**Dismantling the Legacy of Caste:**

**Affirmative Action in Indian Higher Education**

|  |  |  |
| --- | --- | --- |
| **Surendrakumar Bagde** | **Dennis Epple** | **Lowell J. Taylor** |
| Department of Economic Affairs | Carnegie Mellon University | Carnegie Mellon University |
| Ministry of Finance, India |  |  |
|  |  |  |

September, 2011

**Preliminary and Incomplete**

**Please Do Not Quote**

We thank Greg Caetano, Josh Kinsler, Ronni Pavan, and Michael Peress for helpful comments.

**Abstract**

Public policy in modern Indian features affirmative action—policies intended to reduce persistent inequality stemming from a centuries-old caste structure. We study the effects of one such affirmative action program. Specifically, we examine the consequences of an admissions policy to engineering colleges that fixes percentage quotas, common across 214 colleges, for each of six disadvantaged castes. We have obtained access to exceptionally rich data for study of this affirmative action program—data that include the test scores that were used to administer admissions decision rules, as well as detailed independent ability metrics. Our analysis indicates that for targeted students the program has significant and substantial positive effects both on college attendance and first-year academic achievement.

**1. Introduction**

One of the most difficult public policy issues facing any society is how to deal with an historical legacy of discrimination and exclusion based on racial, ethnic, or hereditary categories. One natural response to entrenched inequity is to implement affirmative action policies, which explicitly favor historically disadvantaged groups. The hope is to level the playing field in the short term, and to affect a longer-term transformation whereby society eventually no longer needs affirmative action.[[1]](#footnote-1) Affirmative action policies enacted in almost any situation are bound to generate controversy because they alter the allocation of scarce resources. In the case of higher education, which we study here, the preferential admission status granted one student can result in the exclusion of some other student from a particular college.

In addition, there is controversy about the effectiveness of affirmative action in higher education even for the intended beneficiaries. At issue is the possibility that affirmative action can inadvertently harm targeted students by placing them in academic programs for which they are poorly suited, creating a “mismatch” between the scholastic demands of college and the academic preparation of the student. Students with weak preparation at the pre-college level, so the argument goes, cannot compete in the academically challenging environment of selective institutions; these students would be better off in institutions where their academic preparedness matches the average peer quality. As we discuss below, to date the accumulated empirical evidence informing this debate has been quite modest.

Our paper contributes to the mismatch (or “fit”) literature by examining the college matriculation and academic achievement of students in 214 engineering colleges in a major State in India, where affirmative action based on caste is established policy. For this analysis, we have assembled unique data, which are exceptionally well suited to the study of affirmative action. The data contain extensive information about a large number of colleges, about applicants to those colleges, and about the affirmative action criteria governing the admissions process. Our data include student performance on standardized entrance examinations, as well as prior performance on standardized high-school completion examinations. We know students’ caste, eligibility for preferential admission under a caste-based affirmative action program, and whether the student was admitted through affirmative action. Finally, we have a good measure of progression in college, common across students and colleges, collected after each student’s first year in college.

Because policy establishes the uniform application of affirmative action, our setting and data provide a unique and important feature: We are able to make a precise determination of which students were admitted due to affirmative action policies. Such a determination is often difficult to infer in other settings in which affirmative action has been studied, e.g., in most cases of affirmative action applied in U.S. higher education. The problem is that each university or college has its own admission policy, often based in part on factors that are difficult to quantify, such as student essays and interviews, weight placed on extracurricular activities, recommendation letters, and so forth. To identify beneficiaries of preferential treatment, it is typically necessary to make inferences from information about race or ethnicity, and variables such as prior academic performance and experience of economic hardship.

A second problem is that comparison of academic outcomes across students is often difficult, because subjective grading standards can vary across faculty members, departments, and colleges and universities. Standardized examinations are typically taken only by students seeking admission for advanced degrees.

Our setting provides a way around these two problems for research design. For our study, we use data on admissions to over 210 engineering colleges. The students in all colleges are admitted through a fully transparent common admission system laid out by the state government. Colleges have discretion in the admission for up to 20 percent of their available seats. Admission for the remaining 80 percent of seats is through the common admission system. The affirmative action policies mandate that 50 percent of these seats in each college be set aside for students belonging to six different castes.[[2]](#footnote-2) Moreover, the policy specifies the proportion of those seats that are to be allocated to members of each of those castes. We know for each student his or her caste or social group, the name of the college, and whether the student was admitted under affirmative action. Secondly, most of these colleges are affiliated to one university, which sets a common curriculum for all the disciplines and administers a common examination to all students.[[3]](#footnote-3) This enables us to compare academic performance of all students in these colleges.

Affirmative action can potentially benefit members of disadvantaged castes in two ways. Affirmative action enables some applicants to attend college who could otherwise not have obtained admission to any college, and affirmative action increases priority in selection among college for those who would have attended a college in the absence of affirmative action. We find significant effects of affirmative action on both of these measures. Affirmative action increases college attendance, with effects that are proportionately greatest for the members of the most disadvantaged castes. Similarly, we find that improved priority in college selection improves achievement, again with proportionately greater effects accruing to members of the most disadvantaged castes. Finally, we observe that for members of *all* castes, an increase in priority for college choice results in improved academic achievement in college; there is *no* evidence of mismatch harming intended beneficiaries.

Our paper is organized as follows. In the next section we provide a brief discussion of affirmative action, with particular attention to the Indian case. In Section 3 we describe the data. In Section 4 we formulate our models of attendance and achievement. In Section 5 we present empirical results. Section 6 contains an analysis of the implications of our estimates for affirmative action. Section 7 summarizes our findings and concludes.

**2. Affirmative Action**

Affirmative actionpolicies seek to increase diversity among those selected for productive or developmental opportunities—jobs, slots in school, military positions, government contracts, etc.—often as a means of ameliorating a legacy of discrimination and marginalization by society. Such policies typically entail some form of preferential treatment to a disadvantaged group, identified by gender, religious affiliation, ethnicity, race, and/or caste.

A large literature studies the economic properties of such policies, with an eye toward understanding their impacts in markets, and improving the effectiveness of these policies for accomplishing desired social objectives.[[4]](#footnote-4) Fryer and Loury (2005a), for example, provide an excellent discussion of the trade-offs inherent in affirmative action policies and give reference to earlier literature. Fryer and Loury (2007) study the welfare economics of affirmative action policies, including those that subsidize the skills development of disadvantaged groups. Fang and Moro (2011) give an overview of statistical discrimination and of our current understanding about the role affirmative action plays in markets where statistical discrimination is important.

**2.1. Some Empirical Analysis in the U.S. Case**

Our particular interest here, of course, is the literature that focuses on affirmative action in higher education. A number of papers have provided empirical evidence in this domain, focusing on the impact of affirmative action policies for observed college admission patterns, and then for such subsequent student outcomes as classroom performance, graduation probabilities, and success in the labor market. Much of this research evaluates affirmative action in the U.S. context.

A first observation from that literature is that affirmative action clearly does sway the admission process in the U.S.—increasing by a substantial degree the probability that black students are admitted to elite colleges. For example, in their influential monograph, Bowen and Bok (1998) find that at each 50-point SAT interval the probability of college admission in elite schools is considerably higher for black students than for comparable white students; for some ranges of SAT scores the probability of admission for black students is three times that of white students. In consequence, a race-neutral admission policy would substantially reduce the overall probability of admission for black college applicants to top schools.[[5]](#footnote-5) Long (2004) estimates that accepted minority students in “elite colleges” (colleges in the top decile) would be reduced by 27 percent if preferential admissions for minority students in the U.S. were eliminated. Similarly, Howell (2010) finds advantages in admission for black and Hispanic applicants to 4-year colleges among “most selective colleges” (though no advantages even for “very selective colleges”). She finds that black applicants to the most selective colleges are 23 percentage points more likely than observationally equivalent white applicants to be offered admissions.[[6]](#footnote-6)

As for the impact of affirmative action policies on intended beneficiaries, results vary. Several studies suggest that there is little reason to be concerned about mismatch in U.S. higher education. For instance, Bok and Bowen’s (1998) analysis indicates that within each SAT score interval, graduation rates for black students are positively correlated with the college selectivity. Alon and Tienda (2005) find that students generally benefit from attending more-selective colleges, with minority student gaining more, compared to white students, from attending a “most selective institution.”[[7]](#footnote-7) Similarly, Fischer and Massey (2007) find no adverse impact of affirmative action on GPA or the likelihood a student leaves college by end of the junior year.

On the other hand, Sander (2004) argues that preferences in law school admission harm black students. Affirmative action leads black students into selective schools, he argues, where their grades suffer, leading to poor performance on the bar exam. A sequence of papers—Ayres and Brooks (2005), Ho (2005), Chambers, *et. al* (2005), and Barnes (2007)—evaluates this claim. Rothstein and Yoon (2008) provide a good discussion about the key issue: choices about the construction of counterfactuals matter a great deal here. Under some reasonable choices, the evidence of mismatch disappears. These papers underscore the difficulty of evaluating mismatch in the context of U.S. higher education.

**2.2. How Might Mismatch Occur?**

Before proceeding with our empirical analysis on mismatch, there is value in clarifying how mismatch might occur in the context of education. Consider an environment in which learning takes place according to a production process along the lines of Duflo, Dupas, and Kremer (2011). In particular, let *yi0* be the initial ability level for student *i* and let *yi1* be her subsequent ability if she attends school *j*. We might suppose that

(1) *yi1* = *yi0* + α*j h*(*yj\*- yi0*),

where *h* is a decreasing function in the absolute value of the difference between the student’s initial level of intellectual capital (*yi0*) and the “target teaching level” *yj\** adopted at school *j*. The idea is intuitive: a student will learn most effectively if the curriculum is geared appropriately to her initial ability level. The constant α*j* > 0 varies across schools, reflecting cross-school differences in quality (teacher expertise and dedication, resources, etc.).

Figure 1 shows cases in which students can attend two difference schools: School A, which targets students who are initially high-achieving, and School B, which targets students with lower initial ability levels. The left panel in Figure 1 is a case in which the two schools have similar quality (αA = αB). The right panel illustrates the case in which an “elite college” (A) is more effective than the second school for all levels of initial ability (so αA > αB). In this example School A has particularly high value-added for students with strong preparation, i.e., high initial values of *yi0*.

