

Distance, Gender, and Access to Elite Higher Education

Merve Demirel
Stockholm University

Xuanzhe Hou
University of Michigan

Yash Kanoria*
Columbia Business School

This version: October 21, 2026

Abstract

We study how migration distance shapes students' higher education choices, using administrative data from the 2015 centralized admissions to India's elite public engineering institutes. We document that candidates list programs farther from home less often and rank them lower, and that this pattern is stronger for women, who are assigned to programs 120 km closer to home than men on average. Estimating a discrete choice model of candidates' first-round decisions among the programs they can attain, we find that for a typical male candidate a program 1,000 km farther from home is worth roughly a quarter to a third of a standard deviation of program value less. Women are about 1.6 to 1.7 times as sensitive to distance as men. Holding attainable programs fixed, this difference accounts for 25 percent of the gender gap in the value of the programs candidates can attain. These findings show that spatial frictions create barriers to accessing high-quality education and widen gender inequalities in human capital accumulation, even among high-achieving students at elite STEM institutions.

Keywords: centralized admissions, distance, gender, revealed preference.

JEL codes: I23, I24, J16, R23, C35.

* Author order is alphabetical. This paper supersedes an earlier solo draft by [Demirel \(2025\)](#).

1 Introduction

Higher education is among the most consequential investments an individual can make, yet educational opportunities vary sharply across space. In response to these spatial differences, migrating for education potentially offers large economic gains that will persist over one’s lifetime. Who can access these opportunities therefore depends on who is able to migrate. If the cost of migration is large, distance itself may translate the existing spatial differences in opportunity into persistent inequalities in access by determining who can reap the long-term returns to higher education.

This paper studies how migration distance shapes the educational migration decisions of elite science and engineering students in India. We begin by documenting that distance to college from a student’s home matters in submitted choices. Using complete administrative records from the 2015 cohort of India’s centralized seat allocation mechanism for public engineering colleges (Joint Seat Allocation Authority, hereafter JoSAA), for a sample of 29,024 candidates and 589 programs, we show that students are less likely to include farther-away programs in their rank-order list, and when they do, they rank them lower. This distance gradient is large for men and women alike, but it is systematically steeper for women, who rank distant programs lower than men, are assigned to programs closer to home (536 km on average, compared with 656 km for men), and are less likely to leave their home state to study.

Students may rank and receive closer programs because moving farther is more costly, or because the programs available to them differ systematically with distance. We separate these margins by estimating a structural model of revealed preferences over distance from candidates’ first-round outcomes, conditioning on the programs available to each candidate under the assignment rules. Although the clearinghouse makes truthful reporting safe ([Baswana et al., 2019](#)), identifying preferences from submitted rank-order lists still requires an assumption about how faithfully these lists reflect true preferences. We follow the feasible-choice approach of [Fack et al. \(2019\)](#): we assume only that each candidate chooses her most preferred option among the programs she can attain and an outside option, and we use no other information from how she orders her list. Candidates know their own ranks and past cutoffs and see mock allocations before finalizing their preference lists, so they can tell which programs are within reach; mistakes in how they order or omit programs are therefore unlikely to change the seat most candidates receive.

We estimate a mixed nested logit model for these choices. Our preferred specification includes program fixed effects that capture a program’s overall attractiveness, such as its prestige and quality of education. It also allows women and men in each category to value fields of study differently, and it nests programs by field to capture closer substitution between programs in the same field. Each candidate has a random distance coefficient, with a separate mean and dispersion for each gender-by-category group. The value of the outside option varies with the candidate’s group, home state, and exam rank, and its unobserved component can be correlated with her distance coefficient.

We find that distance is a first-order cost in program choice, even in one of India’s most selective education markets: our preferred estimates imply that, for an average general-category male candidate, avoiding an extra 1,000 km move is worth roughly a quarter to a third of a standard deviation of (estimated) program values. This is a large cost. More than a third of candidates

offered an IIT seat in 2015 had taken the entrance exams the year before, delaying college by a year in pursuit of a better program, and the rank gains from that extra year are worth roughly one standard deviation of program value or less (Section 5.4 and Appendix F).

Women are about 1.6 to 1.7 times as sensitive to distance as men on average. The gender gap holds in every robustness check, including a sample that covers all candidates with a first-round seat, alternative measures of distance, and estimation on only the candidates who accepted a seat.

Affirmative action is central in our context: institutes set aside 49.5 percent of seats for candidates from historically disadvantaged caste and tribal groups, making India’s reservation system one of the largest affirmative action programs in the world. Hence, our setting allows us to compare how gendered distance preferences differ across socioeconomic backgrounds. We refer to candidates eligible for these reserved seats as reserved-category candidates and to all other candidates as general-category candidates. The gender gap is large and statistically significant in both groups, and we cannot reject that it is the same size in each.

Distance preferences account for a substantial share of unequal access at the intensive margin. Holding candidates’ feasible sets fixed, distance preferences explain 25 percent of the female–male access gap. If women were as sensitive to distance as men, their predicted choice distance would rise by 115 km. These results imply that even conditional on feasible opportunities, the cost of migrating to take up those opportunities changes who can access higher education at elite institutions. To the extent that the gendered costs of migration distance reflect constraints rather than preferences over location itself, policies that directly target costs of migration, such as providing hostel access, safer transport, demand-side subsidies, or information about living arrangements, could close a nontrivial part of the gap.

The paper contributes first to work on geography, college choice, and access to higher education. A large literature documents geographic distance as a barrier to college enrollment (Card, 1995; Long, 2004; Alm and Winters, 2009; Spiess and Wrohlich, 2010; Hoxby and Avery, 2013; Acton et al., 2024; Boelmann, 2024; Elsayed and Shirshikova, 2023), and opening campuses closer to home raises educational attainment (Currie and Moretti, 2003; Frenette, 2009; Jagnani and Khanna, 2020), especially among first-generation students (Katzkiewicz et al., 2026). Distance also shapes where students apply (Delaney and Devereux, 2024). A subset of papers highlights the role of migration costs in access to higher education more directly (Fabre, 2026; Ghose, 2026; Ishimaru, 2025; Fu et al., 2022). Estimates of migration costs from this literature are extraordinarily large.¹ They far exceed the observable, monetary costs of relocation, suggesting that they likely represent unobserved preferences and other barriers relating to non-pecuniary aspects of the migration decision. For a young adult, migrating for college means leaving the family home and settling in an unfamiliar place, and it can bring worries about safety, loneliness, or managing daily life alone. We provide revealed-preference estimates of these costs in a lower-income setting, where educational migration

¹Closest to our setting, Ghose (2026) uses district-to-district Census migration flows to estimate that a student moving within an Indian state for college must be compensated with about 91% higher income to be as well off as staying, with crossing a state border adding a further 19 percentage points; and Demirel (2026) elicits stated preferences from students in Maharashtra and finds a non-pecuniary cost of leaving home equivalent to 129% of expected post-graduation salary.

is rare in aggregate,² but a crucial determinant of who can access opportunity (Ghose, 2026). Unlike estimates recovered from aggregate migration flows or from hypothetical choices, ours come from the actual decisions of students at the moment they choose where to study, in a highly mobile (73% of men and 63% of women with a first-round seat are assigned to a college outside of their state) sample. Our results imply that the cost of migration is a critically important part of college choice even among the highest-achieving candidates competing for admission into a highly selective set of colleges, where one might expect distance to matter least.

Second, work on gendered mobility constraints has shown that commuting distance, time, and safety concerns shape women’s economic decisions: women give up earnings or schooling opportunities at a much higher rate compared to men (Borker, 2021; Muralidharan and Prakash, 2017; Cheema et al., 2026; Alba-Vivar, 2025; Le Barbanchon et al., 2021; Field and Vyborny, 2022). We extend this literature from daily commuting and short-run training or job search decisions to a higher-stakes, longer-term mobility decision: where to attend college. Our revealed-preference estimates show that the gendered cost of migration distance is large, both among candidates eligible for preferential admission under affirmative action policies and among other candidates. Importantly, our model separates tastes for field of study from distance, ruling out gender differences in preferred fields as the source of women’s stronger distance aversion.

Our approach builds on the literature on estimating demand in centralized assignment systems (Agarwal and Somaini, 2018; Abdulkadiroğlu et al., 2017; Calsamiglia et al., 2020; Kapor et al., 2020; Agarwal and Somaini, 2020). Its assumption is motivated by evidence that, under deferred acceptance, applicants’ departures from truthful reporting are concentrated on options that are unlikely to affect their assignment (Hassidim et al., 2021; Shorrer and Sóvágó, 2023, 2024; Shorrer et al., 2026; Chrisander and Bjerre-Nielsen, 2023).

The paper is organized as follows. Section 2 describes the institutional setting and the data, and Section 3 presents descriptive evidence. Sections 4–5 set up the identification strategy and the empirical framework. Section 6 reports the main results, interprets the estimated distance coefficients, and decomposes the gender gap in the value of attainable programs. Section 7 reports robustness checks, and Section 8 concludes.

2 Institutional setting and data

2.1 Institutes and entrance exams

India’s federal elite engineering admissions are organized around two groups of institutes: the Indian Institutes of Technology (IITs)³ and a larger group of non-IIT institutes, which in 2015 comprised 31 National Institutes of Technology (NITs), 18 Indian Institutes of Information Technology, and 18 other Government Funded Technical Institutes (Joint Seat Allocation Authority, 2015, Annexure 1).

²The 2011 Census attributes only 1.2 percent of India’s internal migrants to education (Office of the Registrar General & Census Commissioner, India, 2011a), and only 1.19 percent of the college-aged population (15–24) is enumerated as having moved for education (Office of the Registrar General & Census Commissioner, India, 2011b,c).

³Throughout, we include the Indian School of Mines (ISM), Dhanbad, now IIT (ISM) Dhanbad, in the IIT group because it also admitted students through JEE-Advanced in 2015.

In 2015, the Joint Seat Allocation Authority (JoSAA) centralized seat allocation across both groups of institutions, covering 34,074 undergraduate seats ([Joint Seat Allocation Authority, 2015](#), seat matrices, archive pp. 150–153), of which 33,018 are in the 589 science and engineering programs we study.

