses and one in brackets. Standard errors in brackets are robust to correlated errors among students who attended the same institution. The Barron's measure is coded as a continuous measure. Within our sample, this measure ranged from 2 to 5, where "2" represented "Competitive Colleges" and "5" represented "Most Competitive Colleges." Individuals are excluded if they earned less than $13,822 in 2007.

HBCU = historically black college or university.

DRAFT

TABLE 8

PARAMETER ESTIMATES FROM LOG(EARNINGS) REGRESSIONS, ALLOWING THE EFFECT OF AVERAGE
SCHOOL SAT TO VARY BY PARENTAL EDUCATION

	1976 Cohort		1989 Cohort	
	Parameter Estimates			
Variable	Basic	Self - Revelation	Basic	Self - Revelation
School SAT Score/100	0.126	0.041	0.206	0.117
	(.035)	(.036)	(.073)	(.075)
	{.058}	{.055}	{.110}	{.112}
Average Years of Parental Education	0.066	0.063	0.128	0.106
	(.027)	(.027)	(.053)	(.053)
	{.041}	{.038}	{.097}	{.097}
Years of Parental Education*School SAT Score/100	-0.004	-0.004	-0.009	-0.008
	(.002)	(.002)	(.004)	(.004)
	{.0037}	{.003}	{.008}	{.008}
Effect of a 200 Point Increase in School SAT Score If:				
Average Years of Parental Education=12 (equivalent to High School Graduate)	0.148	-0.025	0.193	0.052
Average Years of Parental Education=16 (approximately equivalent to College Graduate)	0.113	-0.060	0.120	-0.009
Average Years of Parental Education=19 (approximately equivalent to Graduate Degree)	0.087	-0.087	0.065	-0.055
Sample Size (Unweighted)	**12,053**	**12,053**	**6,466**	**6,466**

Notes: Parameter estimates drawn from a weighted least squares regression. In addition to the variables listed in the first column of the table, both the basic and self-revelation models control for race, gender, student's SAT score, a dummy indicating if student SAT score was missing, student's high school grade point average, a dummy indicating if high school grade point average was missing, and whether the student was a college athlete; the self-revelation model also controls for the average SAT score of the schools to which the student applied and dummies for the number of applications the student submitted. Two sets of standard errors are reported, one in parentheses and one in brackets. Standard errors in brackets are robust to correlated errors among students who attended the same institution. The dependent variable for the 1989 cohort is the log(2007 earnings), and for the 1976 cohort is the median [log (1993 annual earnings) through log (1997 annual earnings)]. Individuals are excluded if their annual earnings (for the 1989 cohort (or median annual earnings for the 1976 cohort) was less than $13,822 in 2007 dollars. The parental education measure is the average of the mother's and father's education; if data was missing for one parent, the average was set equal to the years of education of the parent with data. Individuals with no parental education data for either parent are excluded from the regression. The average (and standard deviation) of years of parental education was 15.23 (2.70) for the 1976 cohort and 16.18 (3.24) for the 1989 cohort.

DRAFT

DISTRIBUTION OF SCHOOL CHARACTERISTICS FOR COLLEGE AND BEYOND SCHOOL IN SAMPLE

College Characteristic	Number of Schools	Number of Students (Unweighted)	Number of Students (Weighted)
1976 Cohort			
Average SAT Score			
1250 plus	6	2,962	2,962
1175-1250	13	7,999	7,999
<1175	8	6,262	10,219
1976 Net Tuition			
>$3,500	10	5,497	5,497
$2,000-$3,500	12	7,164	7,164
<$2,000	5	4,562	8,519
1978 Barron's Index			
Most Competitive	11	5,503	5,503
Highly Competitive	7	4,833	4,833
Very Competitive or Competitive	9	6,887	10,844
1989 Cohort			
Average SAT Score			
1250 plus	7	3,604	5,944
1175-1250	6	2,857	4,220
<1175	5	2,369	8,488
1990 Net Tuition			
>$12,000	6	2,971	4,116
$10,000-$12,000	7	3,490	6,048
<$10,000	5	2,369	8,488
1992 Barron's Index			
Most Competitive	10	4,931	7,743
Highly Competitive	5	2,651	7,909
Very Competitive or Competitive	3	1,248	3,001

Source: C&B Survey, American Universities and Colleges, 11th, 12th and 14th editions; Barrons Guide.

Notes: The C&B Survey in 1989 included only a subset of schools included in the 1976 Survey. For both cohorts, the sample is weighted to make the sample representative of the population of students attending the C&B institutions. Net tuition for 1970 and 1980 was calculated by subtracting the average aid awarded to undergraduates from the sticker price tuition, as reported in the eleventh and twelfth editions of American Universities and Colleges. Then, the 1976 net tuition was interpolated from the 1970 and 1980 net tuition, assuming an exponential rate of growth. Net tuition for the 1989 cohort was drawn from the 14th Edition of American Universities and Colleges.