As for affirmative action, a policy that admits a targeted student to School A always benefits that student in the second scenario (the right panel of Figure 1). In the first scenario (the left panel), though, there is at least a possibility of mismatch—for a low-ability student who receives preferential admission to School A, but who would have been better off attending School B. Importantly, though, the affirmative action program we study simply increases priority in school choice. Thus any student who is admitted to School A via affirmative action can choose instead to attend School B if she prefers. Affirmative action can therefore create mismatch, but only for students who make unwise decisions about which college to attend. This last observation, about the key role of college choice by targeted students, receives careful attention in the recent work of Arcidiacono, Aucejo, Fang, and Spenner (2009). Their analysis leads to the conclusion that when intended beneficiaries of affirmative action make rational choices, they can be made worse off by affirmative action only when colleges offering preferential admissions have private information about the post-enrollment educational process.[[8]](#footnote-8)

Our model is useful for making an additional point. Even when affirmative action benefits a targeted student, by creating the opportunity to attend a better school, this might create a different kind of mismatch. The beneficiary might receive only a small gain from attending his school of choice, while the non-targeted student he displaces suffers a substantial decline in value-added. For example, in the case illustrated in the right-hand panel in Figure 1, if a very low-achieving targeted student moves from College B to College A, this increases educational value-added to him by only a little. But if a high-ability student thereby moves from College A to College B, this causes a substantial loss to that student. In our work below, we do try for an assessment of the impact of affirmative action not only on targeted students but also on students who are not targeted for preferential admissions.

**2.3. The Caste System and Affirmative Action in India**

As we have mentioned, India is an important and potentially fruitful setting for the study of affirmative action in higher education. The Indian caste system divides society into closed hereditary groups (Shah et al., 2006). The numerous castes in India have been rather carefully classified, graded inequality being a fundamental principal of the system.[[9]](#footnote-9) To address the widespread inequality that has resulted from the caste system, Indian law prohibits discrimination based on caste, and provides for affirmative measures for ameliorating the social and economic conditions of disadvantages castes.

The set of castes known as Scheduled Castes (SC) are at the bottom of the caste hierarchy, suffering the most discrimination in terms of social exclusion and restricted access to educational opportunities.[[10]](#footnote-10) In addition, there are tribal communities, known as Scheduled Tribes (ST), who have lifestyle and religious practices quite distinct from mainstream Indian society (Deshpande, 2005). STs often live in remote and inaccessible places, which make access to education difficult. Finally, there are certain communities or castes that are socially and educationally disadvantaged and are officially designated as Backward Classes (BC). This designation is determined mainly by the extent of educational “backwardness,” position within the hierarchy of castes, and the occupations that members of these communities have traditionally pursued. The BCs are further divided into four distinct groups: BC-A, BC-B, BC-C, BC-D depending on vocation, the degree of backwardness, and other factors. Historically, the BCs have also had difficulty accessing educational opportunities. Relatively speaking, SCs/STs are considered more disadvantaged than the BCs.

The Indian constitution, implemented in 1950, mandates affirmative action for SC and ST groups. Affirmative action has subsequently been extended by law to other disadvantaged castes, with variation across the states of India. Affirmative action operates in such matters as employment in the public sector, recruitment into civil services, and, importantly, in education.

A variety of initiatives are intended to improve access to secondary and higher education for disadvantaged groups. For instance, these initiatives include special residential schools, free accommodation, scholarships, and concessions in tuition. The most important aspect of affirmative action programs is the reservation of a certain percent of seats in institutions of higher education for students from these social groups. The extent of this latter degree of affirmative action is substantial (though the Supreme Court of India has ruled that in higher education total allotments to disadvantaged castes cannot exceed 50%). The Indian state we are studying reserves the following percentage of seats in each college: 15 percent for SCs, 6 percent for STs, 7 percent each for BC-As and BC-Ds, 10 percent for BC-Bs, and 1 percent for BC-Cs. Moreover, one third of seats in each discipline is reserved for female students from each social group.

At present there are only a few studies systematically evaluating the impact of affirmative action in Indian higher education. Darity, Deshpande, and Weisskopf (2011) draw parallels between the Indian and U.S. cases, and evaluate how various affirmative action programs might affect admission prospects for potential beneficiaries in the two societies, but do not address the issue of mismatch. [[11]](#footnote-11) Kirpal and Gupta (1999) examine the performance of students from disadvantaged castes in the prestigious Indian Institutes of Technology (IIT), and though their study also does not directly address the question of mismatch, it is worthwhile to note their findings. They document that academic performance was indeed lower for students in the disadvantaged SC/ST groups. Still, graduation rates for these students were found to be over 80 percent, which suggests that these beneficiaries of reserved seats—and the institutions that are admitting them—are achieving important successes. On the other hand, a recent study by Kochar (2010), who analyses data from one elite college, shows that affirmative action increases the variance of ability within classrooms, and suggests that such increased variance is detrimental to learning of students generally.[[12]](#footnote-12)

Perhaps the most directly relevant paper for our study is Bertrand, Hanna, and Mullainathan’s (2010) analysis of affirmative action in engineering colleges in India. They document that preferential admissions policies have a large effect on admission outcomes by caste—a result that we confirm below. They supplement data on entry examinations and admission for engineering colleges with survey data on subsequent labor market outcomes. The authors estimate positive returns to attending an engineering college (which not surprising). Thus, a policy that reserves seats for members of disadvantaged castes does provide a positive benefit to those individuals. However, these gains come at the social cost of displacing upper-caste applicants. Moreover, the financial returns to attending an engineering school are estimated to be higher for upper-caste than lower-caste individuals, which highlights the possibility that the affirmative action program has a high social cost. Having said that, the study is based on very small samples,[[13]](#footnote-13) so the authors caution against “drawing strong conclusions.”

Against this backdrop, our study holds the potential to make a useful contribution. As we have mentioned, our setting gives us a substantial advantage over researchers who have looked at similar questions in the U.S. context. Specifically, we have detailed information on the admission process, which allows us to precisely trace the impact of the policies on admission; we are able with our data to carefully address the selection process into colleges. Also, in contrast to previous work in the U.S. context, we have an outcome variable—first-year performance on a standardized achievement exam—that is common across all students in all colleges under examination. Finally, we have large samples, which allows for precise estimates of the effects under study.

**3. Data**

We have gained exceptional access to data for 214 engineering colleges in one of the major States in India. This State is amongst top five states in India in terms of population (with a population of more than 75 million) and geographic area. Priority for admission to the engineering colleges is based on rank on a common entry examination and upon caste and gender as noted above. In particular, the highest ranked individual in a caste can choose any college. Then the second highest chooses. This process continues, with choice among colleges limited to those having seats available for the individual’s caste and gender at the time that the individual chooses. In addition to their entry examination scores, we have matched individuals to their high school examination records, and hence we have scores they received on the standardized examination taken at the end of high school. In all we have a sample of 84,795 for whom we have details of their high school examination and entry examination.

**3.1. Admission and Matriculation**

The first part of our analysis investigates the effect of affirmative action on college admission and matriculation. Panel A of Table 1 shows summary statistics for the variables we use in our analysis of college admissions, beginning with the numbers and proportions of observations by caste who participate in the college selection process. In viewing those proportions, it is helpful to bear in mind that 20% of seats are allocated by discretion of the college. Of the remaining 80% of seats, half are reserved for members of disadvantaged castes. Hence, 40% of seats are reserved for disadvantaged castes. We have results of two such examinations: the three-hour entry examination for the engineering colleges (the “Entry Exam”), and an intensive, comprehensive three-day examination taken at the end of high school (summarized by the “High School Score”). Given the difference in time devoted to the two different examinations—approximately 42 hours for the high school examination, compared to three hours for the college entry examination—we expect the high school examination to be the much superior measure of students’ academic aptitude.[[14]](#footnote-14) We return to this point in our subsequent analysis of the data.

As is evident from Table 1, there are substantial differences in performance on these examinations among students of different castes. The ordering of average scores on the entry examination conforms to the commonly accepted ordering of the degree of disadvantage of the caste groups, with Scheduled Tribes (ST) being the most disadvantaged and Scheduled Castes (SC) the next most disadvantaged. Note that the degree of disadvantage among the BC castes is not reflected in the alphabetic ordering (i.e., the A, B, C, D designations). Rather, BC-A and BC-C are generally considered more disadvantaged among the BC’s, and the BC-C and BC-D castes less disadvantaged. As expected, students in the Open category have the best average performance on the entry examination. The ordering of performance of different castes on the high school examination is also in conformity with the generally accepted ranking of relative caste disadvantage.

Since priority in choice of college is based on rank on the entry exam, we also report rank on the college entry exam. We convert all test sore measures to rank and normalize them to the unit interval. Rank of 1 is assigned to the best performing student and rank 0 represents the lowest performing student. All students with the same score receive the same rank. For example, suppose there were 100 students. If 2 of those students tied at the highest score, both would receive an initial rank of 100. If there were then two students with the second-highest score, they would both receive an initial rank of 98, and so forth. When all students are thus ranked, the result would then be normalized by division by 100. The ranking thus preserves the ordinal information in the test scores and facilitates interpretation of the coefficients in analyses that follow.

Importantly, the ranking of students for priority in college choice does *not* follow this procedure. There are two reasons: First, the ranking for admission creates a strict ordering by adoption of a tie breaking algorithm when multiple students within a caste receive the same score on the entry examination.[[15]](#footnote-15) As we discuss below, the somewhat idiosyncratic tie-breaking procedure used to determine rank in college choice is helpful for identifying the effects of affirmative action. Second, student priority is by ranking on the entry exam *and* by caste. We thus create a variable that we term “Effective Rank” to measure priority of college access by caste. For a given caste, members are ranked from highest to lowest on the entry examination score and the result is normalized to the unit interval. As the term implies, effective rank measures the priority in which students of different castes get to select into colleges. Students of different castes with the same effective rank have the same priority in selecting colleges (see Appendix A for details). This construction of effective rank permits direct comparison of the coefficients across castes in the analyses that follow. Moreover, the distinction between rank on the entry exam and effective rank proves crucial in identifying the effects of affirmative action. The higher mean value of effective rank for students eligible for affirmative action show that these students have higher priority of access to colleges compared to Open students. Panel A of Table 1 show, as expected, that the ordering of the average value of effective rank is in reverse order of caste disadvantage, with ST students having highest average value and Open students the lowest.

**3.2. Academic Achievement**

The second part of our analysis investigates the effects of affirmative action on academic achievement. Our analysis of achievement focuses on students who matriculated in private colleges that are all affiliated with a single university. These colleges serve 87% of all students who matriculate. The students in these colleges all take a common set of examinations at the end of their first year—exams that are prepared and graded by the university with which the colleges are affiliated. Our achievement outcome measure is the combined test score based on the performance in seven theory subjects that are taken by all student. The students take approximately three-hour comprehensive test in each of these seven theory subjects. The scores on achievement tests taken at the end of the first year of engineering college are available for 45924 students who were admitted and matriculated in these private engineering colleges. We know the engineering college attended by each student. We also know the caste/social group of each student and the whether the student benefitted by affirmative action program at the time of admission.