Admissions are based primarily on candidates’ ranks in the Joint Entrance Examination (JEE). JEE-Main is the entrance exam for the non-IIT institutes and was taken by roughly 1.3 million candidates in 2015 ([Baswana et al., 2019](#)), competing for 23,092 undergraduate engineering seats. JEE-Advanced is the entrance exam for the IITs: about 117,000 candidates took it in 2015 ([Indian Institute of Technology Bombay, 2015](#)), competing for 9,926 undergraduate engineering seats, or about one seat for every twelve test takers. Each exam ranks candidates on a single list, and almost every candidate in our sample has an overall rank on at least one of the two lists.⁴ Candidates eligible for seats reserved under affirmative action, described below, also receive a category rank: their position among candidates of the same category.

Both exams are national, and JoSAA allocates seats at institutes across the entire country, so a candidate’s choice of where to study is also a choice of how far to move from home. Such moves are not common in India: according to the 2011 Census, while 37 percent of the Indian population moved within the country, only 1.2 percent of all migrants moved for education, compared to 46 percent moving for marriage and 9.1 percent for employment ([Office of the Registrar General & Census Commissioner, India, 2011a](#)). Women comprise only 40 percent of those migrating for education, despite representing 68 percent of all migrants in India. Moreover, this gender gap widens with distance: women make up 43 percent of within-district educational migrants but only 32 percent of those moving across state boundaries. The low rate of migration for education is consistent with broader evidence that internal migration in India other than for marriage is low, particularly across state borders ([Munshi and Rosenzweig, 2016](#); [Kone et al., 2018](#)). The JoSAA admissions process is therefore one of the few large scale settings in India where education driven internal migration can be observed directly.

It is also a highly selected setting. JoSAA candidates are not only a selected group in terms of their academic achievement, but also on dimensions of gender and socioeconomic background. Students and families devote significant amounts of time and resources in preparation for the exam, often by taking private coaching classes for a number of years or attempting the exam multiple times. The upper income share is higher among candidates offered an IIT seat than among those who register for JEE-Advanced: in 2016, 16 percent of the former and 9 percent of the latter reported a parental income above 800,000 rupees ([Joint Implementation Committee, 2016](#)). Gender representation is also starkly unequal. Women have higher mean board exam marks than men, but their representation falls from roughly 26 percent of candidates with a JEE-Main rank to 11 percent of candidates with a JEE-Advanced rank and 9 percent of those assigned an IIT seat in the first round (Appendix D). The candidates we observe have therefore already passed a costly preparation and selection margin before making their JoSAA choices.

⁴JoSAA’s term for each list is the Common Rank List (CRL). Two Scheduled Tribe candidates in our sample have no overall rank on either list (Appendix A.4).

2.2 The JoSAA allocation mechanism

A candidate’s allocation is determined by her submitted list, her exam rank, her caste category, and her home state, with no institutional discretion. Before the first round, each candidate submits a single ordered list of programs, where a program is a major at a specific institute, such as computer science at IIT Bombay. The list cannot be changed afterward.⁵ JoSAA then allocates seats by deferred acceptance over several rounds: in 2015, three rounds for the IITs and four for the non-IIT institutes.⁶

As part of India’s affirmative action policy, institutes set aside seats for candidates from Scheduled Castes (SC, 15 percent), Scheduled Tribes (ST, 7.5 percent), and Other Backward Classes (OBC, 27 percent).⁷ Candidates from these groups are reserved-category candidates; all others are general-category. The remaining seats are open to everyone and are allocated by overall rank. Reserved-category candidates compete for open seats using their overall rank and for reserved seats using their category rank, so they can be admitted through either. A reserved-category candidate admitted through an open seat does not use up a reserved seat.

Outside the IITs, admission also depends on a candidate’s home state, defined as the state in which she passed Class XII or an equivalent examination. The IITs admit candidates from all states on equal terms. The NITs, which account for most non-IIT seats, set aside half of their seats for candidates whose home state is the state in which the institute is located, and some of the other non-IIT institutes have similar quotas. A candidate’s category and home state therefore affect which programs are attainable at a given exam rank.

In each round, JoSAA assigns each candidate to the highest program on her list that can accommodate her, given her rank, her category and home state, and the seats remaining. After the round is run, each type of seat in a program has a realized closing rank, or cutoff: the rank of the last candidate admitted. A program is attainable for a candidate in a given round if she clears the closing rank for at least one type of seat she is eligible for.

After each round except the last, candidates who hold a seat choose one of four options. *Float* keeps the current seat while remaining eligible for an upgrade to any program higher on the list. *Slide* keeps the current seat and allows upgrades only within the same institute. *Freeze* accepts the current seat as final. A candidate assigned to the first program on her list must choose to *Freeze*. *Reject* gives up the seat and exits JoSAA, as does failing to complete acceptance. If choosing *Float* or *Slide* options leads to an upgrade at a later round, the earlier seat is automatically forfeited.

⁵JoSAA revised some candidates’ recorded eligibility inputs between rounds. Sample inclusion and feasible sets use the first-round records, while the groups and covariates in the model use the third-round records (Appendix A.4). Appendix A.2 describes the candidates whose recorded category changed.

⁶A round is one of JoSAA’s successive published allocations. At the IITs, the fourth round allocated only seats in preparatory courses (Baswana et al., 2019). A later special round reallocated remaining vacancies; we use it only to identify the Architecture/Planning exclusions in Appendix A.2.

⁷Among Other Backward Classes, only the Non-Creamy Layer (OBC-NCL) is eligible, a designation that excludes relatively advantaged members of the group (Khanna, 2020). Seats for persons with disabilities (PwD) exist within each category. Sample inclusion uses non-PwD ranks, and PwD seats enter the feasible sets of eligible candidates (Appendix A.3).

The first-round seat is therefore consequential. A candidate who wants to remain in the allocation must accept it, since rejecting it ends her participation. From then on she can only move to a program listed higher on her list, and only if a seat in that program opens up. For 65 percent of the candidates in our sample who accept a first-round seat, that seat is the one they hold at the end of the regular rounds.

When a candidate is upgraded or exits the allocation, the seat she held becomes vacant and is allocated again in the next round. Candidates with lower ranks can then be admitted, so a program that was unattainable for a candidate in the first round can become attainable by the last round. Candidates without a first-round seat remain in the allocation for later rounds.

2.3 Candidates’ information and submitted lists

Candidates know their overall and category ranks when they submit their lists, but not the closing ranks they will face. Closing ranks are only determined once the allocation is complete. However, closing ranks from previous years are publicly available, and in 2015 JoSAA advised candidates to consult those of 2013 and 2014. Before the submission deadline, JoSAA also ran mock allocations using the lists candidates had entered so far, and showed each candidate the seat she would receive ([Joint Seat Allocation Authority, 2015](#)). Candidates could then revise their lists before the deadline, which let them think through the consequences of their reported rankings.

Lists are uncapped and costless to extend: adding another option does not require a separate application fee or supplementary materials beyond entering the option on the online portal, the allocation mechanism makes truthful reporting safe ([Baswana et al., 2019](#)), and JoSAA encourages candidates to list broadly ([Joint Seat Allocation Authority, 2015](#)). This reduces the incentive to omit an attainable program that is preferred. Full submitted lists in our sample have a mean length of about 70 programs, often including aspirational, high-prestige programs whose previous or realized closing ranks are well outside the candidate’s likely reach.

Taken together, these features let candidates assess which programs they can attain and which seat their ranking is likely to lead to, while listing programs beyond their reach costs them nothing and cannot make their assignment worse. Section 4 relies on them only for the assumption that a candidate’s observed first-round outcome is her preferred alternative among her feasible programs and the outside option.

2.4 Sample and attainable programs

We use JoSAA’s administrative records for 2015. The records cover all four regular allocation rounds and the later special round, and contain candidate category, gender, home state, examination ranks, submitted preference list, assignments across rounds, and participation decisions across rounds. Program records provide capacities and closing ranks. We restrict our analysis to 589 science and engineering programs and estimate preferences from candidates’ first-round choices among them.

We restrict the sample to candidates whose ranks clear a sample-inclusion cutoff, which we define as follows. For each rank list, we define the sample-inclusion cutoff as the rank at which the number

of candidates who submitted a preference list reaches the number of seats available through that list (Appendix A.1). The sample thus includes the 29,357 candidates who submitted a preference list and whose overall or category rank on either exam clears at least one cutoff. Inclusion does not require receiving or accepting a first-round seat, since selecting on these outcomes would condition on the choices the model explains and would drop candidates who chose the outside option. This rule keeps the sample broad while restricting it to candidates with a substantial number of attainable engineering programs. We note that the sample size (29,357) is similar to the total number of non-PwD seats used to set the cutoffs (32,076); it is slightly smaller due to candidates who clear multiple sample inclusion cutoffs.

We exclude 191 of these candidates (0.65 percent) who were assigned a seat in a preparatory course or in an architecture or planning program in any round, since our model does not cover these seats. This exclusion is based on candidates' realized assignments. We also exclude 111 candidates whose recorded category changed from reserved to general between the first and third rounds, and 31 foreign nationals. The resulting sample has 29,024 candidates: 21,018 accepted a first-round seat, 4,322 were offered a first-round seat but did not accept it, and 3,684 did not receive one. Appendix A details the complete sample construction.

We use the first-round closing ranks to reconstruct which programs each candidate could attain in the first round, including programs she did not list. These programs form the candidate's feasible set, $S_i^{(1)}$; Appendix A documents its construction.

2.5 Measuring distance

Our main measure of distance is the straight-line distance between a candidate's examination city and the institute. We decode the examination city from the candidate's anonymized roll number (Appendix C). Whenever a candidate's examination city is outside of their home state, we instead measure distance from the centroid of their home state. Many candidates attend residential coaching centers far from home and may take the exam at the most convenient location; in our sample, 4,072 candidates (14.03 percent) took the exam in India outside their home state, and 50 took it overseas. The most common destination among these candidates is Rajasthan: 1,070 candidates from other states took the exam there, 996 of them in one of the two well-known coaching hubs (Kota and Jaipur) (Government of Rajasthan, 2015). Using the examination city for these candidates could treat the city of a coaching center as home. As a robustness check, we use the examination city for all candidates who took the exam in India (Section 7.3).

3 Descriptive evidence

This section documents the descriptive patterns that motivate the structural analysis in Section 4: women are assigned closer to home than men, the gap remains substantial after controlling for academic preparation, category, and home state, and it is mirrored in the submitted preference lists.