DRAFT

APPENDIX TABLE 2

AVERAGE SAT SCORE OF SCHOOL ATTENDED CONDITIONAL ON AVERAGE SAT SCORE OF MOST
SELECTIVE SCHOOL ACCEPTED AT, AMONG STUDENTS
ACCEPTED TO MORE THAN ONE SCHOOL

	1976 Cohort		1989 Cohort	
Variable	Full Sample	Blacks and Hispanics	Full Sample	Blacks and Hispanics
Student SAT Score/100	0.0648	0.122	0.028	0.115
	(.004)	(.021)	(.005)	(.013)
Student SAT missing	-0.074	-0.915	-0.155	-0.592
	(.029)	(.160)	(.103)	(.130)
High School GPA	-0.005	0.229	-0.146	-0.197
	(.020)	(.095)	(.028)	(.062)
High School GPA Missing	0.197	0.173	0.125	0.299
	(.012)	(.070)	(.013)	(.041)
Predicted Parental Income	0.007	-0.132	-0.002	-0.162
	.017)	(.079)	(.018)	(.041)
Female	0.123	0.262	0.070	0.232
	(.011)	(.066)	(.013)	(.036)
Black	-0.193	-0.228	-0.207	-0.172
	(.026)	(.089)	(.025)	(.041)
Hispanic	0.205	(omitted)	0.048	(omitted)
	(.050)		(.035)	
Asian	0.067		0.046	
	(.036)		(.021)	
Other Race	-0.177		0.152	
	(.032)		(.080)	
College Athlete	0.168	0.300	0.108	0.323
	(.021)	(.132)	(.022)	(.069)
Average SAT Score/100 of Most	0.778	0.869	0.838	0.769
Selective School Accepted At	(.007)	(.029)	(.009)	(.020)
Sample Size	**10,068**	**8,335**	**5,642**	**1,275**
R-squared	**0.714**	**0.726**	**0.762**	**0.801**

Source: College and Beyond Survey.

Notes: Parameter estimates drawn from weighted least squares regression models where the dependent variable is the average SAT score of the college attended. Regressions only include students who were accepted to more than one college. Standard errors are in parentheses. Weights were used to make the sample representative of the population of students at C&B schools.

DRAFT

FULL SET OF PARAMETER ESTIMATES FOR SELECTED LOG (EARNINGS) REGRESSIONS
FOR 1976 AND 1989 COHORTS

Variable	1976 Cohort		1989 Cohort	
	Basic	Self-Revelation	Basic	Self-Revelation
School SAT Score/100	0.061	-0.023	0.056	-0.008
	(.013)	(.014)	(.031)	(.034)
Student SAT Score/100	0.022	0.014	0.047	0.033
	(.005)	(.005)	(.008)	(.008)
Student SAT missing	-0.141	-0.122	-0.262	-0.217
	(.030)	(.030)	(.160)	(.160)
Female	-0.479	-0.469	-0.410	-0.412
	(.013)	(.013)	(.020)	(.020)
Black	-0.028	-0.037	0.036	0.022
	(.029)	(.029)	(.040)	(.040)
Hispanic	-0.063	-0.077	-0.060	-0.074
	(.069)	(.069)	(.059)	(.040)
Asian	0.171	0.151	0.154	0.139
	(.046)	(.046)	(.036)	(.036)
Other Race	-0.088	-0.101	-0.363	-0.344
	(.034)	(.034)	(.143)	(.143)
High School GPA	0.218	0.216	0.194	0.188
	(.021)	(.021)	(.042)	(.042)
High School GPA missing	0.015	0.013	0.094	0.092
	(.014)	(.014)	(.021)	(.021)
Predicted Parental Income	0.161	0.140	0.137	0.117
	(.019)	(.017)	(.029)	(.029)
Athlete	0.124	0.123	0.135	0.092
	(.025)	(.037)	(.037)	(.020)
Average SAT Score/100 of Schools Applied to		0.100		0.099
		(.012)		(.014)
One Additional Application		0.062		0.029
		(.017)		(.029)
Two Additional Applications		0.057		0.053
		(.018)		(.028)
Three Additional Applications		0.073		0.084
		(.020)		(.028)
Four Additional Applications		0.085		0.098
		(.034)		(.041)
R-Squared	0.147	0.153	0.122	0.126
Sample Size (Unweighted)	**12,075**		**6,479**	

Source: College and Beyond Survey and Detailed Earnings Records from the Social Security Administration.

Notes: Parameter estimates drawn from weighted least squares regression models where the dependent variable is log(2007 earnings) for the 1989 cohort and the median of [log(2003 earnings) through log(2007 earnings)] for the 1976 cohort. Standard errors are in parentheses, and are robust to correlated errors among students who attended the same institution. Individuals are excluded if annual earnings (for the 1989 cohort) or the median of annual earnings (for the 1976 cohort) was less than $13,822 in 2007 dollars. Weights were used to make the sample representative of the population of students at C&B schools.

DRAFT