Panel B of Table 1 provides summary statistics with respect to achievement, i.e., rank on the first-year exam. This panel shows the number of students for whom we have scores and the mean score by caste.

As for the colleges themselves, we can measure “college selectivity/quality” by mean class entry examination score of Open students. By this measure, colleges with higher mean class entry exam were colleges preferred by those with highest priority in the choice of colleges. The selectivity/quality of 214 colleges in our sample shows substantial variation. The mean of our selectivity statistic is 0.68 and the standard deviation is 0.16. The minimum is 0.29 and the maximum is 0.97.

**4. Modeling College Attendance and Academic Achievement**

Affirmative action improves priority in choice college by students from disadvantaged castes. This policy may increase college attendance by members of disadvantaged castes, and subsequently affect academic achievement once intended beneficiaries are in college. In Section 4.1 we develop the model for academic achievement. Our model for the college attendance is developed in Section 4.2.

**4.1. Modeling Academic Achievement**

The quota system for the engineering schools sets aside seats for six disadvantaged castes. The quotas for each caste, ordered by degree of social and economic disadvantage, are as follows:

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Caste Quotas** | | | | | |
| ST | SC | BC-A | BC-C | BC-B | BC-D |
| 6% | 15% | 7% | 1% | 10% | 7% |

As noted in our discussion of Table 1, academic performance at the high school level follows this ordering.

Additional details with respect to the entry examination and caste rankings are needed to set the stage for our model of achievement. On the entry examination, scores take on 160 discrete values. With more than 140,000 students taking the examination, there are necessarily an enormous number of ties. Ties are broken using the following algorithm: For students with the same score on the entry examination, ties are broken based on the score on the mathematics portion of the examination. For students who have the same score on both of those criteria, ties are broken by score on the physics portion of the examination. Finally, for students who have the same scores on all three of these criteria, ties are broken based on the high school examination score. This idiosyncratic tie-breaking procedure is advantageous for our analysis; it serves to distinguish the effect on college achievement of priority in choice of college from the effect of a student’s own ability.

To develop the logic of our approach, it is useful to begin by assuming that all individuals who gain admission attend and complete the first year. In so doing, we temporarily set aside selection effects associated with students’ decisions about whether to accept college admission. The argument we develop below carries over when we generalize the model to allow for potential selection effects.[[16]](#footnote-16)

Define latent aptitude for engineering, , as follows:

(2)

Here *h* denotes high school examination score, *q* denotes the overall score on the engineering entry examination, *qm* denotes the score on the mathematics component of the entry examination, and *qp* denotes the score on the physics portion of the entry examination. Notice that in this definition of latent aptitude, we normalize the high school score to have a coefficient of one. The high school examination scores are based on a much more extensive set of examinations than the entry examination. Hence, we expect the weight on the entry examination score, , to be substantially less than one. Given that the sub-scores *qm* and *qp* are included in *q*, we expect the coefficients of those sub-scores to be much smaller still. Indeed, we enter the mathematics and physics scores as separate components in the definition of latent engineering aptitude only because they are used in breaking ties on the entry examination, and we want to be certain that the rank variable does not impound any vestige of measures of aptitude.

We adopt the following achievement equation:

(3)  for *j* = O, ST, SC, BC-A, BC-B, BC-C, and BC-D,

where *pij*is the performance of individual *i* in caste *j* on the achievement examination at the completion of the first year of college for those who attend college, ’s are coefficients, and are caste fixed effects. The reference category is the Open Caste (O); caste fixed effects included in the equation capture differences in intercepts for the each of the disadvantaged castes relative to the Open Caste. The key parameter from the perspective of affirmative action is the coefficient r. This parameter captures the effect on achievement of priority in *choice* of engineering college, and hence allows us to draw inference on the impact of affirmative action on achievement. We next discuss identification of this key parameter.

Substituting from (2) into (3), we obtain our achievement equation in estimable form:

(4) 

Our latent aptitude measure is an exceptionally comprehensive measure, in that it includes a high school score that is based on 42 hours of examinations, as well as the scores on the components of the engineering college entry examination. It is still possible, however, that the error term may impound unmeasured components of student ability. Fortunately, such unmeasured components of ability do not give rise to bias in the estimation of the effects of caste rank, for the following reason:[[17]](#footnote-17) Conditional on latent aptitude, , the correlation between *rij* and *u,j* is zero. Intuitively, this arises for the following reason. First, any information about ability that is contained in the high school and entry examinations is captured in via the presence of the high school examination, the overall entry examination score, *qij*, and the tie-breaking sub-scores,  and . The effective rank variable introduces no new information about ability. Hence, the presence of unmeasured components of ability introduces no bias in the estimation of r. The same argument applies to estimating the caste fixed effects  if, as we assume, there is no “cultural bias” in the high school and entry examinations.

We expect the high school score to be a strong predictor of achievement in engineering schools. Nonetheless, the preceding discussion implies that the estimate of the effect of affirmative action should not be sensitive to inclusion of the high school examination score (as the only way in which that score impacts rank is via its rare use as the final tie breaker, after the mathematics and physics scores, in converting entry examination scores into rank). Hence, dropping *hij* from the achievement equation should have minimal effect on the estimates of r or the . In our robustness analysis below, we show that the estimates of r and the are not sensitive to the presence or absence of *hij*.

Identification also requires that the effective rank variable not be linearly dependent on the elements of latent aptitude. Affirmative action assures that this condition is met as shown in Figure 3. Consider individuals scoring at the 40th percentile on the entry examination (i.e., entry exam rank of approximately 0.40). A member of the Scheduled Tribes with this entry exam performance would have an effective rank of 0.88, a member of the Scheduled Castes would have effective rank of 0.74, a member of BC-C would have a rank of 0.65, and a member of the Open castes would have an effective rank of 0.34. Thus, affirmative action gives rise to great variation of effective rank for a given level of performance on the entry examination. This provides strong identification of the coefficient r. A second source of identification is the tie-breaking procedure. This introduces an element of randomness in effective rank, which further aids identification. Hence, even within caste there is considerable variation in effective rank for a given entry examination score. In our robustness analysis, reported in the results section, we exploit this variation to investigate whether the coefficient r varies across castes.

**4.2. College Attendance**

An applicant to an engineering college potentially has three choices: attend an engineering college, attend some other academic institution, or choose the no-college option. We observe whether the applicant attends an engineering college. To frame our choice model, let *Uei,j(ai,j,ri,j,cei,j)* denote utility in the engineering college to which student *i* in caste *j* is admitted, where *ai,j* denotes the student’s academic aptitude, *ri,j* denotes the effective rank of the student in his or her caste, and *cei,j*denotes the cost of attending an engineering college.[[18]](#footnote-18) For this discussion, we do not require that *ai,j* be the latent measure in equation (2). We first provide a general characterization of the choice problem of an applicant and then turn to the details of specifying the empirical counterpart. We write utility of the engineering college option as follows with lower-case *u* denoting the deterministic component of utility:

(5) *Uei,j(ai,j, ri,j , cei,j) = ue,j(ai,j, ri,j, cei,j) + ei,j.*

For a large majority of students in our data, an engineering college will be the best available academic option. However, some exceptionally able students can be expected to gain admission to a more prestigious institution such as an IIT. Let *i,j*=1 for a candidate who is admitted to such an institution (and who also prefers that institution to the no-college option), and *i,j*= 0 otherwise. Hence, we write the utility of the non-engineering college option as follows, where *cai,j* is the cost of attending an alternative academic institution, lower-case u denotes a deterministic component of utility and *’i,j* denotes an idiosyncratic utility shock:

(6) *Uni,j(ai,j,cai,j) = i,j(uaj(ai,j, cai,j) - u0j(ai,j)) + u0j(ai,j) + ’i,j.*

Here *uaj(ai,j,cai,j)* and *u0j(ai,j)* denote, respectively, the deterministic component of utility in an alternative academic institution and in the no-college option. Note that effective caste rank is not included in these two utility expressions since effective rank affects priority only for admission to an engineering college.

Let *pj(ai,j) = E(i,j)* be the probability, conditional on aptitude *ai,j,* that an applicant in caste *j* obtains admission to a more elite institution. Then we can write utility of the non-engineering college option as:

(7) *Un,j(ai,j,cai,j) = pj(ai,j)( uaj(ai,j, cai,j) - u0j(ai,j)) + u0j(ai,j) + ni,j ,*

where

(8) *ni,j = ’i,j + (i,j- pj(ai,j))( uaj(ai,j,cai,j) - u0j(ai,j)).*

Thus, *ni,*j impounds both an idiosyncratic preference shock, *’i,j,* as well as idiosyncratic factors that influence whether a member of caste *j* with aptitude *ai,*j is admitted to a superior alternative academic institution.

Now an applicant who gains admission to an engineering college will matriculate if:

(9) *ui,j = (uej(ai,j,cei,j ri,j) - u0j(ai,j)) - pj(ai,j)( ua,j(ai,j ,cai,j) - u0j(ai,j)) > ni,j - ei,j*.

The first term in the middle expression is the difference in the deterministic components of utility between and engineering college and the no-college option. The second term is the probability of admission to an academic institution preferred to an engineering college multiplied by difference in the deterministic components of utility between the alternative academic institution and the no-college option. We assume that the idiosyncratic shocks are i.i.d. normal, implying a probit specification for the binary choice of attending or not attending an engineering college.

For a large majority of students who take the entry examination for engineering colleges, an engineering college will be their best academic option. The probability of admission to an IIT or comparable institution will be an increasing and convex function that is near zero throughout most of its domain and increases sharply for aptitudes in the far right tail of the distribution. Thus, for most applicants, the second term of the middle expression in equation (9) will be approximately zero, implying a choice between an engineering college and the no-college option:

(10) *ui,j ≈ uej(ai,j,cei,j ,ri,j) - u0j(ai,j).*

If the gain from attending an engineering college is greater for more able students, the preceding expression will be increasing in *ai,j*. The gain will also be increasing in *ri,j* because those of higher rank within their caste have higher priority in engineering college choice. Hence, we expect the probability of matriculation in an engineering college to be increasing in *ai,j* until *ai,j* becomes large enough that the probability of admission to IIT begins to rise above zero. High-aptitude students who gain admission to an IIT will generally prefer attending an IIT to attending an engineering college. Hence, for sufficiently high *ai,j*, we expect the second term in the middle expression in equation (9) will come to dominate. At that point, the probability of matriculation at an engineering college will begin to decline as *ai,j* increases. Overall, then, we expect the probability of matriculation at an engineering college to be increasing in *ri,j* and to be an inverted U-shaped function in *ai,j*. We will estimate the admission probability using a flexible functional form in *ai,j* to capture this anticipated U-shaped pattern. Note that we have retained a subscript *j* on all deterministic components of utility, permitting potentially different valuations of the benefits of college by members of different castes.