3.1 Migration outcomes by gender

Among the 25,340 sample candidates who received a first-round seat, women are assigned to programs about 120 km closer to home than men on average and are about 9 percentage points less likely to be assigned outside their home state. Table 1 reports these first-round outcomes separately for reserved and general-category candidates and by gender. Overall, men are assigned 656.0 km from home on average and women 535.7 km, a female–male gap of -120.3 km. The out-of-state shares are 72.9 and 63.5 percent, respectively, a 9.4 percentage point gap. The distance gap is about 120 km in both categories (128 km among general-category and 112 km among reserved-category candidates); women are 9.3 percentage points less likely to be assigned out of state among general-category candidates and 9.6 percentage points less likely among reserved-category candidates.

Table 1: Migration outcomes by gender and category

Category	Mean assigned distance (km)			Share out-of-state (%)		
	Men	Women	F–M (SE)	Men	Women	F–M (pp, SE)
General	653.9	525.9	-128.0 (21.4)	72.4	63.1	-9.3 (2.3)
Reserved	657.7	545.7	-112.0 (17.3)	73.4	63.9	-9.6 (1.9)
All	656.0	535.7	-120.3 (17.5)	72.9	63.5	-9.4 (2.0)

Notes: $N = 25,340$ sample candidates who received a first-round seat, whether or not they accepted it. F–M is the female–male gap, with standard errors clustered by home state in parentheses; all gaps are significant at the 1 percent level using two-sided tests with cluster-based degrees of freedom. Distance uses the origin defined in Section 2.5. Out-of-state compares the institute’s state with the candidate’s home state. Reserved pools OBC-NCL, SC, and ST candidates; General denotes administrative category GN.

Figure 1 shows the distribution of assignment distance. Both distance distributions have long right tails, but women are more concentrated near home: 55.5 percent of women are assigned within 500 km, compared with 46.0 percent of men; only 15.9 percent of women are assigned more than 1,000 km away, compared with 24.4 percent of men.

The distance gap remains substantial after controlling for category, a rank-based measure of academic ability, and home-state fixed effects. With all controls included, women are assigned 153 km closer to home among general-category candidates and 138 km closer among reserved-category candidates. Appendix E, Table 14, reports the four specifications and their coefficients, with standard errors clustered by home state. The Female $\times$ Reserved interaction is small and statistically insignificant throughout. Section 6 revisits this comparison in the structural model.

A natural concern is that the female–male assignment gap reflects women facing thinner choice sets. The gap persists within each tercile of first-round feasible set size. In the largest tercile, with 464 feasible programs on average, women are assigned about 90 km closer to home and are 6.4 percentage points less likely to be assigned out of state.⁸ Thus, the assignment gap is also present among candidates with many attainable options.

⁸Terciles of first-round feasible set size have means of 165, 301, and 464 programs. Across these terciles, the female–male gaps in first-round assignment distance are -125 , -112 , and -90 km. The out-of-state gap in the largest tercile is not statistically significant ($p = 0.12$).

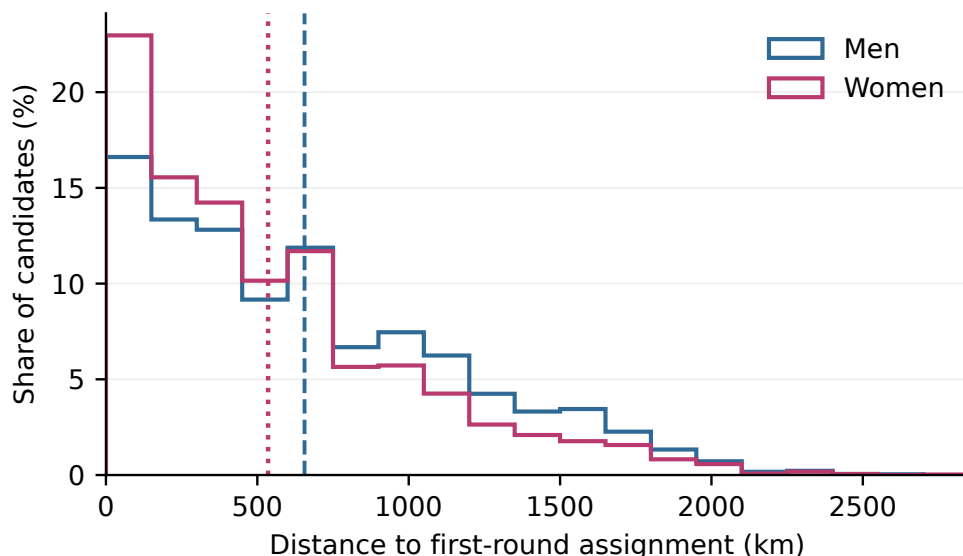


Figure 1: First-round assignment distance by gender, JoSAA 2015

Note: Histograms show distance from the origin in Section 2.5 to the assigned institute for men and women among the 25,340 sample candidates with a first-round seat. Dashed (men) and dotted (women) vertical lines mark the gender-specific means.

3.2 Stated preferences in the submitted lists

The gender gap also appears in both the length and the composition of submitted lists.

List length. Candidates submit relatively long preference lists (even after restricting to the 589 science and engineering programs), with men tending to submit longer lists than women. The median submitted list has 48 programs, 50 for men versus 36 for women (the gender gap is similar for general-category and reserved-category candidates, though the median reserved-category candidate submits a list with length longer by 11). Among candidates with a first-round seat, the average share of the list that is feasible is almost the same by gender but is higher for reserved-category candidates. It is around 27 percent for general-category candidates and 34 percent for reserved-category candidates in the first round.

List composition. Among the listed programs that are feasible in the first round, men list 79.6 percent out-of-state options on average and women 73.0 percent, a 6.6 percentage point gap (Table 2). This list-composition gap is somewhat smaller than the 9.4 percentage point assignment gap.

Among candidates whose first-round feasible set contains both listed and omitted programs, the omitted programs are farther from home than the listed programs by an average of 515 km for women and 343 km for men.⁹

⁹We average each candidate's difference between the mean distance of omitted and listed feasible programs, using the 4,011 women and 21,221 men with both sets nonempty.

Table 2: Out-of-state share on the submitted preference list, by category and gender

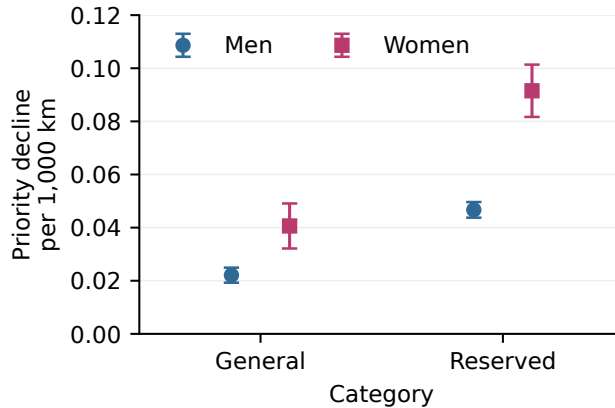
Category	Men (%)	Women (%)	F–M (pp, SE)
General	78.2	71.5	–6.7 (2.0)
Reserved	80.8	74.5	–6.3 (1.5)
All	79.6	73.0	–6.6 (1.7)

Notes: Mean candidate-level share of listed programs whose host institute is outside the candidate’s home state, computed over the portion of each submitted list that is feasible in the first round. The 25,340 candidates who list at least one such program (the candidates with a first-round seat) contribute equally; the median length of this portion is 10 programs. Standard errors clustered by home state are in parentheses.

The gender gap also appears in submitted list order. Women place farther programs lower on their lists, with a larger estimated gradient than men in both categories. Let L_i be the options on candidate i ’s submitted list that belong to the 589 programs, and let r_{ij} be program j ’s position after restricting the list to these programs. For lists with at least two entries, define

$$\text{Priority}_{ij} = 1 - \frac{r_{ij} - 1}{|L_i| - 1},$$

so that priority is one for the first listed program and zero for the last. Within each category-by-gender group, we regress priority on distance with candidate fixed effects, giving each listed candidate–program pair equal weight. The plotted gradient $-\hat{\beta}_g$ measures the decline in priority per additional 1,000 km; standard errors are clustered by candidate.

**Figure 2:** Submitted-list priority and distance, by category and gender

Note: Each point is $-\hat{\beta}_g$ from a candidate fixed effects regression of normalized submitted-list priority on distance, with 95 percent candidate-clustered confidence intervals. The regressions use 2,020,502 listed options of the 28,921 sample candidates who list at least two programs; mean list length is 69.9 programs. These are descriptive associations within submitted lists.

All four gradients in Figure 2 are positive and individually significant: candidates in both categories place farther programs lower on their lists. Women’s estimated gradients exceed men’s in both groups, with statistically significant female–male differences.

4 Preference information from first-round outcomes

Following the feasible-choice approach of [Fack et al. \(2019\)](#), we assume that a candidate’s observed first-round outcome is her preferred alternative among her feasible set $S_i^{(1)}$ and the outside option. The outcome is the program she is assigned if she accepts it, and the outside option if she does not accept it or does not receive a first-round seat. Each candidate contributes one choice observation; we use no other ordering information from her list or from later allocations. This avoids relying on rankings of unreachable programs or safety options that candidates may not have considered carefully, as noted in recent empirical literature (see [Section 1](#)). First-round assignments are a particularly good form of revealed preference information in this context because candidates got feedback from multiple mock allocations before finalizing their preference lists. When we instead treat each candidate’s submitted list as a truthful ranking of her feasible programs, the estimates of program effects and of the outside option are very different ([Section 7.4](#)).

5 Model and estimation

To compare distance sensitivity between women and men, the model allows preferences over fields of study, sensitivity to distance, and the attractiveness of the outside option to differ across four groups: general-category men, general-category women, reserved-category men, and reserved-category women. Reserved-category candidates are those in the reserved social categories, regardless of the type of seat through which a candidate is admitted; [Appendix A.4](#) gives the category definitions. Because similar majors have slightly different names at different institutes, we group programs into 19 fields of study, each combining closely related majors, such as Mechanical, Industrial and Manufacturing; [Table 11](#) lists all 19 fields.

[Section 5.1](#) specifies preferences over programs. Program effects capture each program’s overall attractiveness, including that of its institute and its field, and the value of each field may differ across groups. Each candidate’s distance coefficient is random, with a mean and dispersion that vary by group. [Section 5.2](#) specifies the outside option. Candidates have many alternatives to a JoSAA seat: India had over 1 million undergraduate engineering seats in 2015 ([All India Council for Technical Education, 2017](#)), including seats at state public colleges and many private colleges, so JoSAA seats are a small share of these, albeit among the most desirable. The value of the outside option varies with the candidate’s group, home state, and exam rank, and its unobserved component can be correlated with her distance coefficient. [Section 5.3](#) derives choice probabilities from a nested logit model in which programs in the same field are closer substitutes, and [Section 5.4](#) describes inference and the units in which we report distance sensitivity.

