Measuring College Quality and its Determinants: Evidence from Indian Higher Education

Advait R. Aiyer*
Gaurav Khanna†
Naveen Kumar ‡
Karthik Muralidharan §

October 12, 2024

Abstract

We develop an objective measure of college quality and decompose this effect into its constituent parts. We study the engineering college market in a Indian state that uses a centralized admissions process to match students to programs. Notably, this state mandates a uniform entrance examination and standardized tests for all students statewide each semester, allowing for a unique opportunity to compare cross-college performance. Using a regression discontinuity design, we estimate program-level college outcomes in the neighborhood of program admission cutoffs. We find that marginally admitted students who clear a cutoff experience an increase in peer quality, and a decrease in relative percentile rank upon admission. Over time, these marginally admitted students demonstrate, on average, improvements in academic performance and an increased likelihood of graduation. To objectively gauge college quality, we construct individual program value-added measures. Subsequently, we decompose program value-added into distinct components: college quality, peer quality, and relative rank effects. A unique feature of our study setting is that affirmative action policies generate multiple cutoffs within a single program. We leverage this setting to delineate relative rank effects and isolate this aspect from the value-added originating from college inputs and peer quality. Understanding the contributions of college inputs, peer quality, and relative positions within the classroom to overall value-added is crucial for both academic researchers and policymakers. Such insights are essential for informed decision-making regarding investment strategies and resource allocation within the higher education sector.

JEL: I21, I23, I24, I28
Keywords: College Quality, Value-Added, Rank Effects, Peer Quality, College Inputs

*The Wharton School of the University of Pennsylvania - advaitra@wharton.upenn.edu, Personal Website
†University of California, San Diego - gakhanna@ucsd.edu, Personal Website
‡BGS College of Engineering & Technology, India - nkumar1@bgscet.ac.in, Personal Website
§University of California, San Diego - kamurali@ucsd.edu, Personal Website

1 Introduction

Access to a selective college can expose individuals to high-quality college inputs, high-achieving peers, and better teachers. However, the literature evaluating the impact of selective colleges on marginally admitted students report mixed impacts on academic achievement (Saavedra, 2008; Rubinstein and Sekhri, 2013; Bagde et al., 2016; Francis-Tan and Tannuri-Pianto, 2018; Sekhri, 2020; Dasgupta et al., 2022).[1] A leading explanation for the mixed results is the extent of the trade-off between positive effects of higher-ability peer environments (Jain and Kapoor, 2015; Feld and Zölitz, 2017) and negative impacts resulting from marginally admitted students ranking low among peers (Elsner and Isphording, 2017, 2018; Murphy and Weinhardt, 2020; Dasgupta et al., 2022; Fabregas, 2023).[2] As such, measuring the academic returns to college quality, and decomposing the determinants of college "value-add" is essential.

The objective of our paper is twofold: (i) to determine the effect of going to a preferred program on marginally admitted students, and (ii) to decompose the effect into peer quality, college inputs, and relative rank effects. There is a large literature examining the benefits of having high-performing peers (Sacerdote, 2001; Whitmore, 2005; Carrell et al., 2009; Black et al., 2013; Booij et al., 2017), and the effect of students' ordinal rank within their classroom (Bertoni et al., 2018; Murphy and Weinhardt, 2020; Elsner et al., 2021; Delaney and Devereux, 2022; Denning et al., 2023). The uniqueness of our setting, with multiple admissions cutoffs *within* a program, allows us to separate college "value-add" into effects of relative rank, peer quality, and college inputs.

We access administrative data on all 60,000 applicants appearing for engineering colleges' common entrance test in a Indian state. The common entrance test data gives us students' entrance exam scores, 12^{th}-grade state-standardized exam score, and several observable characteristics such as caste, gender, and age. Students are assigned a state rank using the composite score of the 12^{th}-grade exam scores and the common entrance exam score. The state rank is revealed to students.

Once state ranks are published, students are asked to submit rank-ordered preferences for majors (such as computer science, electronics & communication, and information science engineering) and colleges affiliated to the state university.[3] The state admits students through multiple types of seat reservations for affirmative action based on caste (SC, ST, BC-A, BC-B, BC-C, BC-D, BC-E) and gender (M, F). The declared caste and gender category is used to auto-determine the affirmative

[1] Wage returns for marginally admitted students at a selective or elite college is positive (Zimmerman, 2014; Kirkeboen et al., 2016; Sekhri, 2020)

[2] Research has looked at the negative impact of ranking low on non-cognitive outcomes such as major choice, risk-taking, self-belief, competitiveness, over-confidence, and Big Five personality traits.

[3] We refer to a college and major combination as a program in this article. For example, Computer Science Engineering (CSE) major in College A is a program).

action category of a student. Those not seeking and also seeking admission under reservation categories compete on merit for the seats available in the open category (OC) or general merit category.

A deferred acceptance mechanism, with serial dictatorship, is used to match students to programs in the order of their ranks (Baswana et al., 2019; Gale and Shapley, 1962). In our setting, students are matched to programs in three rounds. Post each round of matching, students are informed of their allocation and they are asked to either accept their allocation or indicate their desire to wait for allocation in the next round. The process is repeated for round 2, and in the final round, students either accept their allocation or opt out of the admission system. The uniqueness of this admission process, combined with seat reservations, results in 16 cutoffs (8 caste categories x 2 gender categories) for each program. In our data, we observe the program allocation for each round and the reservation category under which the seat is allotted.

The state university has 202 affiliated colleges admitting students through the common entrance exam. The university sets the curriculum to be taught in 8 semesters, conducts state-standardized exams at the end of each semester, and grants degrees at the end of the 4-year program. Of the 202 colleges, 192 are private, and 10 are government colleges. We have access to the eight semesters' subject-wise scores, and pass-fail status in each subject for all students.

In our analysis, we first identify the aggregate impact of gaining admission into a preferred program for marginally admitted students using a regression discontinuity design. We estimate the Local Average Treatment Effect (LATE) value-added for each program. Then, guided by the literature on estimating value-added (Chetty et al., 2014; Koedel and Rockoff, 2015), we determine the OLS value-add for each program to compare to the LATE value-added. Finally, we leverage the multiple cutoffs within a program to uniquely estimate a model that decomposes the value added into the effect of college inputs, peer quality, and relative rank.

Several interesting findings emerge from our pooled regression results. As expected, marginally admitted students who clear a cutoff experience a 0.4 standard deviations (σ) increase in peer quality and a decrease in relative percentile rank by 15 percentage points. Turning to academic outcomes, barely getting admitted to a preferred program increases the standardized semester exam total score by 0.117 σ and increases the probability of graduation on time by 3 percentage points. To put this in the context of the literature's mixed findings, our positive results suggest that, on average, the positive impact stemming from college inputs and peer quality is more than the potential negative impact of low-relative ranking.

Moving from pooled regression results, we estimate program-specific LATEs. To draw insights from program-specific value-adds, we rank all programs using the cutoff score from the common entrance test. For example, programs that had the highest cutoff score will be ranked as the most desirable program or are more preferred by students. Looking at program-specific LATEs based on desirability showcases two patterns: (i) 9 out of the top 25 desirable programs do not have positive value-add, whereas 7 out of the bottom 25 desirable programs do not have a positive value add; (ii) only one of top 25 programs has a value-add of 0.4 σ, whereas 11 of the bottom 25 programs have a value-add of more than 0.5 σ.

Compared to the traditional RDD setting where one cutoff per program is used to estimate LATEs, we use up to 16 cutoffs per program and hence up to 16 LATEs to determine the weighted-average LATE of each program. Therefore, our program-specific LATE estimates should be more comparable to OLS estimates. To investigate this further, we determine program-specific value-add using the OLS specification that controls for observables and ability using past exam scores. In general, we observe that OLS and LATE value-add estimates differ considerably. While numerous very desirable programs have a negative LATE estimate, their OLS estimates are large positives. On the other hand, some least desirable programs with high LATE estimates have low OLS estimates. We infer that OLS value-added estimates are often biased by the quality of the incoming students and may not accurately reflect the value-added by the program itself.

Finally, we turn to decomposing the program value-add into relative rank effects, peer effects, and college quality. Now, consider a program with two admission cutoffs, one for the students targeted by affirmative action (AA) policies, and another for students getting admitted under open competitions (OC). We separately identify the expected payoff to clearing an OC cutoff and AA cutoff by taking the difference in outcomes of students around the cutoffs using RDD. Taking the second difference of the expected payoffs of OC and AA gives us relative rank value-add if we assume that the college inputs and peer quality experienced by OC and AA "cutoff-misers" is the same. Since we have already estimated the value-add of each program (program RD VA), we difference out the relative rank value-add (relative rank VA) from program RD VA to determine the combined contribution of college inputs and peer quality, which we refer to as the quality value-add (quality VA).[4]

As such, we can determine program RD VA, relative rank VA, and quality VA. We plot the distributions of all three measures for all programs, and we make the following three observations: (i) we find that program RD VA is made up of two potentially opposing effects (relative rank vs college quality) as one might theorize; (ii) the modal value of the relative rank VA distribution is negative suggesting that on average, the marginally admitted student experiences a negative effect from being the "lowest scoring" student relative to their peers; (iii) the quality VA distribution

[4] We assume that academic outcome of a student is an additive separable function of college inputs, peer quality, relative percentile rank, their ability, and idiosyncratic variation.

has a positive modal value suggesting that once the relative rank VA is separated from overall program RD VA, college inputs and exposure to a higher peer quality has a positive effect on the marginally admitted student.