Let college attendance, *y*, take the value of if a student is admitted and matriculates in an engineering college and 0 otherwise. As explained previously, we have college attendance data for all students. Hence, we can evaluate the effect of affirmative on college attendance. As detailed above, we expect that the probability of matriculation in engineering colleges will be a nonlinear function of an individual’s academic ability. To capture this anticipated nonlinear pattern, we include a polynomial in latent aptitude for engineering, , defined in equation (2). We will permit the coefficients of  to differ between the attendance and achievement equations, since the components of aptitude may differentially effect attendance decisions and performance in engineering colleges. In addition, we include a polynomial in the entry examination scores:

(11) .

Our motivation for including this additional term is the following. The high school examination scores exhibit substantial bunching at the upper end of the distribution as shown in the bottom panel of Figure 2. By contrast, as shown in the bottom panel of Figure 2, the entry examination scores exhibit a great deal of variation at the upper end of the distribution. Thus, the entry examination provides more differentiation among high ability students than does the high school examination. As we demonstrate in the empirical analysis that follows, this is captured quite well by our latent ability measure in equation (2) for explaining performance in engineering colleges. For more prestigious institutions, however, the entry examination score provides an additional source of information for differentiating among applicants. Our empirical analysis shows this to be quite important in the attendance equation.

Let *Iy* be a latent index of propensity to attend an engineering college. We choose a flexible specification with caste fixed effects and high-order polynomials in *r*, , and *s*. As detailed in the empirical section, a specification with a fourth-order polynomial in *r* and third-order polynomials in  and *s* fits matriculation decisions quite well and yields findings that conform with the predictions developed in the derivation of equation (9):

(12) for *j*=O, ST, SC, BC-A, BC-B, BC-C, BC-D.

The open category is the reference category.

As we explained in connection with the achievement model, we have relatively strong priors about the relative magnitudes of the coefficients comprising latent aptitude, , in the achievement equation. For the attendance equation, the polynomials in  and *s* provide a flexible specification for capturing tradeoffs that prospective students make among no-college, engineering college, and other higher education options. Aside from the expectation of an inverted-U relationship of aptitude and attendance, we do not have predictions about the relative magnitudes of the coefficients comprising in  and *s* in the attendance equation.

**4.3. Selection**

In derivation of our achievement equation (4), we argued that the coefficient of effective rank, r, and the caste fixed effects, *j*, are not biased in the presence of unmeasured ability. Hence, from the perspective of studying the effects of affirmative action on achievement, it is not necessary to correct for selection. Nonetheless, as a further investigation of robustness, we estimate the Heckman model taking the attendance model as the selection equation and the achievement model as the outcome equation. Two caveats are in order regarding our application of the Heckman approach. First, our derivation of the attendance equation (12) implies a nonlinear dependence of attendance on ability; this nonlinearity can be expected to extend to any unmeasured ability component. That said, however, we would expect the usual positive selection to prevail for the bulk of the sample, with a favorable draw of unobserved ability increasing the probability of accepting admission to an engineering college. At the very high end of the ability spectrum, a favorable draw of unobserved ability may decrease the probability of attending an engineering school by increasing the likelihood of opting for a more favorable alternative. We expect the proportion of the sample to have the latter type of choice to be small relative to those with the former. Hence, we expect the first caveat will not greatly impact application of the Heckman procedure. The second caveat regards identification. Clearly, exclusion restrictions are desirable—restrictions motivated by a theoretical argument that a variable belongs in the selection equation but not the outcome equation. While we do not have such restrictions, we do have a relatively strong counterpart. As we argued previously, our model implies that matriculation at engineering colleges will be an inverted U-shaped function of student aptitude. By contrast, any plausible characterization of achievement implies a monotone relationship between student aptitude and achievement. Hence, in our framework, the non-monotonic relationship dependence of attendance on  coupled with the presence of *s* in the attendance equation but not the achievement equation provide identification that is similar in spirit to identification by exclusion restrictions (though arguably not as compelling as having an entirely distinct variable appear in the attendance equation but not the achievement equation). In our empirical analysis, we report findings both with and without the Heckman correction to exhibit the robustness of our findings.

**5. Empirical Results**

In this section, we begin by reporting results from estimating the achievement equation. We then turn to attendance.

**5.1. Academic Achievement**

Estimated coefficients are reported in Table 2 along with standard errors clustered by college. Column (1), our preferred model, estimates equation (4). Consider first the coefficients on latent aptitude, reported as the last four coefficients of column (1). Recall, as shown in the table, that the coefficient of the high school examination score, *h*, is normalized to a value of 1 in the definition of latent aptitude. The estimated coefficient of the entry exam score, *q*, is 0.25, conforming to our expectation that the entry examination would have a considerably smaller weight in latent aptitude than the much more comprehensive high school examination. The coefficients of *qm* and *qp* are -0.06 and -0.04 respectively. As expected, these coefficients are modest in magnitude relative to the coefficients of *h* and *q*. Figure 4 plots the relationship of achievement to latent aptitude, revealing that achievement is in increasing convex function of latent aptitude. In interpreting this graph, recall that achievement and each element of latent aptitude are all normalized to have mean zero and variance one. As this graph illustrates, an increase in latent aptitude from 0 to 1.5 yields a predicted increase in achievement of 1 standard deviation. Turning to the coefficients of the caste variables, we see that, the ordering of caste fixed effects follows closely the degree of caste disadvantage. Indeed, with only a modest reduction in magnitude of the coefficient of BC-D, the ordering of the caste fixed effects would follow exactly the ordering of caste disadvantage.

The key coefficient of interest, that on effective rank, is 0.53. Recall that the dependent variable is normalized to have a mean of zero and a standard deviation of one. Hence, for each 0.1 increase in effective rank, the implied increase in achievement is 0.053 standard deviations. As we discuss in more detail below, this is a substantial impact.

The remaining columns of Table 2 provide analyses of robustness of the achievement equation. Column (2) permits the coefficient of effective rank to vary by caste. The coefficients are remarkably similar across castes, with only the coefficient for Scheduled Castes (SC) being noticeably lower than the others. A test of the hypothesis that all these coefficients are equal is not rejected (*p* = 0.05). Hence, we conclude that the benefits of priority in access to college is approximately the same across all castes, including the open group.

In presenting our model of achievement, we argued that unmeasured components of achievement should not bias the estimate of the coefficient of effective rank if those components do not play a role in the calculation of effective rank. Column (3) provides a robustness test with respect to this key feature of the model. Recall that the high school score is based on 42 hours of examinations. This measure plays a minimal role in the determination of effective rank, being the third tie breaker after the math and physics components of the entry examination. Hence, our argument implies that exclusion of this measure should not bias the coefficient of effective rank. Comparing columns (3) and (1), we see that the coefficient of effective rank changes from 0.53 to 0.52. This confirms that the estimate of the effect of affirmative action is robust to the presence of unmeasured components of ability. It is also of interest to note that the magnitude of the caste fixed effects are very little changed between equations (1) and (3). Not surprisingly, the *R2* statistic is markedly higher in column (1) than in column (3). As a result, the coefficient of effective rank and the caste fixed effects are estimated much more precisely in column (1), as evident from the estimated standard errors of the coefficients.

We performed other robustness tests of the achievement equation. In the interest of space, we summarize two of these here without accompanying tables. To investigate the adequacy of the latent achievement measure, we included a quadratic in variable *s*, defined in equation (11) in addition to the variables in column (1). This entails estimation of four additional coefficients. The coefficients of these additional terms were quantitatively small and statistically insignificant, and the coefficient of effective rank changed negligibly, from 0.53 to 0.54. To investigate whether college access priority might have benefit that varies with latent aptitude, we added an interaction of latent aptitude and effective rank to the model in column (1). This coefficient was quantitatively small and statistically insignificant. Hence, there is no evidence that the achievement effect of access priority varies with student aptitude.

Next, we turn to joint estimates of the attendance and achievement equations using the Heckman estimator for selection. Hence, we jointly estimate equations (3) and (12) allowing for correlation of the errors across the equations. This specification thus permits the elements of latent aptitude to have differing importance for college achievement than for decisions about choice between engineering college and other alternatives. The estimated correlation of errors across the attendance and achievement equations is 0.057, and not significantly different from zero. Hence, this estimate suggests that selection on unobserved ability is of limited magnitude. We present the coefficients of the achievement equation from the Heckman model as Column (4) of Table 2 to facilitate comparison to estimates obtained without the correction for selection. By inspection, it is evident that the coefficients in column (4) of Table 2 differ little from those in column (1). This accords with the finding that the estimated correlation of errors across the equations is very small. Note, in particular, the coefficient of effective rank in the achievement equation in column (4) of Table 2 is 0.55, as compared to a value of 0.53 in column (1) of Table 2. This provides still another confirmation of the robustness of the estimated effect of affirmative action on college achievement.

From the perspective of affirmative action, the most important coefficient in the achievement equation is that on effective rank. As we have emphasized above, the coefficient of the effective rank variable measures the effects of priority in college choice, and thereby the effects of affirmative action. Our analysis of the achievement equation indicates that admissions priority has a significant, positive, robust, and quantitatively important effect on student academic achievement in engineering colleges.

**5.2. College Attendance**

We next turn to our model of college attendance—the “selection equation” of our Heckman model. The coefficients are presented in Table 3. The predictive accuracy of the equation is summarized in the following table, which shows that 75 percent of choices are predicted correctly:

|  |  |  |  |
| --- | --- | --- | --- |
| **Actual and Predicted Attendance Rates** | | | |
|  |  | Actual | |
|  |  | No | Yes |
| Predicted | No | 0.410 | 0.120 |
| Yes | 0.130 | 0.340 |

In interpreting the estimates in Table 3, we begin with the caste fixed effects. We anticipate, *ceteris paribus,* that the more disadvantaged castes will be less likely to attend college because they are likely to have more limited family resources than members of less disadvantaged castes. In accord with this expectation, the ordering of the six castes fixed effects reported in the table follows the ordering of degree of caste disadvantage. (Again, the reference category for the caste fixed effects is the Open group, the most advantaged group.) Hence, we would expect the six fixed effects in the table to be negative. Two are positive, however, those for castes BC-B and BC-D. These are the least disadvantaged of the groups benefiting from affirmative action. Moreover, while positive and significant, the coefficients for these two castes are relatively modest in magnitude. Hence, the estimated caste fixed effects they accord reasonably (though not perfectly) well with expectations.