5.1 Program preferences and distance sensitivity

Let $g(i)$ denote candidate i 's group, $k(j)$ program j 's field of study, and $s(i)$ the candidate's home state, the state in which she passed Class XII. The utility index of program j for candidate i is

$$V_{ij}(\nu_i) = \delta_j + \theta_{g(i),k(j)} + b_i d_{ij}, \quad b_i = \mu_{g(i)} + \sigma_{g(i)} \nu_i, \quad \nu_i \sim N(0, 1). \quad (1)$$

Here d_{ij} is distance in units of 1,000 km, measured as in Section 2.5. The program effect δ_j is common to all candidates and captures the appeal of the program. Both the mean μ_g and the SD σ_g of the distance coefficient vary by group.

Let F_i indicate a woman and R_i a reserved-category candidate. The group means of the distance coefficient and the field adjustments take the form

$$\mu_{g(i)} = \beta_d + \beta_F F_i + \beta_R R_i + \beta_{FR} F_i R_i, \quad (2)$$

$$\theta_{g(i),k} = \theta_k^F F_i + \theta_k^R R_i + \theta_k^{FR} F_i R_i. \quad (3)$$

General-category men are the reference group. Since choices depend only on differences in utility, the level of utility must be fixed once overall and once for each of the other three groups. We therefore fix the program effect of Civil Engineering at IIT Bhubaneswar to zero and set the three adjustments for Computer Science and Engineering in equation (3) to zero, so that the other fields are measured relative to this field. The intercepts of the outside option below are expressed relative to the same normalization. Since b_i is normally distributed, some candidates can have a positive distance coefficient, that is, they can prefer programs farther from home.

5.2 The outside option and correlated heterogeneity

The outside option includes attending a college outside JoSAA or retaking the JEE exams in the following year. For candidates who do not receive a first-round seat, it also includes waiting for a seat in a later round. Its utility index is

$$W_i(\xi_i) = \kappa_{g(i)} + \gamma_{u(s(i))} + \alpha_{g(i)} A_i + \tau_{g(i)} \xi_i, \quad \begin{pmatrix} \nu_i \\ \xi_i \end{pmatrix} \mid g(i) = g \sim N \left[\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} 1 & \rho_g \\ \rho_g & 1 \end{pmatrix} \right]. \quad (4)$$

The group term κ_g is common to all home states, and the home-state term $\gamma_{u(s)}$ is common to all four groups. Each home state s is its own unit $u(s)$, except that the eight northeastern states form one unit and the union territories of the Andaman and Nicobar Islands, Dadra and Nagar Haveli, Daman and Diu, and Lakshadweep form another. We pool them because most have very few candidates in our sample, and in Sikkim and the Andaman and Nicobar Islands every candidate accepts a seat. We set $\gamma_u = 0$ for Rajasthan, the home state with the most candidates. Section 7.3 reports the results with a separate intercept for each group and home state.

Ability, A_i , is minus the log of the smaller of the candidate's valid JEE-Main engineering and JEE-Advanced overall ranks, centered at its sample mean; a larger value denotes a better rank. Its coefficient α_g , the SD τ_g of the unobserved component $\tau_g \xi_i$, and the correlation ρ_g all vary by

group. Appendix A.4 describes the records used for the covariates and how we treat candidates with missing Ability.

The correlation allows an unobserved taste for moving farther to be associated with a more or less attractive outside option. A positive ρ_g means that candidates with a higher distance coefficient, and hence less aversion to distance, tend to have a more attractive outside option. The SD τ_g is the marginal dispersion of the unobserved normal component; the usual choice shocks also remain in the model. The model represents the different components of the outside option jointly.

5.3 Choice probabilities and estimation

Programs are grouped into 19 nests, one for each field of study. The utility of each alternative is its utility index, V_{ij} or W_i , plus a choice shock, and the shocks have the generalized extreme value distribution of the nested logit model. Conditional on the two normal draws, define

$$D_{ik}(\nu) = \sum_{j \in S_i^{(1)}: k(j)=k} \exp\left(\frac{V_{ij}(\nu)}{\lambda_{g(i),k}}\right), \quad Z_i(\nu) = \sum_k D_{ik}(\nu)^{\lambda_{g(i),k}}. \quad (5)$$

Empty nests contribute zero. The outside option is a singleton nest. With the upper-level scale normalized to one, the nested-logit probabilities are (Train, 2009)

$$P_{i0}(\nu, \xi) = \frac{\exp W_i(\xi)}{\exp W_i(\xi) + Z_i(\nu)}, \quad (6)$$

$$P_{ij}(\nu, \xi) = \frac{\exp(V_{ij}(\nu)/\lambda_{g(i),k(j)}) D_{i,k(j)}(\nu)^{\lambda_{g(i),k(j)}-1}}{\exp W_i(\xi) + Z_i(\nu)}, \quad j \in S_i^{(1)}. \quad (7)$$

The nesting parameters are modelled as

$$\text{logit}(\lambda_{g(i),k}) = \ell_k + \eta_F F_i + \eta_R R_i + \eta_{FR} F_i R_i. \quad (8)$$

Nine larger fields have separate base parameters ℓ_k ; ten smaller fields share a base parameter but remain distinct nests. The gender and category terms allow substitution patterns to differ across groups.

Writing $y_i \in S_i^{(1)} \cup \{0\}$ for the observed choice, candidate i contributes

$$L_i = \int_{\mathbb{R}^2} P_{i,y_i}(\nu, \xi) \phi_2(\nu, \xi; \rho_{g(i)}) d\nu d\xi. \quad (9)$$

Here $\phi_2(\cdot, \cdot; \rho)$ is the standard bivariate normal density with correlation ρ . For candidates who choose the outside option, the likelihood compares the outside option with every feasible program and imposes no ranking among the programs. All parameters of the program utilities and of the outside option are estimated jointly by maximizing $\sum_i \log L_i$.

The estimates use 15 Gauss–Hermite quadrature nodes for distance and 81 nodes for the remaining independent normal draw, after writing $\xi_i = \rho_g \nu_i + \sqrt{1 - \rho_g^2} \psi_i$ with ψ_i independent standard

normal. There are 686 finite estimated parameters. Alternatives that are never chosen are evaluated at their exact limiting utilities rather than at arbitrarily large finite coefficients. Appendix A.5 describes these cases and the numerical checks.

5.4 Uncertainty and reporting scale

We cluster standard errors by home state, that is, by the 35 states and union territories in which candidates passed Class XII. We use a cluster sandwich estimator that sums candidates' likelihood scores within each state, based on the observed Hessian and with the finite-cluster correction 35/34. Standard errors of nonlinear functions of the parameters use the full joint covariance, and the 95 percent intervals we report use the t_{34} critical value. For the female-to-male ratios, we report Fieller intervals (Fieller, 1954) based on the same covariance. With only 35 clusters, a few of which account for most of the variation in the scores, t -tests based on these standard errors can reject a true null hypothesis too often (Cameron et al., 2008; Cameron and Miller, 2015). For the main comparisons of distance coefficients across groups, we therefore report p -values from the restricted wild cluster bootstrap of Kline and Santos (2012), which estimates the model once with the null hypothesis imposed and uses 99,999 draws of Webb weights over the 35 states (Webb, 2023).

We express distance coefficients in units of the pooled within-group standard deviation of distance-free program values. Let $a_{gj} = \delta_j + \theta_{g,k(j)}$, and let n_{gj} count sample candidates in group g who accept program j in the first round. Define

$$\bar{a}_g = \frac{\sum_{j:n_{gj}>0} n_{gj} a_{gj}}{\sum_j n_{gj}},$$

$$s_{\text{prog}}^2 = \frac{\sum_g \sum_{j:n_{gj}>0} n_{gj} (a_{gj} - \bar{a}_g)^2}{\sum_g \sum_j n_{gj}}.$$

This pools within-group variances, weighting each accepted seat equally. It incorporates group-specific field preferences while remaining invariant to the normalization of group utility levels. Combinations with no acceptances receive zero weight. We use the same s_{prog} to normalize all four groups' distance coefficients. Its value is 3.42642 (standard error 0.164). The delta-method standard errors of the normalized coefficients account for the uncertainty in s_{prog} and for its covariance with the distance coefficients.

Program values closely track admission selectivity. For general-category men, the reference group, a program's value equals its program effect δ_j . Figure 3 plots δ_j against the program's 2015 first-round open-seat closing rank on a logarithmic scale, for the 213 IIT and 280 non-IIT programs that general-category men accept in the first round; bubble areas and regression weights reflect these acceptances. We fit IIT and non-IIT programs separately because their cutoffs come from different entrance exams. Program effects decline approximately linearly with the log closing rank, with weighted R^2 values of 0.924 for IITs and 0.888 for non-IIT programs. Along the fitted lines, a program value one SD lower corresponds to a closing rank 3.20 times larger for IIT programs and 4.02 times larger for non-IIT programs.

One standard deviation of program value is a substantial gain to applicants. Consider candidates who spend an extra year preparing to retake JEE-Advanced, the entrance exam for the IITs. A candidate who retakes the exam starts her degree a year later, giving up a year of educational progress as well as the effort of preparing again.¹⁰ Through the relationship in Figure 3, the average rank gain of candidates in our matched sample corresponds to 0.73 SD in the value of the programs whose cutoffs they clear, and to about one SD among those whose ranks improve (Appendix F). More than a third of the candidates offered IIT seats in the first round of 2015 can be matched to 2014 entrance-exam records, suggesting that many are willing to make these sacrifices to pursue gains of this size.