We make two important contributions. Our first contribution is to the literature studying the effects of going to a selective college. We extend the literature by developing an objective measure of college quality (program value-add) and decomposing this effect into its constituent parts. Understanding the contribution of college inputs, peer quality, and relative position within the classroom to overall value-add is crucial for both academic researchers and policymakers. Such insights are essential to policymakers for informed decision-making regarding investment strategies and resource allocation within the high-education sector.

Methodologically, we speak to the literature on Regression Discontinuity Designs with multiple cutoffs widely used in political science and economics of education (Cattaneo et al., 2021; Bertanha, 2020; Cattaneo et al., 2021). Our setting enables us to exploit the within-classroom variation around multiple cutoff thresholds that exist in a single classroom (Kumar, 2023). This is rare in the school and college quality literature. We advance this nascent literature by developing a conceptual framework to showcase how to fully exploit all the information available in a multi-cutoff RD setup in the economics of education.

The rest of the paper is organized as follows. The data, empirical strategy, validity of RDD, and findings from pooled regression results are discussed in Section 2. The conceptual framework in Section 3 illustrates our approach to identifying the contribution of relative rank, college inputs, and peer quality. We present our insights from program-level RD value-adds and OLS value-adds in Section 4. In Section 5, we plot and interpret the distributions of program RD VA, relative rank VA, and quality VA. We conclude in Section 6.

2 Data and Empirical Strategy

2.1 Overview of Data Sources

We rely on four primary sources of administrative data in this paper. First, we have access to student-level data from a common state-level entrance examination. This dataset contains students' entrance examination marks and their 12^{th} grade marks. It also includes various demographic characteristics like their caste, gender, and affirmative action categories. This student-level dataset allows us to construct admission cutoffs for each program. Since we also have access to the seat category under which a student is admitted to a program, we can construct multiple admission cutoffs for a given program, corresponding to each seat category. Second, we have access to data from the centralized admissions mechanism that is linked to the student-level entrance exam data. This dataset contains the rank-ordered list of program preferences for each student and information on student-program matches or allocations. We are also able to see if students actually enroll in their allotted programs or not, i.e. their "take-up" of the program allocation. Third, we have access to students' semester examination scores for all colleges affiliated with the state's technical university. This accounts for 80% of all colleges in the centralized admissions process.

By linking these semester exam scores to students' entrance exam data, we are able to construct a complete dataset for each student that traces their academic outcomes from high school, through college, until graduation. All colleges affiliated with the state technical university administer the same semester exams within a program, allowing us to construct program value-added metrics that are comparable across colleges within a major. For the remainder of this paper, we focus on the computer science engineering (CSE) major, but our approach can easily be duplicated to any other major in the system. Finally, we also have access to college-level inputs (number of hostels, libraries, laboratories, playgrounds, etc.) from the All India Survey of Higher Education (AISHE) dataset that we have linked to each college in the state technical university.

2.2 Empirical Strategy

We use a regression discontinuity (RD) approach that examines the neighborhood of admission cutoffs to a given program (college + major combination). This allows us to compare the outcomes of students who are barely admitted to a program to counterfactual outcomes had they missed the admissions cutoff and, therefore, been denied admission into their "preferred program". In all regression specifications that follow, i indexes a student and j a program, respectively. $k \in \{OC, AA\}$ indicates the type of cutoff a student faces, i.e., the seat category they were admitted under: Affirmative Action (AA) or Open Category (OC). D_{ij} is an indicator that takes the value 1 if a student i is admitted to classroom j. X_{ij} indicates the value of the running variable for a student, i.e. their distance from the cutoff score. α_j is a *program waitlist* fixed effect and α_k is a *seat category* fixed effect. β^D is the slope coefficient of interest (VAM or change in outcome upon clearing an admission cutoff).

2.3 Validating the RD Approach

2.3.1 First Stage: Take-up of Allotment

We first estimate how marginally clearing a cutoff affects enrollment in the college using the following specification:

$$y_{ij} = \alpha_j + \beta^D D_{ij} + \beta^X X_{ij} + \beta^{DX} D_{ij} X_{ij} + \epsilon_{ij}$$

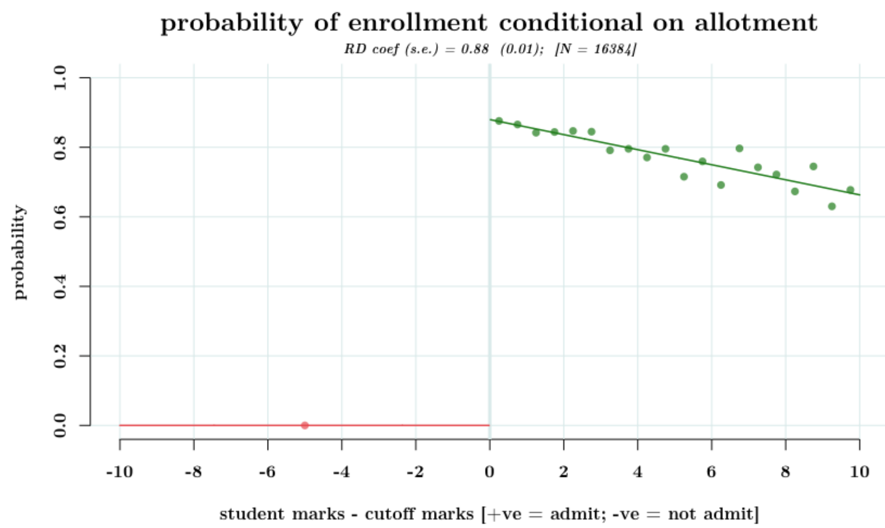

Figure 1: ***Probability of Enrollment.*** *The horizontal axis shows the running variable X_{ij}, which denotes the distance between a student's marks on the entrance exam and the marks of the last student admitted to a program. The vertical axis shows the probability of enrollment conditional on receiving a program allocation in the centralized admissions mechanism.*

In Figure 1, we see that upon clearing a cutoff and receiving a program allocation, the probability of enrolling in the allotted program increases from 0 to 0.88. By construction, the left side of the graph is at zero. On the right, since certain students may choose to leave the state, not go to any college, or join a private college, the probability does not jump all the way to 1. Table A1 shows the RD coefficients for the first stage probability of enrollment.

2.3.2 Covariate Smoothness

The specification used to check for smoothness of covariates at the threshold is as follows.

$$y_{ij}^k = \alpha_j + \alpha_k + \beta^D D_{ij} + \beta^X X_{ij} + \beta^{DX} D_{ij} X_{ij} + \epsilon_{ij}$$

For the purpose of the graphs, first, we regress the raw outcome of interest on *program waitlist* and *seat category* fixed effects, i.e. α_j and α_k respectively. It is necessary to include seat category fixed effects because, owing to affirmative action, we expect covariates like high school marks or demographics like students' age to be continuous within seat categories. Second, we obtain the residuals $\hat{\eta_{ij}}$ from the first regression and regress those on the admission indicator D_{ij}, a polynomial in the running variable X_{ij} and an interaction between D_{ij} and X_{ij}. β^D is the coefficient reported on the plots.

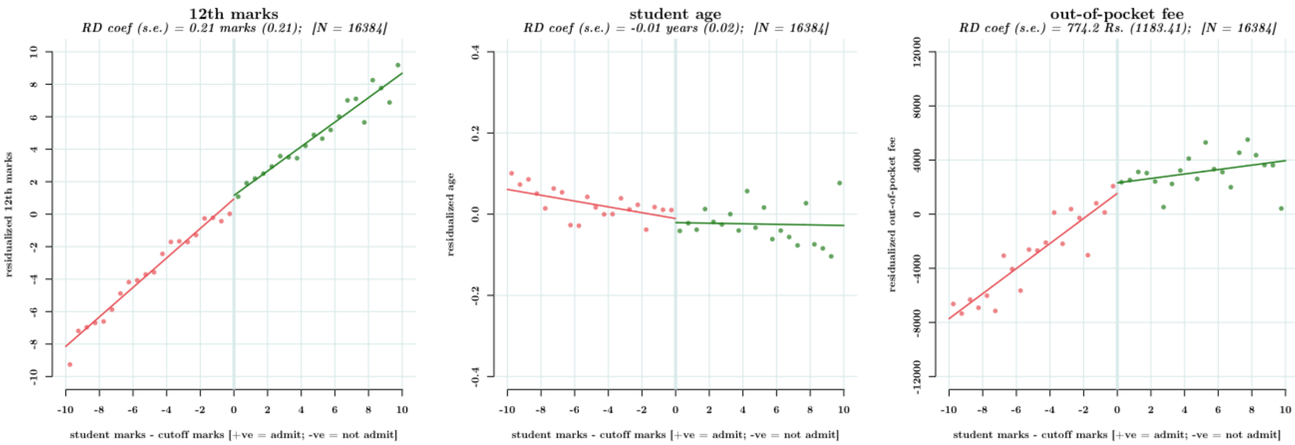

Figure 2: ***Covariate Smoothness.*** *The horizontal axis shows the running variable X_{ij}, which denotes the distance between a student's marks on the entrance exam and the marks of the last student admitted to a program. The vertical axes show the residualized 12^{th} grade marks, student age, and out-of-pocket tuition fee from left to right.*

In Figure 2, we establish that students immediately to the left and right of a program cutoff are comparable based on their ability as measured by their high school graduation scores as well as key demographics like their age and the out-of-pocket tuition fee they would pay at their allotted programs. The RD coefficients are economically and statistically indistinguishable from zero, as expected. Table A2 shows the RD coefficients for the smoothness of covariates at the admission cutoffs.