The remaining coefficients are less readily interpreted because of the presence of high order polynomials in the variables and the presence of two latent variables in and *s* in equation (11). Hence, we adopt a two-pronged approach to interpreting these estimates. Here we provide a graphical summary to illustrate the estimated relationships. In the section that follows, we provide a counter-factual analysis to quantify the effects of affirmative action on attendance.

We expect the probability of college attendance to be a monotone increasing function of priority of access to college, i.e., to effective rank. Figure 5 plots the estimated fourth-order polynomial in effective rank. This exhibits the influence of effective rank on the index of college attendance (*Iy*, in equation (12)). It is evident from Figure 5 that effective rank does in fact have the anticipated monotone, positive effect on college attendance.

Figure 6 illustrates the effects captured by the polynomials in  and *s* in equation (12). These plots are obtained as follows. Recall that in estimating the attendance equation, we substitute the expressions for  and *s* from equations (2) and (11) into the following polynomials from equation (12): . The left panel of Figure 6 contains a plot of the preceding as a function of high school score, *h,* with  *q*, *qm*, and *qp* fixed at their mean values (zero). The right panel varies *q* in these polynomials while fixing *h*, *qm*, and *qp* at zero. In both graphs, the horizontal axis is confined to scores between the 2.5th and 97.5th percentiles so that the range of variation in the graphs is representative of the preponderance of the sample.[[19]](#footnote-19)

Turning to interpretation, we see that both panels exhibit the anticipated inverted-U shaped relationship between aptitude and attendance in an engineering college. Recall that all examination scores are normalized to have a mean of zero and a variance of one. Note that the vertical axis of the left panel has a range of 0.6 while the right panel has a range of 2. In both panels, the horizontal axis is expressed in standard deviations.[[20]](#footnote-20) This difference is a consequence of the patterns in Figure 2. The entry exam score provides more discrimination at the high end, and hence is more indicative than the high school score of opportunities for attendance at other higher education institutions.

We have estimated a quite flexible form for the attendance equation in order to capture the anticipated nonlinear dependence of attendance on aptitude. We also included a quite flexible form with respect to rank so as to detect deviation from monotonicity if such were present. The findings from the attendance equation conform well in all with respects with our expectations.

**5.3. Discussion of Empirical Findings**

Our empirical results provide strong evidence that improved priority in college access increases propensity to attend college and academic achievement in college. The caste fixed effects in our estimated achievement model indicate that more disadvantaged castes perform less well than there more advantaged counterparts. This raises the question of whether these fixed effects point to mismatch, with members of disadvantaged castes “getting in over their heads” by choosing colleges that are too difficult for them. We believe this is not the case, for the following three reasons. First, if the fixed effects reflect mismatch, the implication is that the measured aptitude variables do not adequately capture differences in aptitude across students of different castes, hence the fixed impound these difference. If the fixed effects are capturing unmeasured differences in aptitude by caste, we would expect those fixed effects to become more pronounced if we use a lower quality measure of aptitude.[[21]](#footnote-21) We find, however, that removing the most comprehensive aptitude measure, the high school score, results in virtually no change in the caste fixed effects. This can be seen by comparing the caste fixed effects in columns (1) and (3) of Table 2. Second, if the more disadvantaged castes are using their improved priority to get into colleges for which they are not adequately prepared, we would expect the coefficient of effective rank to be significantly lower for members of the more disadvantaged castes than for the advantaged castes. In column (2) of Table 2, we saw that the coefficients of the interactions of effective rank and caste where not significantly different from each other. Moreover, with the exception of SC, the coefficient of effective rank is quantitatively very similar across all the castes. Third, if the disadvantaged castes choose colleges that are too demanding relative to their ability, we would expect the interaction of effective rank with latent aptitude to be significant, and we would expect including this interaction to result in a decrease in magnitude of the fixed effects for the disadvantaged castes. In fact, the interaction of effective rank with latent aptitude has a quantitative small (0.07), statistically insignificant coefficient, and inclusion of this interaction leads to negligible change in any of the caste fixed effects (none changes in magnitude by more than 0.01).

If the significant negative fixed effects for castes do not reflect mismatch, the question remains as to the reason for these statistically significant fixed effects. At this point, we can only provide speculation. It seems that disadvantages of being in a lower caste extend to college performance, even conditional of measured ability upon college entry. Perhaps faculty in the engineering colleges, either advertently or inadvertently, give less time and attention to members of disadvantaged castes. Or perhaps there are social or psychological elements at work; members of the disadvantaged castes socialize less well or view themselves as inferior in status to members of the advantaged castes (and then, for example, are more reluctant to ask questions and seek help when it is needed). We stress that we have no evidence that either for either of these potential explanations, but neither can we at present rule them out.

**6. Effects of Affirmative Action**

Our analysis of the effects of affirmative action proceeds as follows. We predict achievement and attendance with observed effective rank of each student (i.e., with affirmative action). We then repeat the calculations taking the effective rank of each individual as the rank on the entry exam (i.e., without affirmative action). For attendance, an individual is coded as attending if the predicted probability of attendance is greater than one half and not attending otherwise. For these calculations, we use the Heckman selection model. As we have seen above, the selection correction has minimal effects. Hence, the calculations would yield essentially the same results if we were to use the separately estimated achievement and attendance equations.

**6.1. Effect of Affirmative Action on Achievement**

Our results in Table 4 provide estimates of the mean effects of affirmative action on achievement for each caste. For example, for individuals in the Scheduled Tribes, the median increase in effective rank due to those who attend college under affirmative action is 0.45.[[22]](#footnote-22) Hence, multiplying by the coefficient of effective rank in the achievement equation, we obtain a predicted mean gain in achievement for members of the Scheduled Tribes of 0.23 standard deviations. We see gains for all members of the disadvantaged castes. The gains for the most disadvantaged castes are quantitatively large relative to gains normally found in educational interventions. The members of the open caste lose. While the mean loss is lower than for the advantaged castes, but of course these losses are being spread over a relatively large number of individuals.

**6.2. Effect of Affirmative Action on College Attendance**

Using our college attendance model we predict attendance for each student with and without Affirmative Action. Table 5 reports predicted changes in attendance for each caste. As this table shows, affirmative action increases attendance by the disadvantaged castes while decreasing attendance by the open caste. In Table 6, we provide additional detail. The table shows mean differences in the predicted attendance rates caste and by quintiles on the entry examination. This table exhibits several important features. First, the mean difference in predicted attendance rate varies by the caste rank. All of the disadvantaged castes benefit from affirmative action, as evidenced by the positive mean difference in attendance for all disadvantaged castes for all percentiles of entry exam rank. Moreover, the magnitudes of the effects are monotone in the degree of caste disadvantage, with the most disadvantaged castes exhibiting the greatest increases and the least disadvantaged the smallest increases. Second, the mean differences are negative for the open caste, showing the adverse impact of affirmative action on the members of the open castes. Third, the effects of affirmative action are greatest for individuals in middle quintile of entry exam rank, with magnitude of the effects declining as one moves from the middle to either the higher or lower quintile ranges.

Overall, then, Tables 5 and 6 show that affirmative action increases attendance of all disadvantaged castes, with effects most pronounced near the middle of the range of entry exam rank. For the open castes, affirmative action decreases attendance, with effects most pronounced near the middle of the entry exam range.

**6.3. Discussion of Counterfactual Analysis**

The counterfactual analysis assumes that the college choice and achievement models would remain unchanged with an alternative system for determining effective rank for college admissions. This is a strong assumption. An alternative ranking system will change the allocation of students across colleges. This reallocation of students can be expected to change peer characteristics within colleges. If peer effects operate within colleges, this may lead to further impacts on student achievement. It seems likely, however, that our counterfactual analysis will understate the adverse effect on the disadvantaged castes of eliminating affirmative action. If affirmative action is eliminated, members of disadvantaged castes will have lower priority in college choice. This will tend to concentrate lower-performing, disadvantaged students in a subset of colleges that higher-ranked students perceive as less attractive. Thus, members of disadvantaged castes who continue to attend college in the absence of affirmative action will almost surely attend colleges with weaker peers than they would attend with affirmative action. Such colleges would also then be less likely to be able to attract the more able among college faculty. The above logic suggesting that our counterfactual analysis understates the benefits of affirmative action for disadvantaged castes, also implies that we likely understate the adverse effects of affirmative action on advantaged castes. It is possible, however, that concentrating weaker students in the same colleges can permit better curriculum targeting (Duflo, Dupas, and Kremer (2010)).teachers. The preceding discussion makes clear that the counterfactual calculations, while informative, should be viewed as suggestive with respect to assessing the magnitude of the effects of affirmative.

**7. Conclusions**

We conclude that the exceptional database we have assembled, and our estimated achievement and attendance models, provide compelling evidence that affirmative action in Indian engineering colleges increases college attendance and achievement of members of disadvantaged castes.

**References**

Akerlof, George (1976), “The Economics of Caste and of the Rat Race and Other Woeful Tales,” *Quarterly Journal of Economics,* 90(4), 599-617.

Alon, S. and M. Tienda (2005), “Assessing the “Mismatch” Hypothesis: Differences in College Graduation Rates by Institutional Selectivity,” *Sociology of Education*, 78, 294-315.

Anderson, Siwan (2011), “Caste as an Impediment to Trade,” *American Economic Journal: Applied Economics,* 3(1), 239-263.

Arcidiacono, Peter (2005), “Affirmative Action in Higher Education: How do Admission and Financial Aid Rules Affect Future Earnings?” *Econometrica*, 73(5), 1477–1524.

# [Arcidiacono](http://www.nber.org/people/peter_arcidiacono) Peter, [Esteban M. Aucejo](http://www.nber.org/people/esteban_aucejo), [Hanming Fang](http://www.nber.org/people/hanming_fang), and [Kenneth I. Spenner](http://www.nber.org/people/kenneth_spenner) (2009), “Does Affirmative Action Lead to Mismatch? A New Test and Evidence,” NBER Working Paper.

Arrow, K. (1973), “The Theory of Discrimination,” In Ashenfelter, O. (Ed.), *Discrimination in Labor*

*Markets*. Princeton University Press.

Ayres, Ian and Richard Brooks (2005), “Does Affirmative Action Reduce the Number of Black Lawyers?” *Stanford Law Review,* 57(6), 1807-1854.

Barnes, Katherine Y. (2007), “Is Affirmative Action Responsible for the Achievement Gap Between Black and White Law Students,” *Northwestern University Law Review,* 101(4), 1759-1808.

Becker, G. (1957), *The Economics of Discrimination*. Chicago University Press.