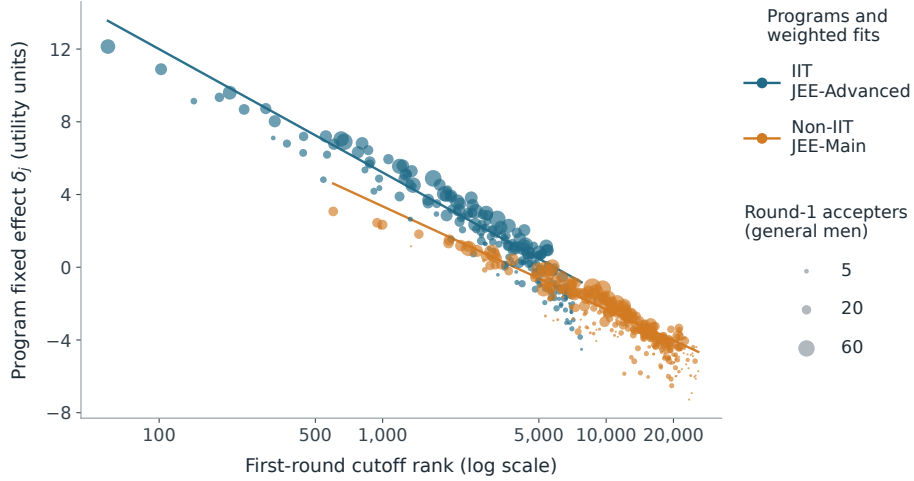


Figure 3: Program effects and first-round closing ranks

Notes: General-category men. Blue denotes 213 programs at IITs; orange denotes 280 non-IIT programs in our sample. Bubble area and regression weights are proportional to n_{gj} , the number of first-round acceptances. Lines are fitted separately: cutoffs use JEE-Advanced ranks for IITs and JEE-Main ranks otherwise. Cutoffs are ordinary open, non-PwD ranks, using other-state seats where available and all-India seats otherwise; six NIT Warangal programs with an Andhra Pradesh quota use home-state-only cutoffs. Programs with zero acceptances are not shown. The vertical axis is δ_j in utility units, relative to Civil Engineering at IIT Bhubaneswar.

6 Distance sensitivity and gender

6.1 Distance gradients and the gender gap

An additional 1,000 km lowers the utility of a program by 0.271 standard deviations (SD) of program values net of distance for the average general-category man, 0.431 for the average general-category woman, 0.329 for the average reserved-category man, and 0.564 for the average reserved-category woman (Table 3). On average, women are more sensitive to distance than men in each category, and this holds while allowing preferences over fields of study, substitution between programs, and the outside option to differ across the four groups.

¹⁰Without continued preparation, we would expect forgetting and loss of fluency to erode exam readiness. Experimental evidence on mathematical knowledge supports the role of continued practice in retention (Rohrer and Taylor, 2006).

Table 3: Distance preferences and the outside option: correlated main model

Coefficient	General		Reserved	
	Men	Women	Men	Women
Mean distance coefficient, normalized	−0.271 (0.029)	−0.431 (0.037)	−0.329 (0.023)	−0.564 (0.034)
Distance-coefficient SD, raw (σ_g)	0.655 (0.105)	0.875 (0.131)	0.964 (0.138)	1.512 (0.174)
Outside Ability coefficient (α_g)	1.011 (0.086)	0.698 (0.108)	0.017 (0.124)	0.105 (0.130)
Outside SD (τ_g)	2.129 (0.251)	2.073 (0.530)	2.985 (0.286)	2.859 (0.387)
Distance–outside correlation (ρ_g)	0.432 (0.299)	0.388 (0.441)	0.351 (0.174)	0.626 (0.131)

Standard errors clustered by home state are in parentheses. Distance is measured per 1,000 km. Normalized means divide raw distance means by the pooled within-group SD of program values net of distance, weighted by first-round acceptances ($s_{\text{prog}} = 3.42642$); their SEs propagate uncertainty in this denominator. The two reported SDs are raw utility parameters. The outside SD is marginal rather than conditional on the distance coefficient. The usual choice shocks remain in the model.

Women’s mean distance sensitivity is 1.588 times men’s among general-category candidates (95 percent interval 1.324–1.974) and 1.716 times men’s among reserved-category candidates (1.511–1.949). The difference between women and men is statistically significant in both groups: the restricted bootstrap rejects equal mean distance coefficients among general-category candidates ($p = 0.009$) and among reserved-category candidates ($p = 0.001$). The two ratios are similar, and we cannot reject that they are equal ($p = 0.51$). These ratios do not depend on the units in which the distance coefficients are expressed.

The distributions of distance sensitivity for women and men also overlap substantially. In raw utility units, the SD of the distance coefficient is 0.655 and 0.875 for general-category men and women and 0.964 and 1.512 for reserved-category men and women, compared with mean coefficients of -0.930 , -1.477 , -1.126 , and -1.932 (Table 9). The group means therefore do not imply that every woman is more averse to distance than every man.

Reserved-category candidates are also more sensitive to distance on average than general-category candidates of the same gender, although these differences are smaller than the gender gaps. Mean distance sensitivity is about 21 percent larger for reserved-category men than for general-category men (restricted bootstrap $p = 0.06$), and about 31 percent larger for reserved-category women than for general-category women ($p = 0.08$).

6.2 Field preferences and substitution

The model allows the value of each field of study to differ by gender, by category, and by their interaction. These three adjustments are estimated for each of the 18 fields other than the reference field, Computer Science and Engineering. Compared with men, women value Mechanical, Industrial and Manufacturing programs less: relative to the reference field, the difference is -2.038 among

general-category candidates and -1.542 among reserved-category candidates. They value Bioengineering and Biotechnology programs more, with corresponding differences of 0.938 and 0.967 . The gender gap in distance sensitivity therefore exists alongside differences between women and men in the fields they prefer. Programs in the same field are closer substitutes than programs in different fields: the estimated nesting parameters range from 0.270 to 0.637 . Appendix B reports the adjustments for all fields and the nesting parameters of all 19 fields for each group.

6.3 The outside option and its relationship to distance

Among general-category candidates, the outside option is more attractive for candidates with better exam ranks: the coefficient on Ability is 1.011 for men and 0.698 for women. Among reserved-category candidates, the corresponding coefficients, 0.017 and 0.105 , are close to zero. These coefficients describe the value of the outside option given a candidate’s feasible programs and the intercept for her group and home state. They do not distinguish between the components of the outside option, such as retaking the exam and attending another college.

Candidates also differ substantially in the value of their outside option beyond what these variables capture: the SD of the unobserved component ranges from 2.073 to 2.985 across groups. Its estimated correlation with the distance coefficient is 0.432 for general-category men, 0.388 for general-category women, 0.351 for reserved-category men, and 0.626 for reserved-category women. The correlations are imprecisely estimated for general-category candidates and more precisely estimated for reserved-category candidates. A positive correlation means that, given the included covariates, candidates who are less averse to distance tend to have more attractive outside options.

6.4 Interpreting the magnitudes

To put these magnitudes in perspective, consider two programs in the same field, one of which is 500 km farther from home. At the group’s mean distance coefficient, the farther program would need a program value that is larger by 0.136 SD for general-category men, 0.215 for general-category women, 0.164 for reserved-category men, and 0.282 for reserved-category women to be valued equally. This is a comparison of utilities, not of choice probabilities: how much an increase in distance changes the probability of choosing a program also depends on the candidate’s other feasible programs, her outside option, how closely programs in the same field substitute for one another, and her individual tastes. Section 6.5 translates the gender difference in distance sensitivity into predicted choice distance and attainable value.

Distance sensitivity can reflect travel costs, housing, safety, information, family constraints, and tastes. Our estimates capture the combined influence of these channels and do not separate them.

6.5 Distance and the gender gap in attainable value

A candidate’s eligibility determines which programs she can attain; how much these programs are worth to her also depends on how far away they are. To compare this value across the four groups, we assume that the same Computer Science and Engineering program has the same value, net of

distance, for all of them. The resulting comparison reflects differences across groups in feasible sets, field preferences, and substitution patterns.

For candidate i , let $\mathcal{I}_i(\nu) = \log Z_i(\nu)$ be the nested-logit index of her feasible set, with $Z_i(\nu)$ given by equation (5). We define her attainable value as

$$\mathcal{M}_i = \frac{\mathbb{E}_\nu[\mathcal{I}_i(\nu)] - \theta_{g(i),\text{CS}}}{s_{\text{prog}}},$$

where subtracting her group’s Computer Science adjustment imposes the assumption above and s_{prog} is the SD of program values from Section 5.4. Under the normalization in Section 5.1, the Computer Science adjustment is zero for every group, so this term drops out of the estimates. Attainable value excludes the outside option, so the outside-option intercepts and heterogeneity do not enter it, and the expectation is taken over the normal distribution of ν . A constant common to all groups cancels from the gender gap; the Computer Science assumption fixes the group-specific constants, which are otherwise not identified. We interpret $\mathcal{M}_i$ as an index of access adjusted for distance, rather than as a measure of earnings or educational quality.

In the counterfactual, we set the mean and standard deviation of women’s distance coefficient equal to those of men in the same category. We hold women’s feasible sets, field preferences, and nesting parameters fixed. We therefore do not recompute assignments, which would change through competition for seats if women’s choices changed; the exercise isolates the role of distance preferences given the programs women can attain. If $\bar{\mathcal{M}}_M$ and $\bar{\mathcal{M}}_W$ denote the baseline means for men and women, the share of the gap that is closed is

$$\frac{\bar{\mathcal{M}}_W^{\text{cf}} - \bar{\mathcal{M}}_W}{\bar{\mathcal{M}}_M - \bar{\mathcal{M}}_W}. \quad (10)$$

Overall means use each gender’s observed category composition. We also calculate the expected distance of the program a candidate chooses, by integrating her choice probabilities within the feasible set, $P_{ij}/(1 - P_{i0})$, over ν .

Table 4 reports the results. At baseline, men’s attainable value exceeds women’s by 0.398 SD. If women were as sensitive to distance as men in the same category, their attainable value would rise by 0.098 SD, closing 24.6 percent of the gap (95 percent Fieller interval 19.3–29.6 percent). The share is 19.3 percent among general-category candidates and 30.3 percent among reserved-category candidates; the difference between the two is not statistically significant. Women’s predicted choice distance would rise by 115 km overall, by 119 km for general-category women and by 111 km for reserved-category women.

The comparison of values across groups depends on the Computer Science assumption. If we instead assume that Electrical, Electronics and Instrumentation programs have the same value net of distance across groups, the share of the gap closed is 21.0 percent overall, 17.6 percent among general-category candidates, and 24.1 percent among reserved-category candidates. This alternative changes the baseline gap but not the increase in women’s attainable value or their predicted choices. Rescaling women’s distance coefficients proportionally to match men’s mean,

instead of replacing the whole distribution, closes 23.2 percent of the overall gap under the Computer Science assumption.