2.3.3 McCrary Density Tests

Using local polynomial regressions, in Figure 3, we see the density of students to the left and right of cutoffs is similar, i.e. students cannot endogenously choose which side of a cutoff they land on. We examine different bandwidths around the cutoff, and the results fail to show any sharp discontinuities in the density, suggesting manipulation of the running variable (or sorting) is unlikely.

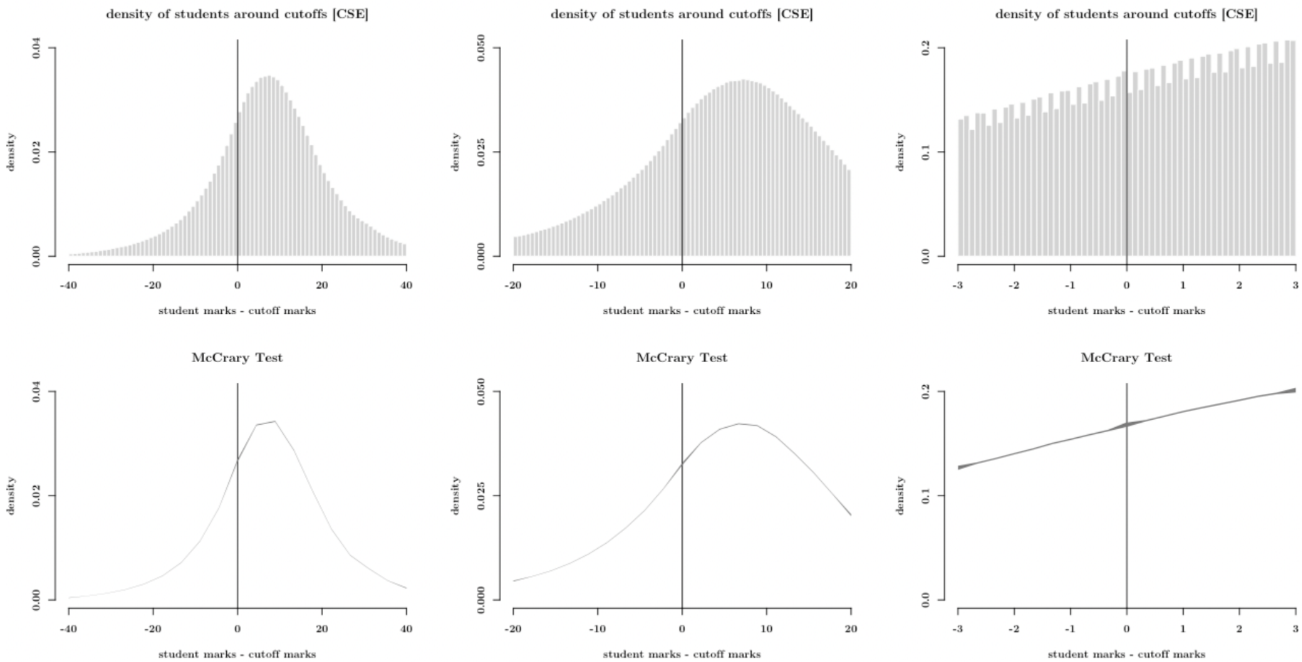

Figure 3: ***McCrary Density Tests using local polynomial regression.*** *The horizontal axes in all plots show the bandwidth around the RD admission cutoff. The vertical axes show the density of the distribution of students around the cutoff. The first, second, and third columns show a bandwidth of ± 40, ± 20, and ± 3 respectively. The top panel shows the density histogram of students on either side of an admissions cutoff. The bottom panel shows the results of a local polynomial regression representing the mass at various points of the density distribution.*

2.4 Pooled Regression Results

In this section, we present the pooled regression results, aggregating over all student types and program cutoffs. The objective is to determine the effect of clearing an admission cutoff and gaining admission into a preferred program. We examine intermediate outcomes like peer quality and relative percentile rank and, thereafter, academic outcomes of students during college. We estimate the following regression specification.

$$y_{ij} = \alpha_j + \beta^D D_{ij} + \beta^X X_{ij} + \beta^{DX} D_{ij} X_{ij} + \epsilon_{ij}$$

2.4.1 Peer Quality and Relative Rank

In Figure 4 (top panel), we do not include seat category fixed effects because our objective here is to estimate the aggregate change in outcomes upon gaining admission into a classroom pooling across all seat categories. Yet, our results are almost identical to including seat category fixed effects as well (bottom panel of Figure 4). In Figure 4, we see that upon gaining admission to a preferred program, a student is exposed to a higher peer quality as measured by the average of all students' entrance exam marks excluding that individual student, i.e. leave-own-out peer quality.

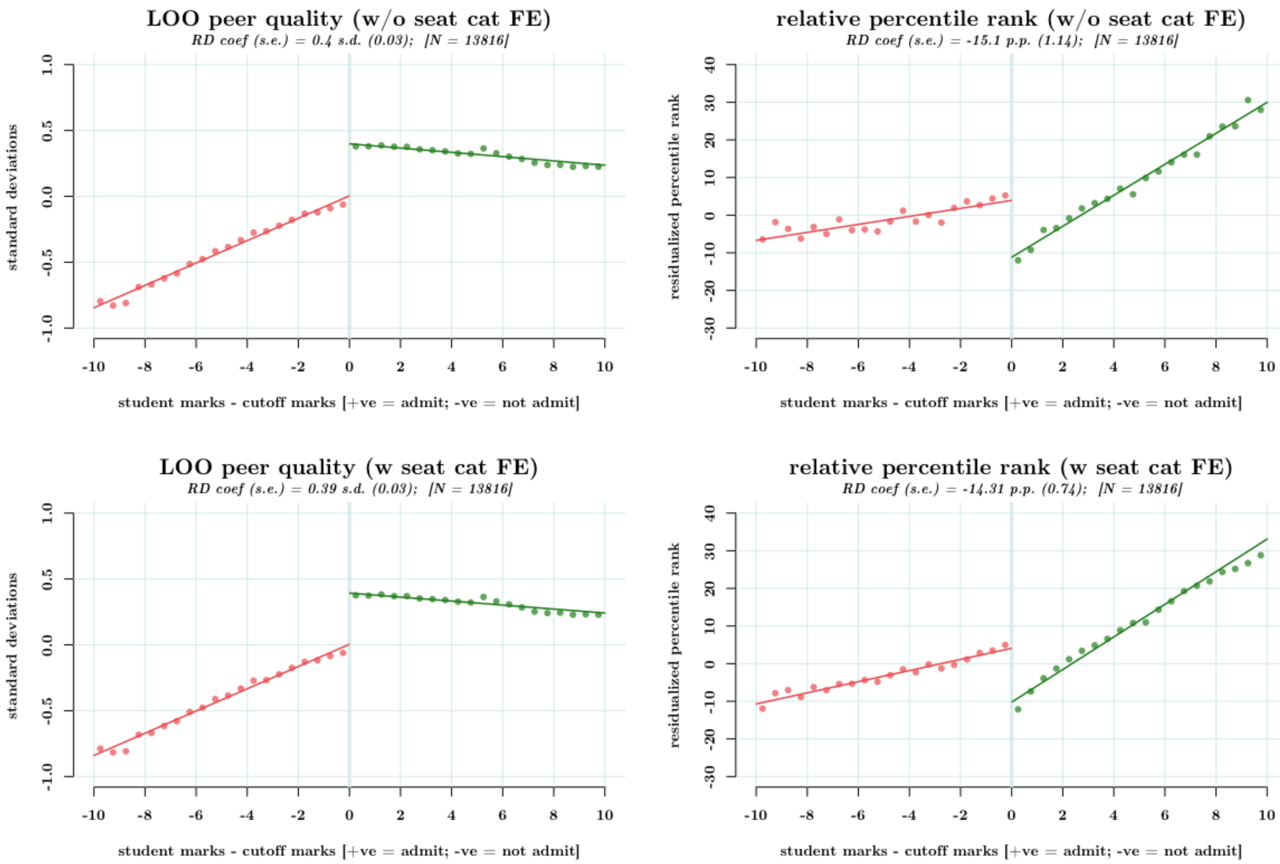

Figure 4: ***Leave-own-out Peer Quality and Relative Rank.*** *The horizontal axis shows the running variable X_{ij}, which denotes the distance between a student's marks on the entrance exam and the marks of the last student admitted to a program. The vertical axes in the first column show the residualized leave-own-out average entrance exam marks in a program, standardized around the mean. The vertical axes in the second column show the residualized relative percentile rank of a student.*

Additionally, a student who is barely admitted to a program, occupies a lower percentile rank in their program relative to the counterfactual student who was not admitted to the preferred program. As we expect, clearing a cutoff implies a student faces a significant increase in peer quality of approximately 0.4 s.d. and a decrease in their relative percentile rank of approximately 15 p.p. Table A3 shows the peer quality and relative rank effects experienced by marginally admitted students.