Banerjee, A., E. Duflo, M. Ghatak, and J. Lafortune (2009), “Marry for What? Caste and Mate Selection in Modern India,” [*NBER Working Paper*,](http://ideas.repec.org/s/nbr/nberwo.html) 14958.

Bertrand, Marianne & Hanna, Rema & Mullainathan, Sendhil (2010), “[Affirmative Action in Education: Evidence from Engineering College Admissions in India](http://ideas.repec.org/a/eee/pubeco/v94y2010i1-2p16-29.html),” [*Journal of Public Economics*](http://ideas.repec.org/s/eee/pubeco.html), Vol. 94(1-2), 16-29.

Betts, J., and D. Morell (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*, 34(2), 268-293.

Bourguignon F., Fournier M. and Gurgand M. (2007), “Selection Bias Corrections Based on the Multinomial Logit Model: Monte-Carlo Comparisons,” *Journal of Economic Surveys*, 21(1).

Bowen, W.G. and D. Bok (2000), *The Shape of the River*. New Jersey: Princeton University Press.

Card, D. and Krueger, A. (2004). “Would the Elimination of Affirmative Action Affect Highly Qualified Minority Applicants? Evidence from California and Texas,” *Industrial and Labor Relation Review,* 58(3), 416-434.

Caucutt, E. (2002), “Educational Policy When There Are Peer Group Effects: Size Matters,” *International Economic Review*, 43, 195–222.

Chambers, David L., Timothy T. Clydesdale, William C. Kidder, and Richard O. Lempert (2005), “The Real Impact of Eliminating Affirmative Action in American Law Schools: An Empirical Critique of Richard Sanders’ Study,” *Stanford Law Review,* 57(6), 1855-1898.

Chalam, K. (2007), “Caste-Based Reservations and Human Development in India,” *Sage Publications*, New Delhi, India.

Chan, J. and Eyster, E. (2003), “Does Banning Affirmative Action Harm College Quality?” *American*

*Economic Review*, 93(3), 858–873.

Chung, K. (2000), “Role Models and Arguments for Affirmative Action,” *American Economic Review*, 90, 640–648.

Coate, S. and Loury, G. (1993), “Will Affirmative Action Policies Eliminate Negative Stereotypes?” *American Economic Review*, 83, 1220–1240.

Dale, S., and A. Krueger (2002), “Estimating the Payoff to Attending a More Selective College: an Application of Selection on Observables and Unobservables,” *Quarterly Journal of Economics*, 1491-1527.

Darity, William, Ashwini Deshpande, and thomas Weisskopf (2011), “Who Is Eligible? Should Affirmative Action be Group- or Class-Based?” *American Journal of Economics and Sociology,* 70(1), 238-268.

Deshpande, Ashwini (2001), “Caste at Birth? Redefining disparity in India,” *Review of Development Economics*, 5(1), 130-144.

Deshpande, A. (2005), “Affirmative action in India and the United States,” Background Paper, *World Development Report*.

Duflo, Esther, Pascaline Dupas, and Michael Kremer (2010), “Peer Effects, Teacher Incentives, and the Impact of Tracking: Evidence from a Randomized Evaluation in Kenya,” working paper, MIT.

Epple, D., Romano, R., and Sieg, H. (2002), “On the Demographic Composition of Colleges and

Universities in Market Equilibrium,” *American Economic Review: Papers and Proceedings*, 92 (2),

310–314.

\_\_\_\_\_\_\_\_\_ (2006), “Admission, Tuition, and Financial Aid Policies in the Market for Higher Education,” *Econometrica*, 74(4), 885-928.

\_\_\_\_\_\_\_\_\_ (2008), “Diversity and Affirmative Action in Higher Education,” *Journal of Public Economic Theory*, 10(4), 475-501.

Fernandez, R. and Rogerson, R. (2001), “Sorting and Long Run Inequality,” *Quarterly Journal of*

*Economics*, 116, 1305–41.

Fischer, M., and D. Massey (2007), “The Effects of Affirmative Action in Higher Education,” *Social Science Research,* 36, 531-549.

Fryer, R. and Loury, G. (2005), “Affirmative Action and its Mythology,” *Journal of Economic Perspectives,* 19(3), 147-162.

Fryer, R., G. Loury, and T. Yuret (2007), “An Economic Analysis of Color-Blind Affirmative Action”, *Journal of Law, Economics, and Organization*, 24(2), 319-355.

Ghure, G.S. (2000), “The Caste and Race in India,” *Popular Publication*, Mumbai.

Hanushek, E. (2006), “School Resources,” in *Handbook of the Economics of Education, Volume 2,* Elsevier B.V*.*

Ho, Daniel (2005), “Why Affirmative Action Does Not Cause Black Students to Fail the Bar,” *Yale Law Review,* 114(8), 1997-2004.

Howell, J. (2010), “Assessing the Impact of Eliminating Affirmative Action in Higher Education,” *Journal of Labor Economics*, 28(1), 113-166.

Kane, T. (1998), “Do Test Scores Matter? Racial and Ethnic Preferences,” in Jencks, C. and Phillip, M.

(eds.), *The Black-White Test Score Gap*. Brookings Institution Press.

Kirpal, V., and M. Gupta (1999), “Equality Through Reservations,” *Rawat Publications*, Jaipur.

Kochar, Anjini (2010), “Affirmative Action through Quotas: The Effect on Learning in India,” working paper, Stanford Center for International Development.

Kumar, D. (1992), “The Affirmative Action Debate in India,” *Asian Survey*, 32, 290-302.

Lee, L.F. (1983), “Generalized Econometric Models with Selectivity,” *Econometrica*, 51, 501-512.

Long, M. (2004), “Race and College Admissions: An Alternative to Affirmative Action,” *The Review of Economics and Statistics*, 86(4), 1020-1033.

Lundberg, S. (1991), “The Enforcement of Equal Opportunity Laws Under Imperfect Information: Affirmative Action and Alternatives,” *Quarterly Journal of Economics*, 106(1), 309–26.

Manski, C. (1991), “Educational Choice (Vouchers) and Social Mobility,” *Economics of Education Review*, 11(4), 351–369.

McFadden, D.L. and J. A. Dubin (1984), “An Econometric Analysis of Residential Electric Appliance Holdings and Consumption,” *Econometrica*, 52, 345-362.

Moro, A. and Norman, P. (2003), “Affirmative Action in a Competitive Economy,” *Journal of Public*

*Economics*, 87, 567–94.

\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_ (2004), “A General Equilibrium Model of Statistical Discrimination,” *Journal of Economic Theory*, 114, 1–30.

Munshi, Kaivan and Mark Rosenzweig (2006), “Traditional Institutions Meet the Modern World: Caste, Gender, and Schooling Choice in a Globalizing Economy,” *American Economic Review,* 96(4), 1225-1252.

Muralidharan, Karthik and Venkatesh Sundararaman (2011), “Teacher Performance Pay: Experimental Evidence from India,” *Journal of Political Economy,* 119(1), 39-77.

Puhani, P. (2000), “The Heckman Correction for Sample Selection and Its Critique,” *Journal of Economic Surveys*, 14(1), 53-68.

Rothstein, Jesse and Albert Yoon (2008), “Affirmative Action in Law School Admissions: What Do Racial Preferences Do?” *University of Chicago Law Review,* 75(2), 649-714.

Sander, Richard H. (2004), “A Systemic Analysis of Affirmative Action in American Law Schools,” *Stanford Law Rev*iew, 367.

Shah, G., H. Mander, S. Thorat, S. Deshpande, and A. Baviskar (2006), “Untouchability in Rural India,” *Sage Publications*, New Delhi, India.

Sowell, T. (2004), *Affirmative Action Around the World: An Empirical Study*. Yale University Press.

Srinivas, M.N. (2009), “Social Change in Modern India,” *Orient Black Swan*, New Delhi.

Weisskopf, T. (2004), “Impact of Reservation on Admissions to Higher Education in India,” *Economic and Political Weekly*.

Zwart, F. (2000), “The Logic of Affirmative Action: Caste, Class and Quotas in India,” *Acta Sociologica*, 43, 235-249.

**Appendix A: Ranking Within Caste and Effective Rank**

The rank of a student in the entry examination depends on his/her score in the entry examination, i.e., caste of a student is not a relevant factor for determining the rank in the entry examination (caste of a student is not taken into account while ranking students in the entry examination). However, due to affirmative action, a certain percent of seats are set aside or reserved for caste students. Therefore, the students in different caste essentially constitute separate groups as far as priority in which students get to select into colleges is concerned. For example, if top three SC students have a rank of 135, 700, and 789. Then these students would be first three to be called for filling seats reserved for SC students. (They would still be top three even if their ranks were instead 120, 379, and 679.) Thus, in filling the quota for each caste group what matters is rank within that caste group and not their ranking in the entry examination. Therefore we rank individuals in each caste separately and the numerical ranking a student in a caste receives is based on all members of the caste who took the entry examination.

The capacity in each discipline in each college is divided amongst seven category/caste groups (Open, SC, ST, BC-A, BC-B, BC-C, BC-D) in proportion to their allocated quota. Thus value of a given ranking within a caste depends on the proportion of seats allocated to the caste. For example, a ranking of 50 for a caste assigned 500 seats has the same value as a ranking of 30 for a caste assigned 300 seats. The ranking within caste will not account for this feature of the affirmative action. We solve this problem by dividing rank within the caste by the seat share for that caste, which we call *effective rank*. Let be ranking of an individual in caste *c*. Let be allocated seat share of caste. Then effective rank is defined as



For example, the rank within caste is divided by 0.15 for SC students, as 15 percent seats are reserved for these students. As explained in the admission section, due to specific regulation of the state government, the realized share of quota for some caste exceeds their allocated quota. We divide ranking within caste by the realized quota *except for SC and ST*. In the top-ranked colleges we observe the seat share of SC and ST being equal to their allocated share and therefore if we divide ranking within caste by the realized quota for SC and ST, then this will have effect of reducing their seat share below the allocated quota in the top-ranked colleges. *The effective rank measures the priority in which students of different caste get to select into colleges. This construction of effective rank permits direct comparison of the coefficients across castes.*

To get a sense of how affirmative action affects priority in admission, notice from the table below that the 2nd ranked SC student ranked 383rd on the entry exam, but due to affirmative action was treated roughly the same as an open-category student ranked 13th.

|  |  |  |
| --- | --- | --- |
| Table A1. Ranking of Top SC Students | | |
| Rank in Entry Examination | Ranking within the Scheduled Caste | Equivalent Rank in Open Category |
| 78 | 1 | 6.7 |
| 383 | 2 | 13.3 |
| 512 | 3 | 20.0 |
| 542 | 4 | 26.7 |
| 554 | 5 | 33.3 |
| 587 | 6 | 40.0 |
| 638 | 7 | 46.7 |

**Appendix B. Unbiased Estimation of the Coefficient of Effective Rank**

Our achievement equation is:

B1. 