Table 4: Distance sensitivity and the gender gap in attainable value

	All	General	Reserved
Initial attainable-value gap (SD)	0.398 [0.356, 0.439]	0.436 [0.386, 0.486]	0.366 [0.317, 0.416]
Increase in women’s value (SD)	0.098 [0.073, 0.123]	0.084 [0.046, 0.122]	0.111 [0.076, 0.147]
Attainable-value gap closed (%)	24.6 [19.3, 29.6]	19.3 [10.8, 27.9]	30.3 [22.8, 36.9]
Increase in women’s choice distance (km)	115 [82, 148]	119 [65, 173]	111 [74, 147]

The same Computer Science and Engineering program is assumed to have the same value, net of distance, for all four groups. Both the mean and SD of women’s distance coefficient are replaced by those of men in the same category; feasible sets, field preferences, and nesting parameters remain fixed. The attainable value is the expected nested-logit index of the first-round feasible programs, excluding the outside option. The gap is men minus women. Brackets contain 95 percent intervals using the full covariance clustered by home state and t_{34} critical values: Fieller intervals for shares and delta-method intervals otherwise. SD denotes the pooled within-group standard deviation of program values net of distance, weighted by first-round acceptances. Overall means use each gender’s observed category composition.

7 Robustness

This section reports how women’s mean distance sensitivity relative to men’s changes when we expand the sample (Section 7.1), add Class XII board marks to the outside option (Section 7.2), and vary the sample, the distance measure, the specification, and the treatment of first-round acceptances (Section 7.3). Section 7.4 compares our estimates with those based on the full submitted lists.

7.1 Expanding the sample

We widen the rank cutoffs that define the sample so that it covers every candidate in the scope of the study who received a first-round seat. The expanded sample includes all candidates who submitted a preference list and clear the expanded cutoffs, and it retains the exclusions in Appendix A.2. For each of the eight rank lists, the expanded cutoff is the largest rank on that list among the candidates who received a first-round seat, remain after these exclusions, and whose seat came from that list’s group of institutes and type of seat. Candidates who received a PwD seat are counted at their rank on the corresponding non-PwD list when they have one; the others clear the cutoff of another list. The JEE-Main cutoffs in Table 6 become 1,178,563, 367,599, 116,772, and 45,591 for the CRL, OBC-NCL, SC, and ST lists, and the JEE-Advanced cutoffs become 20,723, 5,434, and 2,537 for the CRL, OBC-NCL, and SC lists; the JEE-Advanced ST cutoff stays at 755.

The expanded cutoffs therefore cover 140,239 of the 142,709 candidates who submitted a preference list. The exclusions remove 1,984 candidates who received a seat in a preparatory course or in an architecture or planning program, 168 whose recorded category changed after the first round, and 34 foreign nationals. The resulting sample has 138,053 candidates: 25,808 accepted a first-round seat, 6,734 were offered a first-round seat but did not accept it, and 105,511 did not receive one. Of the latter, 92,149 have no feasible program in the first round and do not contribute to the likelihood. One other candidate also does not contribute: a reserved-category woman whose only feasible programs are in Pharmacy, a field in which no reserved-category woman accepts a seat. The estimates therefore use 45,903 candidates.

When we reestimate the model on this sample, women’s mean distance sensitivity is 1.498 times men’s among general-category candidates (95 percent interval 1.308–1.727) and 1.679 times men’s among reserved-category candidates (1.515–1.862), compared with 1.588 and 1.716 in the main sample. Women remain more sensitive to distance than men on average in both categories.

7.2 Board marks in the outside option

We add the Class XII board mark to the utility of the outside option, with a separate coefficient for each of the four groups. Like the other covariates, the mark is taken from the third-round records; it is centered at its sample mean and measured in units of 10 points. The nine candidates without a positive recorded mark receive an indicator instead. All four coefficients are positive: a higher board mark raises the value of the outside option, given a candidate’s feasible programs, the intercept for her group and home state, and Ability. Women’s mean distance sensitivity is 1.589 times men’s among general-category candidates and 1.720 times men’s among reserved-category candidates, compared with 1.588 and 1.716 in the main specification.

7.3 Other sensitivity analyses

In Appendix A.2, we exclude 111 candidates whose recorded category changed after the first round. When we reestimate the model keeping these candidates in their third-round groups, women’s mean distance sensitivity is 1.600 times men’s among general-category candidates and 1.716 times men’s among reserved-category candidates, compared with 1.588 and 1.716 in the main sample. Using the first-round records of home state and overall ranks instead of the third-round records gives ratios of 1.587 and 1.714; the two sets of records differ in home state for 73 candidates and in Ability for 116.

In Section 2.5, we measure distance from the centroid of the home state for candidates who took the exam outside their home state. If we instead use the examination city for all candidates who took the exam in India, women’s mean distance sensitivity is 1.552 times men’s among general-category candidates and 1.835 times men’s among reserved-category candidates, compared with 1.588 and 1.716 in the main specification. Measuring distance from the home-state centroid for all candidates gives ratios of 1.604 and 1.652.

We also add interactions of distance with an indicator for IIT programs, with Ability, and with their product; these interactions are common to all four groups. At mean Ability, women’s mean distance sensitivity is then 1.353 times men’s among general-category candidates and 1.499 times men’s among reserved-category candidates for other programs, and 1.562 and 1.736 times men’s for IIT programs. Distance matters less for IIT programs, and the coefficients on the terms involving Ability are close to zero. Women remain more sensitive to distance than men on average in both categories and for both types of program.

In Section 5.2, we model the intercept of the outside option as the sum of a group term and a home-state term. When we instead estimate a separate intercept for each combination of group and home state, women’s mean distance sensitivity is 1.646 times men’s among general-category candidates (95 percent interval 1.306–2.085) and 1.750 times men’s among reserved-category candidates (1.528–1.996), compared with 1.588 and 1.716 in the main specification.

A candidate who is offered a first-round seat must accept it to remain in the allocation, and candidates who choose *Float* or *Slide* keep the chance of an upgrade. Accepting a first-round seat may therefore partly reflect the value of this option. When we estimate the model only on the 21,018 candidates who accepted a first-round seat, without the outside option, women’s mean distance sensitivity is 1.561 times men’s among general-category candidates and 1.833 times men’s among reserved-category candidates, compared with 1.588 and 1.716 in the main specification. When we remove from every feasible set the 179 programs at which more than 30 percent of the candidates assigned an ordinary non-PwD seat in the first round did not accept it, and drop the 2,111 candidates who accepted a seat in one of these programs, the ratios are 1.676 and 1.791. Of these 2,111 candidates, 79 percent chose *Float*, compared with 58 percent of the other candidates who accepted a seat.

In every check in Sections 7.1–7.3, women remain more sensitive to distance than men on average in both categories. The female-to-male ratio ranges from 1.353 to 1.676 among general-category candidates and from 1.499 to 1.835 among reserved-category candidates.

7.4 Using the full preference lists

We also estimate the model treating each candidate’s submitted list as a truthful ranking of her feasible programs, as in the Plackett–Luce model (Luce, 1959; Plackett, 1975). Besides the first-round outcome, the likelihood then uses the order of the feasible programs a candidate listed below her seat, or of all those she listed if she did not accept it. Each listed program is treated as a nested-logit choice among the remaining feasible programs, listed or not. These choices exclude the outside option and use the same distance coefficient b_i as the first-round outcome. Without nests, they follow from the Plackett–Luce model once the unobserved position of the outside option in the ranking is summed out. With nests, they are a modeling assumption. The lists add 439,481 such choices, and we reestimate all parameters.

The SD of program effects falls from 3.574 to 1.346 and the SD of program values from 3.426 to 1.264 (Table 5). The rank correlation between the two sets of program effects is 0.856. Mean distance sensitivity falls by 47 to 60 percent in raw units but rises in all four groups once normalized,

Table 5: Estimates from first-round outcomes and from full preference lists

	General		Reserved	
	Men	Women	Men	Women
SD of program effects				
Main model		3.574		
Full lists		1.346		
SD of program values (s_{prog})				
Main model		3.426		
Full lists		1.264		
Mean distance coefficient, raw				
Main model	−0.930	−1.477	−1.126	−1.932
Full lists	−0.419	−0.783	−0.449	−0.792
Mean distance coefficient, normalized				
Main model	−0.271	−0.431	−0.329	−0.564
Full lists	−0.331	−0.619	−0.355	−0.627
Women’s mean distance sensitivity relative to men’s				
Main model	1.588		1.716	
	[1.324, 1.974]		[1.511, 1.949]	
Full lists	1.870		1.764	
	[1.511, 2.382]		[1.655, 1.894]	
Outside Ability coefficient (α_g)				
Main model	1.011	0.698	0.017	0.105
Full lists	−0.148	−0.215	−0.145	−0.076
Outside SD (τ_g)				
Main model	2.129	2.073	2.985	2.859
Full lists	0.543	0.452	0.395	0.453
Distance–outside correlation (ρ_g)				
Main model	0.432	0.388	0.351	0.626
Full lists	1.000	1.000	1.000	1.000

Both sets of estimates use the same 29,024 candidates and feasible sets and estimate the same parameters. The SD of program effects is taken over the 572 finite program effects. Distance is measured per 1,000 km. Normalized means divide raw means by the SD of program values from the same estimates (Section 5.4). Brackets contain 95 percent Fieller intervals based on standard errors clustered by home state.

because the SD of program values falls by 63 percent. Program effects and mean distance coefficients are thus both much smaller relative to the choice shocks, as expected if candidates rank programs lower on their lists with more noise (Hausman and Ruud, 1987). Women’s mean distance sensitivity is 1.870 times men’s among general-category candidates (95 percent interval 1.511–2.382) and 1.764 times men’s among reserved-category candidates (1.655–1.894), compared with 1.588 and 1.716 at the main estimates. Women remain more sensitive to distance than men on average in both categories. Mean distance sensitivity is 7 percent larger for reserved-category men than for general-category men and 1 percent larger for reserved-category women than for general-category women, compared with 21 and 31 percent.

In the outside option, the Ability coefficient becomes negative in all four groups, the SD of the unobserved component falls to between 0.395 and 0.543, and its correlation with the distance coefficient is one in every group. At these estimates, the log likelihood of the first-round outcomes

is 9,067 lower than at the main estimates. Of this difference, 337 comes from the choice between a program and the outside option and the rest from the choice of program by the candidates who accepted a seat.

8 Conclusion

This paper shows that distance is a first-order cost in where students choose to study in India, even among candidates competing for seats at the country’s most selective engineering institutes. Using administrative data from India’s 2015 centralized engineering admissions, we find that for a typical male candidate, avoiding an extra 1,000 km move is worth roughly a quarter to a third of a standard deviation of program values. Women are about 1.6 to 1.7 times as sensitive to distance as men. Holding attainable programs fixed, distance preferences account for 25 percent of the female–male access gap.