2.4.2 Academic Outcomes

In Figure 5, we see that, relative to students who just missed a program cutoff, students who are marginally admitted to their preferred programs perform significantly better on their examinations in college and are more likely to graduate from their program. These results suggest that, on average, going to your desired college, over the next best available option leads to higher marks

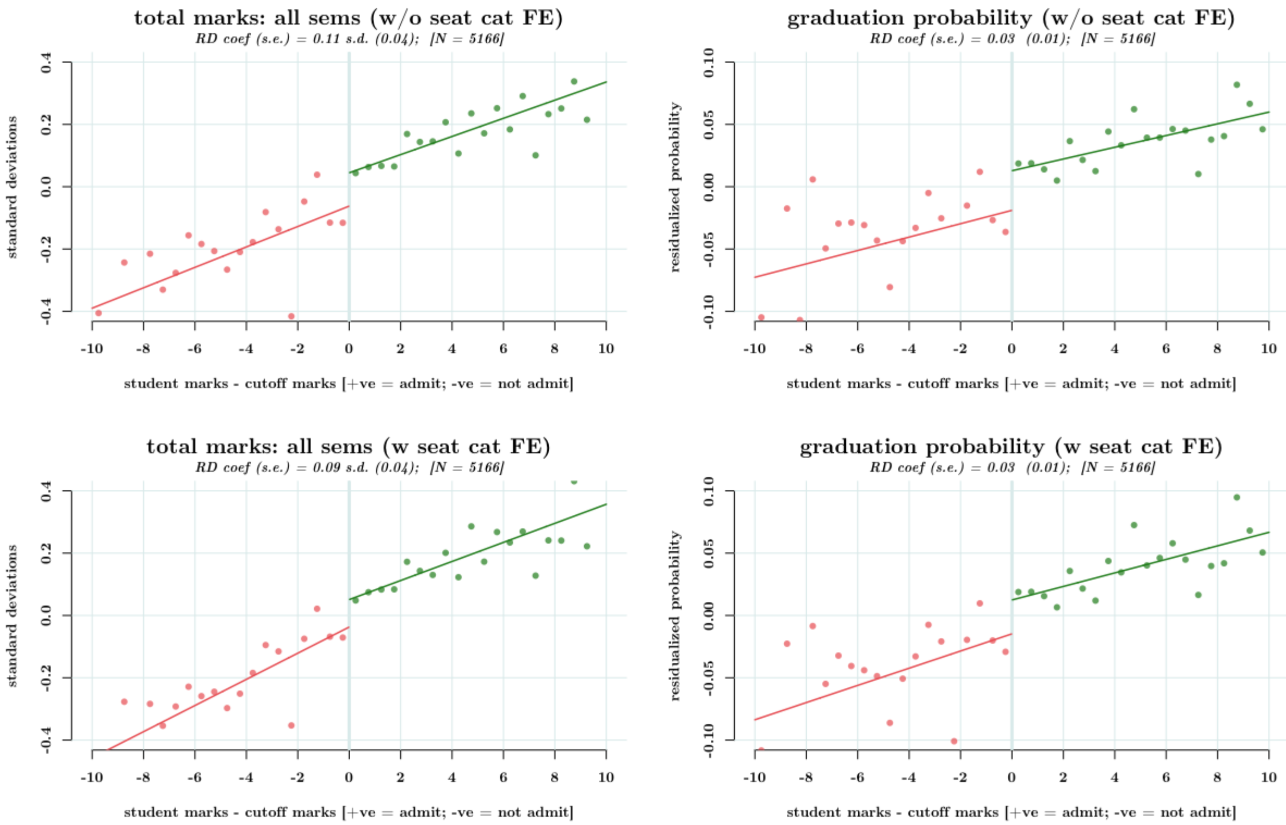

Figure 5: ***Academic Outcomes.*** *The horizontal axis shows the running variable X_{ij} which denotes the distance between a student's marks on the entrance exam and the marks of the last student admitted to a program. The vertical axes in the first column show the residualized total marks scored by a student while enrolled in a program, standardized around the mean. The vertical axes in the second column show the residualized probability of graduating from a program.*

and graduation probability. This may not necessarily be what one would have expected if the high-ability students were simply selecting into the most desirable colleges. Yet, these average aggregate effects hide meaningful heterogeneity. We show below, that even though the average "value added" is positive, there are a large number of (even top-ranked programs) with negative "value added". Table A4 shows the RD coefficients for academic outcomes as well as tests the robustness of these estimates to different bandwidths.

3 Model: Conceptual Framework

Thus far, using an RD approach and comparing students with similar abilities, we have illustrated the aggregate impact of gaining admission into a preferred program. Students who clear an admissions cutoff are exposed to a higher peer quality and are at a lower relative percentile rank in the classroom they are admitted to. Additionally, these students experience a gain in their academic outcomes as they score higher marks in the college semester examinations, and are more likely to graduate, relative to their counterparts who barely missed an admissions cutoff for a given program.

Going forward, our objective is to estimate the "value-added" for each individual program and separately identify the contribution of relative rank effects, college inputs, and peer quality in order to understand how each of these components contributes to an overall measure of program value-added. Equation 1 denotes the outcome of a student i admitted to program j. We assume that y_{ij} is an additive separable function of college inputs Q_j, peer quality PQ_j, relative percentile rank of a student in a program RR_{ij}, their ability A_i and idiosyncratic variation ϵ_{ij}.

$$y_{ij} = \underbrace{\gamma^{C} Q_j}_{\text{college inputs}} + \underbrace{\gamma^{PQ} PQ_j}_{\text{peer quality}} + \underbrace{\gamma^{RR} RR_{ij}}_{\text{relative rank}} + \underbrace{A_i}_{\text{ability}} + \underbrace{\epsilon_{ij}}_{\text{unobservables}} \tag{1}$$

- Step 0: Using an RD design to compare students who fall on either side of an admissions cutoff, we control for ability A_i while estimating program value-added. The underlying RD design assumption is that A_i varies smoothly across the admissions cutoff.

- Step 1: Estimate the *program specific* LATE value added, γ_j, upon admission into a "preferred" program j

$$\begin{aligned}\gamma_j &= y_{ij} - y_{\neg i \neg j} \\ &= \gamma^{C}\Delta\text{college inputs} + \gamma^{PQ}\Delta\text{peer quality} + \gamma^{RR}\Delta\text{relative rank}\end{aligned}$$

  Here, the Δ terms represent changes in college inputs, peer quality, and relative rank for the marginally admitted student as a result of clearing the admissions cutoff.

- Step 2: Separately estimate the relative rank effect for each program, γ_j^{RR}, with multiple cutoffs

$$\gamma_j = \underbrace{\gamma^{C}\Delta\text{college inputs} + \gamma^{PQ}\Delta\text{peer quality}}_{\gamma_j^{Q}} + \underbrace{\gamma^{RR}\Delta\text{relative rank}}_{\gamma_j^{RR}} \tag{2}$$

  We are able to leverage the fact that there are multiple admissions cutoffs created by affirmative action policies, to separately identify γ^{RR}. Consider a program j with two admission cutoffs, one for the students targeted by affirmative action (AA) policies and another for students getting admitted under open competition (OC). For the purpose of exposition, we refer to programs with lower cutoffs than program j as $\neg j$. Let y_{ij}^{OC} and $y_{\neg i \neg j^{OC}}^{OC}$ respectively represent the outcomes for students who clear and miss the OC cutoff for program j. The terms y_{ij}^{AA} and $y_{\neg i \neg j^{AA}}^{AA}$ are defined similarly.

The expected payoff to clearing an OC cutoff is

$$\begin{aligned}\Delta y_j^{OC} &= y_{ij}^{OC} - y_{\neg i \neg j^{OC}}^{OC} \\ &= \gamma^C (Q_j - Q_{\neg j^{OC}}) + \gamma^{PQ}(PQ_j - PQ_{\neg j^{OC}}) + \gamma^{RR}(RR_{ij}^{OC} - RR_{\neg i \neg j^{OC}}^{OC})\end{aligned}$$

The expected payoff to clearing an AA cutoff is

$$\begin{aligned}\Delta y_j^{AA} &= y_{ij}^{AA} - y_{\neg i \neg j^{AA}}^{AA} \\ &= \gamma^C (Q_j - Q_{\neg j^{AA}}) + \gamma^{PQ}(PQ_j - PQ_{\neg j^{AA}}) + \gamma^{RR}(RR_{ij}^{AA} - RR_{\neg i \neg j^{AA}}^{AA})\end{aligned}$$

Now consider the following second difference:

$$\begin{aligned}\Delta^2 y_j &= \Delta y_j^{OC} - \Delta y_j^{AA} \\ &= \gamma^C [(Q_j - Q_{\neg j^{OC}}) - (Q_j - Q_{\neg j^{AA}})] \\ &+ \gamma^{PQ}[(PQ_j - PQ_{\neg j^{OC}}) - (PQ_j - PQ_{\neg j^{AA}})] \\ &+ \gamma^{RR}[(RR_{ij}^{OC} - RR_{\neg i \neg j^{OC}}^{OC}) - (RR_{ij}^{AA} - RR_{\neg i \neg j^{AA}}^{AA})]\end{aligned}$$

We make the following assumptions,

1. A1: $Q_{\neg j^{OC}} = Q_{\neg j^{AA}}$
   A1 posits that the program quality (e.g., from inputs) experienced by OC and AA "cutoff-missers" is the same. This will hold true if OC and AA students have similar underlying preferences for programs and, therefore, are exposed to similar program quality when they miss a cutoff. This assumption can be tested using college input data.
2. A2: $PQ_{\neg j^{OC}} = PQ_{\neg j^{AA}}$
   A2 posits that the peer quality experienced by OC and AA "cutoff-missers" is the same. This assumption is directly testable from the data and holds in practice.