The goal is to obtain an unbiased estimate of .and the caste fixed effects . The purpose of this appendix is to demonstrate that least squares estimation of equation (B1) yields a consistent estimate of this coefficient even if there are unmeasured components of ability in .

Let  be the rank of student , on the entry examination,the seat share of caste *j,* and let  be the effective rank of the student. Function  makes explicit the dependence of caste rank on individual rank and caste share. Hence, the achievement equation is rewritten:

B2. 

The unmeasured components of ability are impounded in , creating a correlation of , with *h*, *q*, *qm*, and/or *qp*. and thereby with the latent aptitude variable. This can be expressed as:

B3.  

Here  has mean zero, uncorrelated with , and, and independent of,. The key point here is independence of  from . This follows because  contains *no* information about ability beyond that impounded in , and. Hence, conditional on the latter,  is independent of . By hypothesis there are no caste-specific test effects (i.e. no “cultural bias” in the high school and entry exams). Hence,  is independent of the caste fixed effects and of .

Substituting (B3) into (B2), we have:

B4. 

Independence of  from  and  implies that  is independent of . Thus, it follows from (B4) that least squares estimation of (B1) yields an unbiased estimators of  and the caste fixed effects . From (B4) it is also evident that the least squares estimators of the intercept, , and in (B1) may be biased. However, those coefficients are not the focus of the analysis.

Figure 1. Impact of School Choice on

Improvement in Ability (Increase from *y0* to *y1*)

School A

School B

School A

School B

*y0*

*y0*

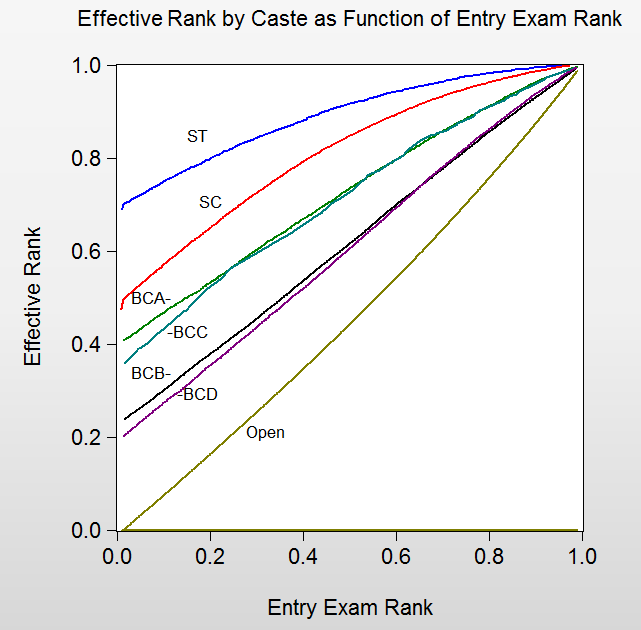
*y1*

*y1*

In the situation depicted on the left, high-ability students (high values of *y0*)are best served by School A and low-ability students by School B. In the case on the right, value added is higher in School A for all students.



Figure 3.









|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 1. Summary Statistics** | | | | | | | | | | |
|  | Full Sample | | | SC | ST | BC-A | BC-B | BC-C | BC-D | Open |
|  | | | | | | | | | |  |
| **A. Variables Used to Study Attendance and Matriculation** | | | | | | | | | |  |
|  | | | | | | | | | |  |
| Observations  % of Obs. in Caste | | 131,290  100 | 13,165 10.0 | | 3137 2.4 | 7484 5.7 | 18,863 14.4 | 1055 0.8 | 14,318 10.9 | 73,268  55.8 |
|  | |  |  | |  |  |  |  |  |  |
| Mean Entry Exam Score | | 60.3 | 51.3 | | 50.5 | 58.2 | 60.5 | 57.2 | 59.7 | 63.3 |
| (Standard Deviation) | | (18.3) | (13.0) | | (12.0) | (15.9) | (17.5) | (15.9) | (16.6) | (19.3) |
|  | |  |  | |  |  |  |  |  |  |
| Mean High School Score | | 747.4 | 680.5 | | 668.2 | 728.7 | 751.0 | 711 | 742.9 | 763.4 |
| (Standard Deviation) | | (155) | (158) | | (153) | (155) | (152) | (162) | (154) | (151) |
|  | |  |  | |  |  |  |  |  |  |
| High School Rank  (Standard Deviation) | | 0.75 (0.21) | 0.66  (0.23) | | 0.64 (0.23) | 0.72 (0.22) | 0.75 (0.20) | 0.70 (0.23) | 0.74 (0.21) | 0.77 (0.20) |
|  | |  |  | |  |  |  |  |  |  |
| Entry Exam Rank  (Standard Deviation) | | 0.50 (0.29) | 0.33 (0.26) | | 0.32 (0.26) | 0.47 (0.27) | 0.50 (0.28) | 0.45 (0.28) | 0.49 (0.28) | 0.54 (0.29) |
|  | |  |  | |  |  |  |  |  |  |
| Effective (Caste) Rank  (Standard Deviation) | | 0.57 (0.27) | 0.71 (0.16) | | 0.83 (0.09) | 0.70 (0.17) | 0.62 (0.22) | 0.68 (0.19) | 0.60 (0.23) | 0.50 (0.29) |
|  | |  |  | |  |  |  |  |  |  |
| Number Attending | | 57,899 | 6898 | | 1514 | 3937 | 9772 | 543 | 7378 | 27,857 |
|  | |  |  | |  |  |  |  |  |  |
| **B. Variables Used to Study Achievement** | | | | | | | | | |  |
|  | |  |  | |  |  |  |  |  |  |
| Observations | | 45,924 | 5031 | | 979 | 2975 | 7796 | 411 | 6016 | 22,716 |
|  | |  |  | |  |  |  |  |  |  |
| Rank of First-Year Score | | 0.64 (0.25) | 0.48 (0.22) | | 0.45 (0.20) | 0.60 (0.23) | 0.63 (0.24) | 0.58 (0.25) | 0.62 (0.24) | 0.71 (0.25) |
|  | | | | | | | | | | |
| *Notes:* The High School Score is available for only 125,472 of our observations. | | | | | | | | | | |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 2. Regression Results: Academic Achievement** | | | | | | | | | | | | | | | |
|  | Non-linear Least Squares | | | | | | | | | | | | Heckman Model | | |
| Independent Variables | (1) | | | (2) | | | | (3) | | | | | (4) | | |
|  | Coef. | (s.e.) | | Coef. | | (s.e.) | | Coef. | | (s.e.) | | | Coef. | (s.e.) | |
| Constant | -0.69 | (0.05) | | -0.71 | | (0.05) | | -0.46 | | (0.09) | | | -0.68 | (0.07) | |
| ST | -0.44 | (0.04) | | -0.46 | | (0.21) | | -0.45 | | (0.08) | | | -0.36 | (0.03) | |
| SC | -0.28 | (0.03) | | -0.12 | | (0.07) | | -0.30 | | (0.06) | | | -0.23 | (0.01) | |
| BC-A | -0.19 | (0.02) | | -0.18 | | (0.08) | | -0.20 | | (0.04) | | | -0.18 | (0.01) | |
| BC-B | -0.10 | (0.01) | | -0.08 | | (0.04) | | -0.10 | | (0.02) | | | -0.09 | (0.01) | |
| BC-C | -0.12 | (0.04) | | -0.17 | | (0.21) | | -0.14 | | (0.06) | | | -0.10 | (0.03) | |
| BC-D | -0.14 | (0.01) | | -0.12 | | (0.05) | | -0.15 | | (0.02) | | | -0.13 | (0.01) | |
| Effective rank | 0.53 | (0.08) | |  | |  | | 0.52 | | (0.15) | | | 0.55 | (0.08) | |
| Effective rank\*OPEN |  |  | | 0.56 | | (0.09) | |  | |  | | |  |  | |
| Effective rank\*ST |  |  | | 0.58 | | (0.25) | |  | |  | | |  |  | |
| Effective rank\*SC |  |  | | 0.35 | | (0.09) | |  | |  | | |  |  | |
| Effective rank\*BC-A |  |  | | 0.55 | | (0.12) | |  | |  | | |  |  | |
| Effective rank\*BC-B |  |  | | 0.54 | | (0.08) | |  | |  | | |  |  | |
| Effective rank\*BC-C |  |  | | 0.64 | | (0.26) | |  | |  | | |  |  | |
| Effective rank\*BC-D |  |  | | 0.54 | | (0.09) | |  | |  | | |  |  | |
| Latent aptitude | 0.64 | (0.01) | | 0.64 | | (0.01) | | 0.92 | | (0.05) | | | 0.58 | (0.01) | |
| Square of Latent aptitude | 0.13 | (0.01) | | 0.13 | | (0.01) | | -0.16 | | (0.01) | | | 0.16 | (0.00) | |
| Latent Aptitude Components: |  |  | |  | |  | |  | |  | | |  |  | |
| High School Score | 1.00 | -- | | 1.00 | | -- | | 0.00 | | -- | | | 1.00 | -- | |
| Entry Score Total | 0.25 | (0.03) | | 0.25 | | (0.03) | | 1.00 | | -- | | | 0.26 | (0.02) | |
| Entry Score Math | -0.06 | (0.01) | | -0.06 | | (0.02) | | -0.16 | | (0.02) | | | -0.07 | (0.01) | |
| Entry Score Physics | -0.04 | (0.01) | | -0.04 | | (0.01) | | -0.17 | | (0.01) | | | -0.05 | (0.01) | |
|  |  | | |  | | | |  | | | | |  | | |
| R-Squared | 0.57 | | | 0.57 | | | | 0.45 | | | | |  | | |
|  |  | | |  | | | |  | | | | |  | | |
| *Notes:* *n =* 41,451. Dependent variable is the score on the first year exam. Standard errors, reported in parentheses, are clustered by college in regressions (1), (2), and (3), but not in (4). There are 117, 101 observation in the selection equation for model (4). | | | | | | | | | | | | | | | |
|  | | | | |  | |  | |  | |  |  | |  |  | |
|  | | | | |  | |  | |  | |  |  | |  |  | |
|  | | |  | |  | |  | |  | |  |  | |  |  | |