These estimates rest on a substantially weaker assumption than treating the full submitted list as truthful. Our identification strategy assumes only that each candidate chooses her most preferred option among the programs she could attain and an outside option; it uses no other ordering information from her list.

Our estimates likely understate the cost of distance for students more broadly. First, our sample consists of candidates who invested heavily in exam preparation and cleared a highly competitive exam. Migration costs are likely higher for students from more disadvantaged backgrounds or with less academic preparation. Second, students who find distance most costly may not take part in the allocation at all. Our data cover only candidates who participate, so they miss those deterred from applying, whose distance costs are likely the highest.

Distance is likely a proxy for several barriers, including safety concerns, parental preferences, and social norms around women’s mobility. Our data do not allow us to distinguish among these barriers or to identify why they bind more strongly for women. To the extent that gendered distance costs reflect constraints rather than preferences over location itself, policies that lower the difficulty of moving could allow more women to enroll in better programs, and would let institutes attract high-achieving students from a wider geographical area.

References

- Atila Abdulkadiroğlu, Nikhil Agarwal, and Parag A Pathak. The welfare effects of coordinated assignment: Evidence from the New York City high school match. *American Economic Review*, 107(12):3635–3689, 2017.
- Riley Acton, Kalena E. Cortes, Lois Miller, and Camila Morales. Distance to degrees: How college proximity shapes students’ enrollment choices and attainment across race-ethnicity and socioeconomic status. EdWorkingPaper 24-1055, Annenberg Institute, Brown University, 2024. <https://doi.org/10.26300/vjyg-ta27>.
- Nikhil Agarwal and Paulo Somaini. Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444, 2018.
- Nikhil Agarwal and Paulo Somaini. Revealed preference analysis of school choice models. *Annual Review of Economics*, 12:471–501, 2020.
- Fabiola M. Alba-Vivar. Opportunity bound: Transport and access to college in a megacity. Working Paper 128, Wake Forest University, Department of Economics, 2025.
- All India Council for Technical Education. Annual report 2015–16. Technical report, All India Council for Technical Education, New Delhi, 2017. Section 3.3. <https://web.archive.org/web/20170918120747/http://www.aicte-india.org/downloads/annual%20report%202015-16.pdf>.
- James Alm and John V. Winters. Distance and intrastate college student migration. *Economics of Education Review*, 28(6):728–738, 2009.
- Surender Baswana, Partha Pratim Chakrabarti, Sharat Chandran, Yashodhan Kanoria, and Utkarsh Patange. Centralized admissions for engineering colleges in India. *INFORMS Journal on Applied Analytics*, 49(5):338–354, 2019. doi: 10.1287/inte.2019.1007.
- Barbara Boelmann. Women’s missing mobility and the gender gap in higher education: Evidence from Germany’s university expansion. CRC TR 224 Discussion Paper 518, University of Bonn and University of Mannheim, 2024. URL <https://www.crctr224.de/research/discussion-papers/archive/dp518>.
- Girija Borker. Safety first: Perceived risk of street harassment and educational choices of women. Policy Research Working Paper 9731, World Bank, 2021.
- Caterina Calsamiglia, Chao Fu, and Maia Güell. Structural estimation of a model of school choices: The Boston mechanism versus its alternatives. *Journal of Political Economy*, 128(2):642–680, 2020.
- A Colin Cameron and Douglas L Miller. A practitioner’s guide to cluster-robust inference. *Journal of Human Resources*, 50(2):317–372, 2015.
- A Colin Cameron, Jonah B Gelbach, and Douglas L Miller. Bootstrap-based improvements for inference with clustered errors. *Review of Economics and Statistics*, 90(3):414–427, 2008.
- David Card. Using geographic variation in college proximity to estimate the return to schooling. In Louis N Christofides, E Kenneth Grant, and Robert Swidinsky, editors, *Aspects of Labour Market Behaviour: Essays in Honour of John Vanderkamp*, pages 201–222. University of Toronto Press, 1995.
- Central Board of Secondary Education. Joint entrance examination (main) 2015: Information bulletin, 2015. Appendix 2; official bulletin archived by the [Internet Archive](https://www.archive.org/details/central-board-of-secondary-education-joint-entrance-examination-main-2015-information-bulletin).
- Ali Cheema, Asim Ijaz Khwaja, M. Farooq Naseer, and Jacob N. Shapiro. Glass Walls: Experimental evidence on constraints faced by women in accessing valuable skilling opportunities. *Journal of Political Economy*, 134(6):1775–1845, 2026. doi: 10.1086/739830.
- Emil Chrisander and Andreas Bjerre-Nielsen. Why do students lie and should we worry? An analysis of non-truthful reporting. Technical Report 2302.13718, arXiv, 2023.
- Janet Currie and Enrico Moretti. Mother’s education and the intergenerational transmission of human capital: Evidence from college openings. *The Quarterly Journal of Economics*, 118(4):1495–1532, 2003.
- Judith Delaney and Paul J. Devereux. The effect of distance to colleges on application behaviour. *Applied Economics Letters*, 2024. doi: 10.1080/13504851.2024.2407996.
- Merve Demirel. How far would you go? Higher education and migration decisions in India. Working paper, Stockholm University, 2025. <https://doi.org/10.2139/ssrn.5174906>.

- Merve Demirel. Peer networks and educational migration: Experimental evidence from India. Working paper, Institute for International Economic Studies, Stockholm University, 2026. URL <https://www.demirelmerve.com/>.
- Ahmed Elsayed and Alina Shirshikova. The women-empowering effect of higher education. *Journal of Development Economics*, 163:103101, 2023. doi: 10.1016/j.jdeveco.2023.103101. URL <https://doi.org/10.1016/j.jdeveco.2023.103101>.
- Anaïs Fabre. The geography of higher education and spatial inequalities. Working paper, revised March 2026, 2026. URL <https://www.anaisfabre.com/>.
- Gabrielle Fack, Julien Grenet, and Yinghua He. Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529, 2019. doi: 10.1257/aer.20151422. URL <https://doi.org/10.1257/aer.20151422>.
- Erica Field and Kate Vyborny. Women’s mobility and labor supply: Experimental evidence from Pakistan. ADB Economics Working Paper Series 655, Asian Development Bank, 2022. <https://doi.org/10.22617/WPS220166-2>.
- Edgar C Fieller. Some problems in interval estimation. *Journal of the Royal Statistical Society, Series B*, 16(2):175–185, 1954.
- Marc Frenette. Do universities benefit local youth? Evidence from the creation of new universities. *Economics of Education Review*, 28(3):318–328, 2009.
- Chao Fu, Junjie Guo, Adam J. Smith, and Alan T. Sorensen. Students’ heterogeneous preferences and the uneven spatial distribution of colleges. *Journal of Monetary Economics*, 129:49–64, 2022. doi: 10.1016/j.jmoneco.2022.03.001. URL <https://doi.org/10.1016/j.jmoneco.2022.03.001>.
- Devaki Ghose. Trade, internal migration, and human capital: Who gains from India’s IT boom? *The Review of Economics and Statistics*, 2026. doi: 10.1162/REST.a.1734. Forthcoming; author manuscript dated January 27, 2026.
- Government of Rajasthan. Industrial potential survey 2015–2016: Kota. Technical report, Department of Industries, District Industries Centre, Kota, 2015. p. 14. [Government survey](#).
- Avinatan Hassidim, Assaf Romm, and Ran I Shorrer. The limits of incentives in economic matching procedures. *Management Science*, 67(2):951–963, 2021. doi: 10.1287/mnsc.2020.3591.
- Jerry A Hausman and Paul A Ruud. Specifying and testing econometric models for rank-ordered data. *Journal of Econometrics*, 34(1-2):83–104, 1987.
- Caroline Hoxby and Christopher Avery. The missing “one-offs”: The hidden supply of high-achieving, low-income students. *Brookings Papers on Economic Activity*, 2013(1):1–65, 2013.
- Indian Institute of Technology Bombay. JEE (Advanced) 2015: Results at a glance, 2015. June 17; organizing institute’s summary, archived by the [Internet Archive](#).
- Shoya Ishimaru. Geographic mobility of youth and spatial gaps in local college and labor market opportunities. *Journal of Labor Economics*, 43(4):1251–1294, 2025. doi: 10.1086/730542. URL <https://doi.org/10.1086/730542>.
- Maulik Jagnani and Gaurav Khanna. The effects of elite public colleges on primary and secondary schooling markets in India. *Journal of Development Economics*, 146:102512, 2020. doi: 10.1016/j.jdeveco.2020.102512.
- Joint Implementation Committee. JEE (Advanced) 2016: Joint implementation committee report, volume I. Technical report, Indian Institute of Technology Guwahati, 2016. Section 14.3, Parents’ Income. <https://jeeadv.ac.in/reports/2016.pdf>.
- Joint Seat Allocation Authority. [Business Rules for Joint Seat Allocation, Academic Year 2015–16](#). Technical report, Joint Seat Allocation Authority, 2015. Rules dated 27 June 2015, in the official 2015 archive. <https://jeeadv.ac.in/reports/2015.pdf>.
- Adam J Kapor, Christopher A Neilson, and Seth D Zimmerman. Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315, 2020.
- Noemí Katzkowicz, Victor Lavy, Martina Querejeta, and Tatiana Rosá. Human capital consequences of expanding the supply of higher education institutions in a developing country. *Journal of Public Economics*, 260, 2026. doi: 10.1016/j.jpubeco.2026.105706.