Using A1 and A2, we isolate the parameter of interest, γ^{RR} as follows:

$$\Delta^2 y_j = 0 + 0 + \gamma^{RR} \Big[\underbrace{(RR_{ij}^{OC} - RR_{\neg i \neg j^{OC}}^{OC})}_{\Delta RR^{OC}} - \underbrace{(RR_{ij}^{AA} - RR_{\neg i \neg j^{AA}}^{OC})}_{\Delta RR^{AA}} \Big] \tag{3}$$

The LHS of equation 3 can be obtained by computing the difference between the program value-added for students admitted to program j through OC and AA cutoffs. Similarly, the RHS of equation 3 can be obtained by computing the difference between the relative rank effects for students admitted to program j through OC and AA cutoffs.[5] Thereafter we

[5]Heterogeneous cutoff-specific program VA and relative rank effects are visible in Figure 10.

isolate γ^{RR} by simply rearranging terms. We can then compute γ_j^{RR} as defined in equation 2.

- Step 3: Rearranging the terms in equation 2, we can isolate the joint contribution of college inputs and peer quality γ_j^Q to program VA according to the equation

$$\underbrace{\gamma_j^Q}_{\text{Quality VA}} := \underbrace{\gamma_j}_{\text{Program RD VAM}} - \underbrace{\gamma_j^{RR}}_{\text{Relative Rank VA}}$$

- Step 4: Regress γ_j^Q on all available college inputs and peer quality within each classroom. This allows us to decompose γ_j^Q into peer quality effects, and the impacts of more college inputs.

4 Estimating Program Level Value-Added: γ_j

In this section, we compute program-level value-added measures (VAM) using an RD approach and compare these RD VAM to traditionally computed OLS VAM (Chetty et al., 2014).

4.1 RD Estimation of Program Value Added

The regression specification for estimating individual RD program value added is as follows.

$$y_{ij} = \alpha_j + \sum_{l=1}^{J} \beta_l^D \times D_{ij} \times \mathbb{1}[l = j] + \beta^X X_{ij} + \beta^{DX} D_{ij} X_{ij} + \epsilon_{ij}$$

Figure 6 plots the distribution of β_j^D, i.e. the RD VAM (estimated with and without caste fixed effects) for each program individually. Note that the distribution of RD VAMs does not change drastically upon including caste fixed effects. The modal value added is around 0, and the average is positive because certain colleges have a very high value-added measure.

4.2 Comparing OLS and RD VAM Distributions

Table 1 compares the regression specifications used to compute OLS and RD value added measures. We include student caste fixed effects for the OLS regression in order to control for demographic characteristics of students. Our approach here is guided by the literature on estimating value-added (Chetty et al. 2014; Koedel and Rockoff 2015). Therefore our preferred OLS VAM specification includes caste fixed effects and our preferred RD VAM specification does not include caste fixed effects.

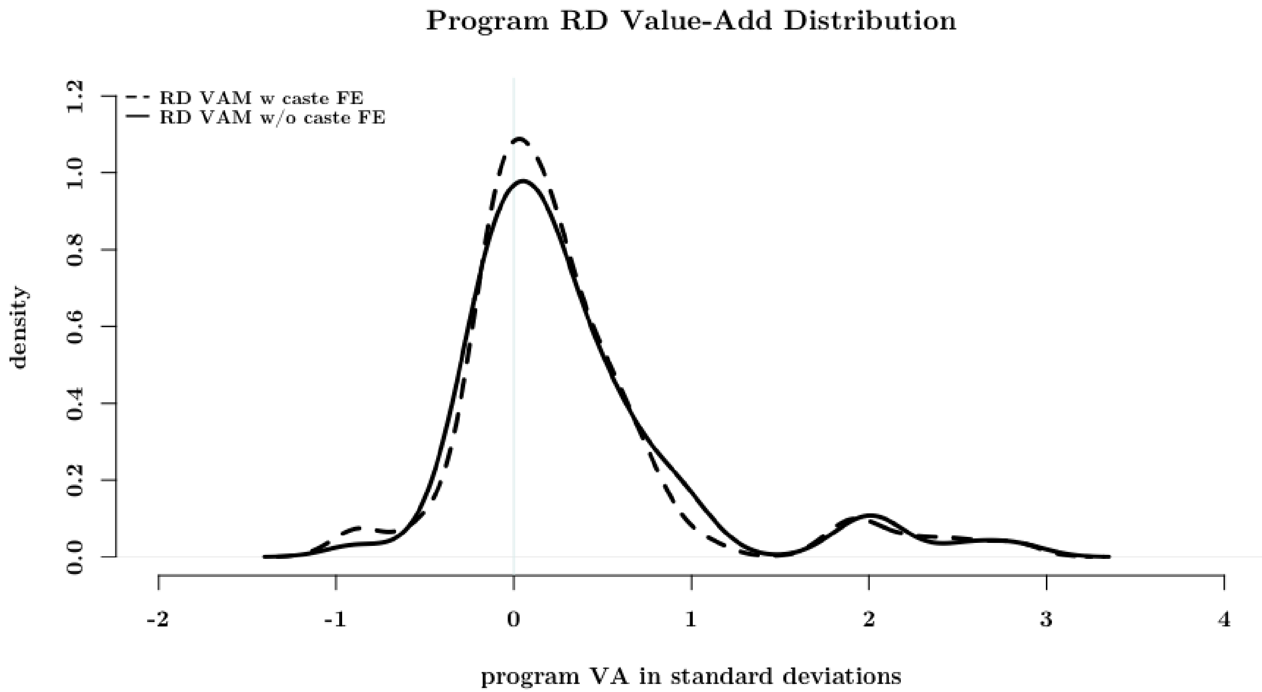

Figure 6: ***Distribution of Program RD Value-Added Measures***. *The horizontal axis shows the program RD VAM or the change in total marks of marginally admitted students, measured in standard deviations, relative to students who just missed a program cutoff. The vertical axis shows the density of the distribution.*

Table 1: OLS and RD Specifications

	OLS	RD
Outcome	total exam scores (s.d.)	total exam scores (s.d.)
FE	actual program, caste	program waitlist
Controls	high school, EAMCET scores (s.d.)	RV, RV interacted with treatment indicator
SE cluster level	actual program	program waitlist

In Figure 7, we see that the OLS and RD VAM distributions are centered around zero, however there are some programs that have a very high RD VAM, to the order of $\geq 1.5\sigma$, which extends the right hand tail of the RD VAM distribution.

4.3 VAM for Most and Least Preferred Programs

Figures 8 and 9 respectively plot the OLS and RD VAM estimates for the top and bottom 25 programs in the system. The programs are chosen based on the highest cutoff marks in that program. For example, in Figure 8, the program **JNKR** has the highest cutoff marks, followed by **GNTW** and so on. Similarly, in Figure 9, the program **SDEW** has the lowest cutoff mark. Owing to the serial dictatorship property of the centralized admissions mechanism, we can infer

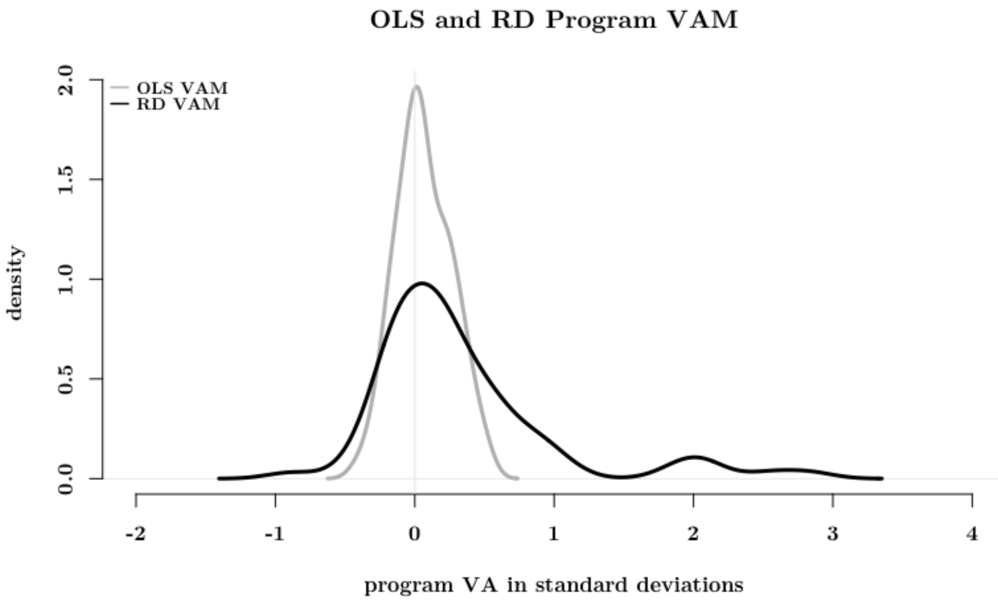

Figure 7: ***Comparing OLS and RD Program VAM Distribution***. *The horizontal axis shows the RD and OLS VAM measures. The vertical axis shows the density of each distribution.*

that programs with higher cutoff marks are more preferred by students. In general, we observe