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Table 3. Regression Results: Attendance** | | | | | |
|  | Heckman Model | | Probit | | |
| Independent Variables | Coef. | (s.e.) | Coef. | | (s.e.) |
| Constant | -1.62 | (0.044) | -1.62 | (0.043) | |
| ST | -0.89 | (0.031) | -0.89 | (0.031) | |
| SC | -0.35 | (0.017) | -0.35 | (0.018) | |
| BC-A | -0.25 | (0.020) | -0.25 | (0.019) | |
| BC-B | 0.061 | (0.013) | 0.061 | (0.013) | |
| BC-C | -0.16 | (0.046) | -0.15 | (0.047) | |
| BC-D | 0.11 | (0.014) | 0.11 | (0.014) | |
| Effective Rank | -0.80 | (0.482) | -0.76 | (0.473) | |
| Square of Effective Rank | 12.93 | (1.747) | 12.72 | (1.722) | |
| Cube of Effective Rank | -14.8 | (2.466) | -14.5 | (2.437) | |
| Fourth Pwr of Effective Rank | 6.02 | (1.188) | 5.85 | (1.176) | |
| Latent Aptitude | -0.051 | (0.008) | -0.050 | (0.009) | |
| Square of Latent Aptitude | -0.22 | (0.009) | -0.22 | (0.009) | |
| Cube of Latent Aptitude | -0.027 | (0.003) | -0.028 | (0.004) | |
| Latent Score | -0.16 | (0.015) | -0.16 | (0.015) | |
| Square of Latent Score | -0.22 | (0.024) | -0.22 | (0.024) | |
| Cube of Latent Score | 0.005 | (0.004) | 0.005 | (0.004) | |
| Latent Aptitude Components: |  |  |  |  | |
| High School Score | 1.00 | -- | 1.00 | -- | |
| Entry Score Total | -0.37 | (0.062) | -0.38 | (0.063) | |
| Entry Score Math | -0.12 | (0.055) | -0.12 | (0.056) | |
| Entry Score Physics | -0.014 | (0.020) | -0.014 | (0.020) | |
| Latent Score Components: |  |  |  |  | |
| Entry Score Total | 1.00 | -- | 1.00 | -- | |
| Entry Score Math | -0.71 | (0.019) | -0.71 | (0.020) | |
| Entry Score Physics | 0.30 | (0.032) | 0.30 | (0.032) | |
|  |  |  |  |  | |
| Correlation Across Equations | 0.057 | (0.054) | 0.00 | -- | |
|  |  |  |  | |  |
| McFadden R-squared |  |  | 0.267 | |  |
| *Notes:* *n* = 117,101. The dependent variable is 1 for attend (53,730 cases) and 0 not attend (63,371 cases). | | | | | |

|  |  |  |
| --- | --- | --- |
| **Table 4. Effects of Affirmative Action (AA) on Achievement** | | |
| Caste | Mean Change in Effective  Rank Due to AA | Mean Predicted Change in Achievement From AA |
|  |  |  |
| Scheduled Tribes | 0.41 | 0.246 |
| Scheduled Castes | 0.29 | 0.171 |
| BC-A | 0.16 | 0.096 |
| BC-B | 0.05 | 0.028 |
| BC-C | 0.15 | 0.092 |
| BC-D | 0.04 | 0.023 |
| Open | -0.10 | -0.060 |

|  |  |  |  |
| --- | --- | --- | --- |
| **Table 5. Changes in Attendance from Affirmative Action** | | | |
|  | (1) | (2) | (3) |
| Attend With Affirmative Action? | Yes | Yes | No |
| Attend Without Affirmative Action? | Yes | No | Yes |
| Scheduled Tribes | 67 | 1333 | 0 |
| Scheduled Castes | 1707 | 4484 | 0 |
| BC-A | 2202 | 1754 | 0 |
| BC-B | 8543 | 1165 | 0 |
| BC-C | 310 | 231 | 0 |
| BC-D | 6603 | 650 | 0 |
| Open | 26,707 | 0 | 9229 |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| **Table 6. Mean Difference (MD) in Predicted**  **Attendance: Rates By Caste and Entry Exam Rank** | | | | | | | | | | | | | | |
|  | | | | | | | | | | | | | | |
|  | ST | | SC | | BC-A | | BC-B | | BC-C | | BC-D | | Open | |
| Rank (R) on Entry Exam | MD | *n* | MD | *n* | MD | *n* | MD | *n* | MD | *n* | MD | *n* | MD | *n* |
| 0 < R < .2 | 0.32 | 1002 | 0.24 | 3722 | 0.17 | 1262 | 0.05 | 2872 | 0.15 | 203 | 0.03 | 2266 | -0.04 | 9,804 |
| .2 < R < .4 | 0.47 | 639 | 0.43 | 2606 | 0.26 | 1490 | 0.08 | 3368 | 0.26 | 172 | 0.06 | 2532 | -0.12 | 11,686 |
| .4 < R < .6 | 0.50 | 523 | 0.43 | 2049 | 0.26 | 1457 | 0.08 | 3729 | 0.25 | 210 | 0.06 | 2851 | -0.19 | 12,930 |
| .6 < R < .8 | 0.38 | 335 | 0.29 | 1478 | 0.17 | 1356 | 0.05 | 3684 | 0.17 | 170 | 0.05 | 2866 | -0.14 | 14,728 |
| .8 < R < 1 | 0.18 | 147 | 0.12 | 813 | 0.07 | 1094 | 0.02 | 3431 | 0.06 | 141 | 0.02 | 2411 | -0.05 | 17,074 |
| Total *n* |  | 2,646 |  | 10,668 |  | 6,659 |  | 17,084 |  | 896 |  | 12,926 |  | 66,222 |

*Notes:* MD is the mean difference in predicted attendance due to the affirmative action policy.

1. As Harry Blackmun famously suggested, in his legal opinion on affirmative action in higher education in the U.S., “I yield to no one in my earnest hope that the time will come when an affirmative action program is unnecessary and is, in truth, only a relic of the past” (*Regents of the University of California vs. Bakke,* 1978). [↑](#footnote-ref-1)
2. As discussed below, the Indian constitution, implemented in 1950, mandates affirmative action for Scheduled Castes (historically “untouchables”) subjected to the most severe economic and social discrimination, and Scheduled Tribes, isolated both geographically and culturally. Affirmative action has subsequently been extended by law to other disadvantaged castes, with variation across the states of India. [↑](#footnote-ref-2)
3. A small number of colleges admit students under the common admission system, but are not affiliated to this university. As detailed below, we include these students in our analysis of the effects of affirmative action on admission, but we do not include these students in our analysis of the effects of affirmative action on achievement. [↑](#footnote-ref-3)
4. The economics of affirmative action is developed in such work as Lundberg and Startz (1983), Coate and Loury (1993), Moro and Norman (2003), and Fryer and Loury (2005b). [↑](#footnote-ref-4)
5. Bowen and Bok (1998) note that their claim is consistent with the observed outcome for the University of California, Berkeley, when it switched from a race-sensitive to race-neutral admission policy, as that policy change reduced admission rates for black applicants from 49 percent to 16 percent. See Card and Krueger (2005) for more evidence along these lines. [↑](#footnote-ref-5)
6. Estimates from her structural model indicate that a nationwide ban on affirmative action would have only a modest impact on admissions for blacks and Hispanics in colleges generally, but reduce their admission by approximately 10% in the most selective colleges. [↑](#footnote-ref-6)
7. A notable feature of their analysis is an effort to treat bias in estimates that might arise due to the decision process of the students themselves into particular institutions. [↑](#footnote-ref-7)
8. Arcidiacono, *et. al* examine evidence from one institution, Duke University, and find that Duke does indeed have relevant information that the student does not. While this finding cannot be used to establish that mismatch occurs at Duke, it does indicate that the possibility of mismatch exists. [↑](#footnote-ref-8)
9. Economists have studied a variety of aspects related to the caste system. Most famously, Akerlof (1976) provides thoughts about how a rigid caste structure could persist as a long-run equilibrium. Among many other examples, Munshi and Rosenzweig (2006) study how the caste system shapes career choices for Indian men and women, and Anderson (2011) studies the possibility that the caste structure leads to a breakdown in the trade involving water rights at the village level. Each of these papers provides a valuable discussion of the nature and economic consequences of the caste system. [↑](#footnote-ref-9)
10. In fact, Scheduled Castes (untouchables) are outside the traditional four-fold division of Hindu society. [↑](#footnote-ref-10)
11. These authorsalso provide reference to earlier relevant literature. For a nice example of research that gives a sense of challenges in the Indian educational system at the pre-college level, see Muralidhara and Sundararaman (2011). [↑](#footnote-ref-11)
12. Kochar (2010) emphasizes that the results are based on one highly-selective college only. As she says, “The extent to which they would generalize to data on students from the middle of the ability distribution of all students is an open question.” [↑](#footnote-ref-12)
13. The baseline labor market regressions have 273 upper-caste individuals and 380 lower-caste individuals. [↑](#footnote-ref-13)
14. For the “high school score,” we use combined test scores of 11th and 12th grade, for which students take three hour long examinations for each of the 12 papers and practical examinations in two or three subjects. These examinations are spread over about two weeks. [↑](#footnote-ref-14)
15. On the entry examination, scores take on 160 discrete values. With tens of thousands of students taking the examination, there are necessarily an enormous number of ties. Ties are broken using the following algorithm. For students with the same score on the entry examination, ties are broken based on the score on the mathematics portion of the examination. For students who have the same score on both of those criteria, ties are broken by score on the physics portion of the examination. Finally, for students who have the same scores on all three of these criteria, ties are broken based on the high school examination score. [↑](#footnote-ref-15)
16. Also, as we will see below, estimates that correct for selection yield very similar results to those that do not. [↑](#footnote-ref-16)
17. This is developed formally in Appendix B. [↑](#footnote-ref-17)
18. While tuition is the same for all students in all castes, the opportunity cost of resources for college attendance and the psychic costs may create variation across individuals in the utility cost of college attendance. [↑](#footnote-ref-18)
19. The shape of the graph is not sensitive to this restriction, only the range of the values. [↑](#footnote-ref-19)
20. The mean and standard deviation of scores is based on all applicants who took the engineering examination. Those with scores below \_\_ were deemed not qualified for college and were not given an entry rank and hence are not in the sample used to estimate the attendance equation. This and our exclusion of the lowest 2.5% of observations from the plot explain the truncation of the engineering score at -1 on the right panel of Figure 6. [↑](#footnote-ref-20)
21. This can be established formally in appendix B by allowing to be correlated with caste. [↑](#footnote-ref-21)
22. We report in the text mean changes in effective rank for members of Scheduled Tribes who attend college (.45) rather than for all members of Scheduled Tribes in the sample (.50). We do the same for all disadvantaged castes. For members of the Open caste, we use the mean change in effective rank is for those predicted by our model to attend college if there were no affirmative action, though this is virtually the same as the mean change for those who attended college in the presence of affirmative action. [↑](#footnote-ref-22)