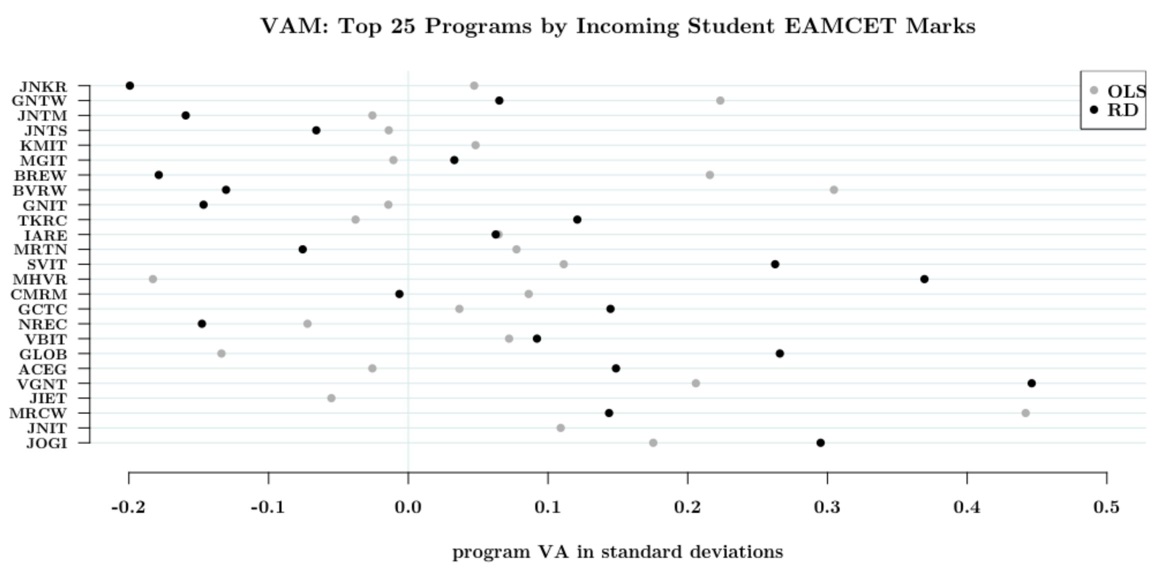

Figure 8: ***Comparing the Top 25 Programs' OLS and RD VAM***. *The horizontal axis shows the RD and OLS VAM measures. The vertical axis shows the corresponding programs ranked from best to worst (top to bottom) based on incoming student quality. Black and gray dots represent the RD and OLS program VAM, respectively.*

that OLS and RD VAM estimates differ considerably. In Figure 8 we see that numerous very desirable programs have a negative RD VAM, but a large positive OLS VAM. On the other hand, in Figure 9, we observe that some of the least desirable programs have a high RD VAM relative to their OLS VAM. Therefore, we can conclude that OLS VAM estimates are often biased by the quality of the incoming students and may not accurately reflect the value-added by a program itself.

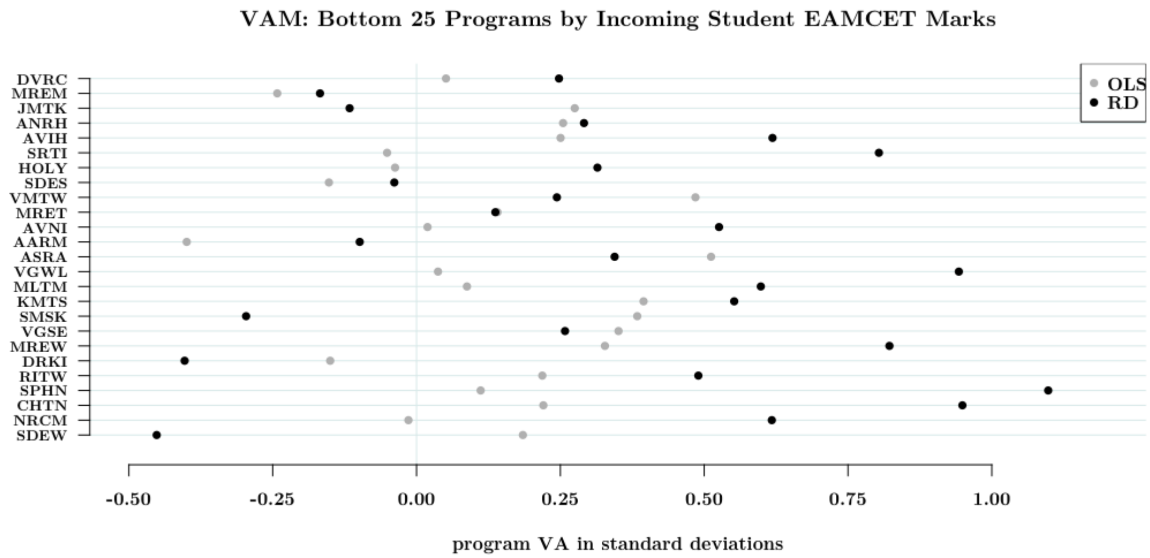

Figure 9: ***Comparing the Bottom 25 Programs' OLS and RD VAM***. *The horizontal axis shows the RD and OLS VAM measures. The vertical axis shows the corresponding programs ranked from best to worst (top to bottom) based on incoming student quality. Black and gray dots represent the RD and OLS program VAM respectively.*

5 Decomposing Program Value-Added

Program VAM (defined as γ_j in equation 2) measures pick up a combination of three things, namely college inputs, peer quality, and relative percentile rank. First, college inputs may be better in top colleges, and inputs may improve test score outcomes. Second, high-ability peers select into top colleges, and the peer effects may either improve or decrease one's performance. Third, students who are marginally admitted to a preferred program experience a significant drop in their relative percentile rank relative to their counterfactual student who missed an admission cutoff, but is the best in their admitted classroom. This change in the relative position of a student can have an impact on their performance in college.

$$\gamma_j = \underbrace{\gamma^C[\Delta\text{college inputs}] + \gamma^{PQ}[\Delta\text{peer quality}]}_{\gamma_j^Q} + \underbrace{\gamma^{RR}[\Delta\text{relative rank}]}_{\gamma_j^{RR}} \tag{4}$$

Consider a program j. Using an RD approach, we have estimated the overall value added by program j, namely γ_j. Next, we want to examine the various components that make up this VA measure. We posit that program VA can be decomposed into the effect of college inputs, peer quality, and relative rank. Δcollege inputs, Δpeer quality, and Δrelative rank represent the change in college inputs, peer quality, and relative percentile rank experienced by a student who clears a cutoff and is marginally admitted to program j. Hereafter we will refer to γ_j as **program RD VA**, γ_j^Q as **quality VA**, and γ_j^{RR} as **relative rank VA**.

5.1 Heterogeneous Effects by Program-Cutoff Type

In Section 3, we developed a new method to distinguish the relative rank VA γ_j^{RR} from the overall program VA, γ_j. The advantage of our setting is that we have multiple cutoffs within a program, owing to affirmative action rules that are uniform across all programs. This enables us to estimate both, RD VAM and the change in relative rank at different points of the score distribution *within* a classroom. This unique feature of our setting allows us to separately estimate relative rank VA γ_j^{RR}, unlike what others have been able to estimate in different settings.

Figure 10 plots the cutoff type specific density distributions of program VA and relative rank, i.e. the left-hand panel corresponds to the distribution of Δy_j^{OC} and Δy_j^{AA} and the right-hand panel corresponds to the distribution of ΔRR^{OC} and ΔRR^{AA} as defined in Section 3. We observe that although the program VA and relative rank effect distributions are similar for students admitted through OC and AA cutoffs, these distributions are not exactly the same. Therefore, computing the second differences $\Delta y_j^{OC} - \Delta y_j^{AA}$ and $\Delta RR^{OC} - \Delta RR^{AA}$ is plausible, and these differences do not reduce to zero. The plausibly exogenous variation in these differences enables us to compute γ^{RR} and subsequently the relative rank VA, $\gamma_j^{RR} := \gamma^{RR}\Delta$relative rank as outlined in equation 4.

5.2 Components of Program Value-Added

Thus far, we have posited that program RD VA (γ_j) is made up of three individual components, namely, college inputs, peer quality, and relative rank. The first two are jointly called quality VA (γ_j^Q), and we have separately computed the third component, relative rank VA (γ_j^{RR}), by leveraging the uniqueness of our setting. Figure 11 shows the density distributions of program RD VA, relative rank VA, and quality VA. We can see that program RD VA is made up of two potentially opposing effects as one might theorize. The modal value of the relative rank VA distribution is negative, suggesting that, on average, the marginally admitted student experiences a negative effect from being the “lowest scoring” student relative to their peers. On the other hand, the quality VA distribution has a positive modal value, suggesting that once the relative rank VA effects are separated from overall program VA, college inputs and exposure to a higher

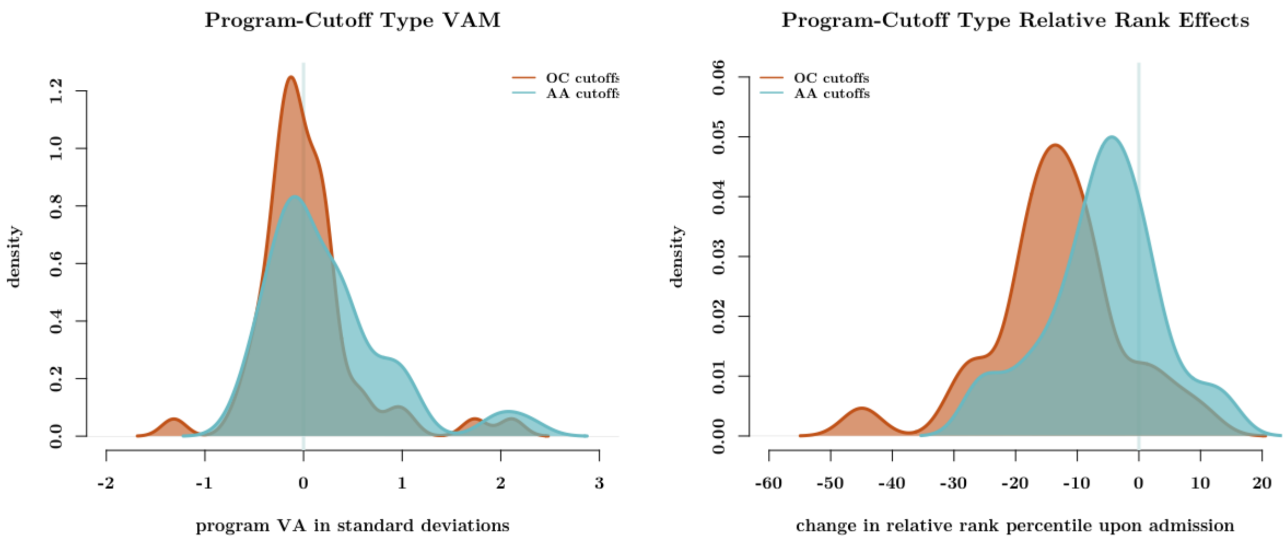

Figure 10: ***Distribution of Program-Cutoff Type RD VAM and Relative Rank Effects***. *The left-hand panel shows the distributions of Δy_j^{OC} and Δy_j^{AA} in orange and blue, respectively. The horizontal axis shows the VA measured in standard deviations. The right-hand panel shows the distribution of ΔRR^{OC} and ΔRR^{AA} in orange and blue, respectively. The horizontal axis shows the change in relative percentile rank upon admission to a preferred program. The vertical axes in both panels show the density of each distribution.*

peer quality have a positive effect on the marginally admitted student. I.e., rearranging the terms in equation 4, we isolate quality VA γ_j^Q that can be regressed on a combination of college inputs and peer quality to quantify the unexplained variation in program VA owing to factors other than the relative rank VA.

$$\underbrace{\gamma_j^Q}_{\text{Quality VA}} := \underbrace{\gamma_j}_{\text{Program RD VAM}} - \underbrace{\gamma_j^{RR}}_{\text{Relative Rank VA}}$$

6 Conclusion

In this paper, our primary objectives are (i) to compute a metric of education quality, namely program (college + major combination) value-added, and (ii) to decompose this metric into the effects of college inputs, peer quality, and relative rank. In our setting, students give a common state-level entrance examination, receive their scores, and submit their rank-ordered list of program preferences. Students and programs are matched using a candidate proposing Deferred Acceptance Mechanism, with serial dictatorship, that incentivizes truth telling. This generates admission cutoffs for each program. Using a regression discontinuity approach we compare students with entrance exam scores in the neighborhood of these admission cutoffs to estimate the

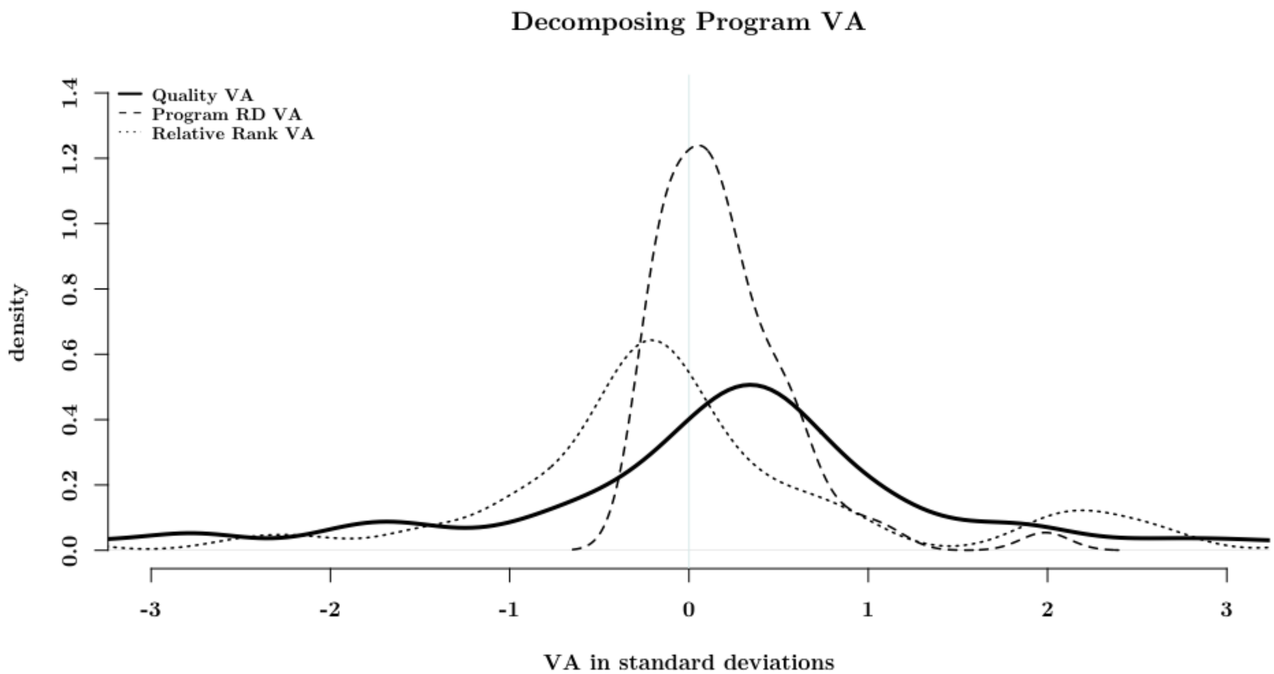

Figure 11: ***Decomposing Program RD VAM***. *The horizontal axis shows the VA in standard deviations. The vertical axis shows the density of each distribution. The densities of program VA* γ_j*, relative rank VA* γ_j^{RR}*, and quality VA* γ_j^{Q} *are represented by the dashed, dotted, and solid lines, respectively.*

effect of admission into a preferred program while controlling for students' academic ability at the time of admission. On average, we find that being marginally admitted to a preferred program exposes students to a higher quality of peers and causes a decrease in the relative percentile rank in their classroom. Moreover, we also find that gaining admission into a preferred classroom has a positive effect on the academic achievement of students as measured by their total marks in college, the number of times they fail, and their probability of graduation.

Since all colleges in our setting administer the same semester exams within a program, we are able to use the academic achievement of students to develop a program value-added metric for individual programs in this setting. Thereafter we posit that these program value-added metrics are composed of college inputs, peer quality, and relative rank effects. By leveraging the fact that affirmative action policies result in multiple cutoffs within a single classroom, we are able to separately identify the contribution of relative rank to the overall program value-added. We will now proceed to regress the residual value-added that comes from factors other than relative rank on college inputs and peer quality changes that students experience upon being admitted to a preferred program.

This paper makes two important contributions to the literature. First, we develop and estimate an education quality metric in one of the largest higher education markets in India. Our method can be applied to higher education settings across the world which follow centralized admissions processes. Second, to our knowledge, this is the first paper that separately estimates the contribution of relative rank effects to overall program value-added. This will enable administrators and policymakers to understand the actual impact that various college inputs and peer quality have on program value-added as they make important allocation and investment decisions.

References

Bagde, S., Epple, D., and Taylor, L. (2016). Does affirmative action work? caste, gender, college quality, and academic success in India. *American Economic Review*, 106(6):1495–1521.

Baswana, S., Chakrabarti, P. P., Chandran, S., Kanoria, Y., and Patange, U. (2019). Centralized admissions for engineering colleges in India. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 323–324.

Bertanha, M. (2020). Regression discontinuity design with many thresholds. *Journal of Econometrics*, 218(1):216–241.

Bertoni, M., Nisticò, R., et al. (2018). Rank concerns, peer effects, and ability tracking in university: Evidence from a randomized experiment. *Centre for Studies in Economics and Finance (CSEF) Working Paper*, (506).

Black, S. E., Devereux, P. J., and Salvanes, K. G. (2013). Under pressure? The effect of peers on outcomes of young adults. *Journal of Labor Economics*, 31(1):119–153.

Booij, A. S., Leuven, E., and Oosterbeek, H. (2017). Ability peer effects in university: Evidence from a randomized experiment. *The Review of Economic Studies*, 84(2):547–578.

Carrell, S. E., Fullerton, R. L., and West, J. E. (2009). Does your cohort matter? Measuring peer effects in college achievement. *Journal of Labor Economics*, 27(3):439–464.

Cattaneo, M. D., Keele, L., Titiunik, R., and Vazquez-Bare, G. (2021). Extrapolating treatment effects in multi-cutoff regression discontinuity designs. *Journal of the American Statistical Association*, 116(536):1941–1952.

Chetty, R., Friedman, J. N., and Rockoff, J. E. (2014). Measuring the impacts of teachers i: Evaluating bias in teacher value-added estimates. *American Economic Review*, 104(9):2593–2632.

Dasgupta, U., Mani, S., Sharma, S., and Singhal, S. (2022). Effects of peers and rank on cognition, preferences, and personality. *Review of Economics and Statistics*, 104(3):587–601.

Delaney, J. M. and Devereux, P. J. (2022). Rank effects in education: What do we know so far? *Handbook of Labor, Human Resources and Population Economics*, pages 1–24.

Denning, J. T., Murphy, R., and Weinhardt, F. (2023). Class rank and long-run outcomes. *Review of Economics and Statistics*, 105(6):1426–1441.

Elsner, B. and Isphording, I. E. (2017). A big fish in a small pond: Ability rank and human capital investment. *Journal of Labor Economics*, 35(3):787–828.

Elsner, B. and Isphording, I. E. (2018). Rank, sex, drugs, and crime. *Journal of Human Resources*, 53(2):356–381.

Elsner, B., Isphording, I. E., and Zölitz, U. (2021). Achievement rank affects performance and major choices in college. *The Economic Journal*, 131(640):3182–3206.

Fabregas, R. (2023). Trade-Offs of Attending Better Schools: Achievement, Self-Perceptions and Educational Trajectories. *The Economic Journal*, 133(655):2709–2737.

Feld, J. and Zölitz, U. (2017). Understanding peer effects: On the nature, estimation, and channels of peer effects. *Journal of Labor Economics*, 35(2):387–428.

Francis-Tan, A. and Tannuri-Pianto, M. (2018). Black movement: Using discontinuities in admissions to study the effects of college quality and affirmative action. *Journal of Development Economics*, 135:97–116.

Gale, D. and Shapley, L. S. (1962). College admissions and the stability of marriage. *The American Mathematical Monthly*, 69(1):9–15.

Jain, T. and Kapoor, M. (2015). The impact of study groups and roommates on academic performance. *Review of Economics and Statistics*, 97(1):44–54.

Kirkeboen, L. J., Leuven, E., and Mogstad, M. (2016). Field of study, earnings, and self-selection. *The Quarterly Journal of Economics*, 131(3):1057–1111.

Koedel, C. and Rockoff, J. E. (2015). Value-added modeling: A review. *Economics of Education Review*, 47:180–195.

Kumar, G. N. (2023). Improving public school productivity: Evidence from model schools in india. *Economics of Education Review*, 97:102465.

Murphy, R. and Weinhardt, F. (2020). Top of the class: The importance of ordinal rank. *The Review of Economic Studies*, 87(6):2777–2826.

Rubinstein, Y. and Sekhri, S. (2013). Value-added versus positive sorting in higher education: Evidence from general education sector in India. Technical report, Citeseer.

Saavedra, J. E. (2008). The returns to college quality: A regression discontinuity analysis. *Harvard University Manuscript*.

Sacerdote, B. (2001). Peer effects with random assignment: Results for Dartmouth roommates. *The Quarterly Journal of Economics*, 116(2):681–704.

Sekhri, S. (2020). Prestige matters: Wage premium and value addition in elite colleges. *American Economic Journal: Applied Economics*, 12(3):207–225.

Whitmore, D. (2005). Resource and peer impacts on girls' academic achievement: Evidence from a randomized experiment. *American Economic Review*, 95(2):199–203.

Zimmerman, S. D. (2014). The returns to college admission for academically marginal students. *Journal of Labor Economics*, 32(4):711–754.

Appendix

Table A1: First Stage - Probability of Enrollment: ±10 bandwidth

	$\mathbb{P}$(Enrollment = 1)
$\mathbb{1}\{X_{ijk} \geq 0\}$	0.878***
	(0.009)
Observations	16,384
R^2	0.721
Adjusted R^2	0.721
Note:	*p<0.1; **p<0.05; ***p<0.01

Table A1 shows the probability that a student enrolls in their allotted program. We see that approximately 88% of students "take-up" their admission offer and choose to enroll in their allotted program.

Table A2: Covariate Balance at Cutoffs: ±10 bandwidth

	12^{th} Marks (marks)	Student Age (years)	Out-of-pocket Expense (Rupees)
$\mathbb{1}\{X_{ijk} \geq 0\}$	0.214	−0.010	774.200
	(0.209)	(0.017)	(1, 183.414)
Observations	16,384	16,384	16,384
R^2	0.208	0.002	0.018
Adjusted R^2	0.208	0.002	0.018
Note:			*p<0.1; **p<0.05; ***p<0.01

Table A2 shows the balance in covariates in the neighborhood of admissions cutoffs. I.e., for the RD design to be a valid identification strategy, we require that students to the left and right of each admission cutoff are comparable. We see that this assumption holds up since students have similar ability as measured by their high school graduation scores, are of a similar age, and pay a similar amount of tuition fee out-of-pocket, regardless of which side of the admissions cutoff they occupy. The RD coefficients obtained from regressing the covariates on an indicator for admission are not significantly different from zero.

Table A3 shows the changes in peer quality and relative percentile rank experienced by marginally admitted students relative to the students who barely miss the admissions cutoff for a program.

Table A3: Peer Quality and Relative Percentile Rank: ±10 bandwidth

	Peer Quality		Relative Rank	
	All (s.d.)	Affiliated (s.d.)	All (p.p.)	Affiliated (p.p.)
$\mathbb{1}\{X_{ijk} \geq 0\}$	0.317***	0.395***	−15.649***	−15.102***
	(0.023)	(0.027)	(1.094)	(1.144)
Observations	16,383	13,816	16,383	13,816
R^2	0.528	0.557	0.075	0.080
Adjusted R^2	0.528	0.556	0.075	0.080
Note:			*p<0.1; **p<0.05; ***p<0.01	

Columns titled "All" and "Affiliated" respectively refer to all programs that are part of the centralized admissions mechanism and the 80% of programs that are affiliated with the state's technical university. Since we have semester examination (outcome) data only for the latter sub-sample of programs, we show that changes in peer quality and relative rank are consistent in both groups of programs. Marginally admitted students experience an increase in peer quality (0.32 - 0.4 σ) and a decrease in their relative percentile rank of a little over 15 p.p.

Table A4 shows the changes in academic outcomes for marginally admitted students relative to the students who barely miss the admissions cutoff for a program. We see that students see a significant improvement in their total semester examination scores (approximately 0.12 σ) across all semesters as well as their scores on exams each year they are in college. Marginally admitted students also have a significantly higher probability of graduating from their program. These results are fairly robust across different bandwidths around a program's admissions cutoff.

Table A4: Academic Outcomes

	Total Marks (s.d.)	$\mathbb{P}$(Graduation=1)	Yr 2 Marks (s.d.)	Yr 3 Marks (s.d.)	Yr 4 Marks (s.d.)
Panel A: ± 6 ***bandwidth***					
$\mathbb{1}\{X_{ijk} \geq 0\}$	0.117**	0.033*	0.093	0.118**	0.121**
	(0.060)	(0.018)	(0.059)	(0.056)	(0.058)
Observations	4,033	4,033	4,033	4,033	4,033
R^2	0.137	0.066	0.136	0.113	0.104
Adjusted R^2	0.121	0.048	0.120	0.096	0.087
Panel B: ± 10 ***preferred bandwidth***					
$\mathbb{1}\{X_{ijk} \geq 0\}$	0.114**	0.034**	0.087*	0.123***	0.113**
	(0.046)	(0.015)	(0.045)	(0.043)	(0.045)
Observations	5,166	5,166	5,166	5,166	5,166
R^2	0.141	0.063	0.138	0.114	0.105
Adjusted R^2	0.127	0.048	0.124	0.100	0.091
Panel C: ± 14 ***bandwidth***					
$\mathbb{1}\{X_{ijk} \geq 0\}$	0.108**	0.028**	0.081*	0.116***	0.096**
	(0.044)	(0.013)	(0.044)	(0.043)	(0.042)
Observations	5,592	5,592	5,592	5,592	5,592
R^2	0.146	0.065	0.144	0.117	0.109
Adjusted R^2	0.134	0.052	0.132	0.104	0.096
Panel D: ± 18 ***bandwidth***					
$\mathbb{1}\{X_{ijk} \geq 0\}$	0.119***	0.030**	0.096**	0.123***	0.107***
	(0.040)	(0.013)	(0.039)	(0.040)	(0.040)
Observations	5,773	5,773	5,773	5,773	5,773
R^2	0.148	0.065	0.145	0.118	0.109
Adjusted R^2	0.136	0.052	0.133	0.106	0.097

Note: *p<0.1; **p<0.05; ***p<0.01