- Gaurav Khanna. Does affirmative action incentivize schooling? Evidence from India. *Review of Economics and Statistics*, 102(2):219–233, 2020.
- Patrick Kline and Andres Santos. A score based approach to wild bootstrap inference. *Journal of Econometric Methods*, 1(1):23–41, 2012. doi: 10.1515/2156-6674.1006. URL <https://doi.org/10.1515/2156-6674.1006>.
- Zovanga L. Kone, Maggie Y. Liu, Aaditya Mattoo, Caglar Ozden, and Siddharth Sharma. Internal borders and migration in India. *Journal of Economic Geography*, 18(4):729–759, 2018. doi: 10.1093/jeg/lbx045.
- Thomas Le Barbanchon, Roland Rathelot, and Alexandra Roulet. Gender differences in job search: Trading off commute against wage. *Quarterly Journal of Economics*, 136(1):381–426, 2021.
- Bridget Terry Long. How have college decisions changed over time? An application of the conditional logistic choice model. *Journal of Econometrics*, 121(1-2):271–296, 2004.
- R Duncan Luce. *Individual Choice Behavior*. Wiley, 1959.
- Kaivan Munshi and Mark Rosenzweig. Networks and misallocation: Insurance, migration, and the rural-urban wage gap. *American Economic Review*, 106(1):46–98, 2016. doi: 10.1257/aer.20131365.
- Karthik Muralidharan and Nishith Prakash. Cycling to school: Increasing secondary school enrollment for girls in India. *American Economic Journal: Applied Economics*, 9(3):321–350, 2017.
- Office of the Registrar General & Census Commissioner, India. Census of India 2011: Table D-03, migrants by place of last residence, duration of residence and reason for migration, 2011a. URL <https://censusindia.gov.in/nada/index.php/catalog/10840>. Government of India.
- Office of the Registrar General & Census Commissioner, India. Census of India 2011: Table C-14, population in five year age-group by residence and sex, 2011b. URL <https://censusindia.gov.in/nada/index.php/catalog/1541>. Government of India.
- Office of the Registrar General & Census Commissioner, India. Census of India 2011: Table D-05, migrants by place of last residence, age, sex, reason for migration and duration of residence, 2011c. URL <https://censusindia.gov.in/nada/index.php/catalog/11030>. Government of India.
- Robin L Plackett. The analysis of permutations. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 24(2):193–202, 1975.
- Doug Rohrer and Kelli Taylor. The effects of overlearning and distributed practise on the retention of mathematics knowledge. *Applied Cognitive Psychology*, 20(9):1209–1224, 2006. doi: 10.1002/acp.1266.
- Ran I Shorrer and Sándor Sóvágó. Dominated choices in a strategically simple college admissions environment. *Journal of Political Economy Microeconomics*, 1(4):781–807, 2023. doi: 10.1086/726226.
- Ran I Shorrer and Sándor Sóvágó. Dominated choices under deferred acceptance mechanism: The effect of admission selectivity. *Games and Economic Behavior*, 144:167–182, 2024. doi: 10.1016/j.geb.2024.01.002.
- Ran I Shorrer, Assaf Romm, and Avinatan Hassidim. Explaining the Too-Good-To-Be-True puzzle in two-sided matching. Working paper, January 7, 2026, 2026. URL https://assafromm.weebly.com/uploads/2/4/4/5/24450164/2g2btrue_srh.pdf.
- C. Katharina Spiess and Katharina Wrohlich. Does distance determine who attends a university in Germany? *Economics of Education Review*, 29(3):470–479, 2010. doi: 10.1016/j.econedurev.2009.10.009.
- Kenneth E Train. *Discrete Choice Methods with Simulation*. Cambridge University Press, 2nd edition, 2009.
- Matthew D Webb. Reworking wild bootstrap-based inference for clustered errors. *Canadian Journal of Economics*, 56(3):839–858, 2023. doi: 10.1111/caje.12661. URL <https://doi.org/10.1111/caje.12661>.

A Sample construction and first-round feasible sets

A.1 Rank cutoffs and seat capacities

We define the sample using eight fixed rank cutoffs, one for each combination of exam (JEE-Main or JEE-Advanced) and non-PwD rank list (CRL, OBC-NCL, SC, or ST). A candidate qualifies if she submitted a preference list and her rank is no larger than the cutoff in Table 6 on at least one list that applies to her. Ties are included. CRL ranks are matched to open seats, while OBC-NCL, SC, and ST ranks are matched to the corresponding reserved seats. Qualifying through any one list is sufficient, and we add no further check on birth category. A candidate does not need to receive or accept a first-round seat.

The cutoffs are set to match the total number of non-PwD seats. For each JEE-Advanced list, we count the candidates who submitted a preference list in rank order, after setting aside those assigned a non-IIT seat in the first round; for each JEE-Main list, we instead set aside those assigned an IIT seat. The cutoff is the rank at which this count first reaches the corresponding seat total, including all ties. Preparatory seats count as IIT seats here, so candidates assigned a preparatory seat in the first round are set aside from the JEE-Main counts; those who receive a preparatory seat only in a later round remain in these counts. Candidates who clear a cutoff and receive a seat in the other group of institutes remain eligible for the sample; in particular, being assigned an IIT seat does not prevent a candidate from qualifying through a JEE-Main rank. The direct JEE-Advanced ST list ends before its seat total is reached, so its last rank is the cutoff. For the JEE-Advanced lists, the seat total includes IIT Architecture because these programs also use JEE-Advanced ranks. For the JEE-Main lists, it excludes non-IIT Architecture and Planning, which use separate Architecture rank lists. The cutoffs are set separately for each list, so a candidate can be counted toward more than one seat total before we combine the candidates who clear the eight cutoffs. The candidate exclusions below are applied only after these eight cutoffs have been fixed.

Table 6: Fixed rank cutoffs for sample inclusion

Rank list	Non-PwD seats	Cutoff rank
Advanced CRL	4,900	5,086
Advanced OBC-NCL	2,619	3,290
Advanced SC	1,454	1,637
Advanced ST	733	755
Main CRL	11,473	19,370
Main OBC-NCL	5,616	9,010
Main SC	3,325	5,987
Main ST	1,956	3,359

Notes: A candidate is included if any of her valid non-PwD ranks clears its cutoff. The Advanced ST list ends at rank 755, with 673 counted candidates, 60 short of its 733-seat total. All sample exclusions are applied after the cutoffs are set.

A.2 Exclusions and sample accounting

The cutoffs are cleared by 29,357 candidates who submitted a preference list. We exclude 70 candidates who were assigned a seat in a preparatory course, identified by flag P, in any observed round, and 121 additional candidates who were assigned a seat in an architecture or planning program. These exclusions apply whether or not the candidate accepted the seat, and they include the special round (Round 5). They are endogenous exclusions, because they are based on realized assignments; we make them because our model of engineering program choice does not cover these seats. Together they remove 191 candidates (0.65 percent of the candidates who clear a cutoff). We also exclude 111 candidates who were recorded in a reserved social category in the first round but in the general category in the third round (109 OBC-NCL, one SC, and one ST). All of them accepted a first-round seat. Their first-round feasible sets include reserved seats, whereas their group in the model follows the third-round record. Finally, we exclude 31 foreign nationals, who could be assigned only IIT seats under separate rules ([Joint Seat Allocation Authority, 2015](#), rules 25–28), so the feasible sets in Appendix A.3 do not describe their opportunities. These two exclusions remove a further 142 candidates (0.48 percent of the candidates who clear a cutoff). The resulting sample contains 29,024 candidates (Table 7): 1,246 qualify only through a JEE-Advanced list, 19,614 only through a JEE-Main list, and 8,164 through both.

Of these candidates, 21,018 accepted a first-round seat, 4,322 were offered a first-round seat but did not accept it, and 3,684 did not receive one. The latter two groups account for the 8,006 choices of the outside option (Table 8). Accepting a seat includes *Float*, *Slide*, and *Freeze*. Later participation decisions are not used as additional observations of preferences.

Table 7: Construction and outcomes of the main sample

Sample step or outcome	Candidates
Candidates who submitted a preference list	142,709
Clearing at least one fixed rank cutoff	29,357
Less: assigned a preparatory seat in any observed round	70
Less: additionally assigned an architecture or planning seat	121
Less: category revised from reserved to general after the first round	111
Less: foreign nationals	31
Main sample	29,024
Accepted a first-round seat	21,018
Offered a first-round seat but did not accept	4,322
No first-round seat	3,684

Notes: The architecture and planning exclusion covers 116 additional candidates assigned such a seat in the regular rounds and five assigned one only in the special round. Four other candidates assigned an architecture seat were already excluded because they were assigned a preparatory seat. Later rounds are used for these exclusions, not as observations of preferences.

Table 8: First-round choices by group

Group	Candidates	Accepted	Outside	Outside share
General-category men	11,208	7,973	3,235	28.86%
General-category women	2,363	1,616	747	31.61%
Reserved-category men	13,049	9,794	3,255	24.94%
Reserved-category women	2,404	1,635	769	31.99%
Total	29,024	21,018	8,006	27.58%

A.3 Reconstructing first-round feasible sets

We consider 589 science and engineering programs: 214 at IITs and 375 at non-IIT institutes, grouped into 19 fields of study. Architecture and Planning are excluded; Naval Architecture remains included.

For each candidate, we reconstruct the full set $S_i^{(1)}$ of programs that are feasible for her in the first round, using her ranks, the types of seat that apply to her, geographic quotas, and administrative validity rules. This set includes feasible programs that she did not list. Although inclusion in the sample requires a non-PwD rank, candidates in the sample can use every type of seat for which they qualify, including PwD and defence-supernumerary seats.

Whether a program is feasible for a candidate is determined from her first-round ranks and the closing ranks of the corresponding types of seat, including home-state eligibility and restrictions on program validity. A candidate can qualify through any valid rank that applies to her; we add no second check on birth category to the rank test. Non-PwD ranks determine inclusion in the sample, while all relevant types of seat determine the feasible set. The sample therefore retains 44 candidates with PwD status, 42 of whom accepted a seat, and their feasible sets include the PwD seats that apply to them. For defence-supernumerary seats, eligibility uses the institute-level closing ranks of these seats for candidates with the corresponding eligibility flag and a valid JEE-Advanced overall rank.

We additionally apply program-specific gender and medical restrictions using the first-round eligibility records. Women are ineligible for the three Mining-related programs at ISM Dhanbad, the four engineering programs at Gurukula Kangri Vishwavidyalaya, and Mining Engineering at NIT Raipur. Candidates with the one-eyed-vision flag are ineligible for the eight IIT Mining and Mining Machinery degree variants covered by the corresponding rule. The disability restriction for Mining Engineering at NIT Raipur is applied specifically to that program.¹¹

After applying these restrictions, the median feasible set contains 286 programs, and the median number of feasible institutes is 63. Every candidate has at least 20 feasible programs across six institutes before the limiting utilities described below are applied. Every first-round seat offered to

¹¹The ISM Dhanbad gender restriction and the one-eyed-vision restriction appear directly in the [JEE Advanced 2015 information brochure](#), Sections 27 and 26 (printed pages 30 and 29). For Gurukula Kangri and NIT Raipur, 2015 applicability is inferred from the [CSAB 2014 information brochure](#), Table 4 (printed pages 38 and 37), and the 2015 administrative choice and assignment records; a March 2016 [Gurukula Kangri newsletter](#) (page 8) also describes its four engineering programs as male-only. Direct 2015 restriction text was not located for these latter rules.

a candidate in the sample is feasible for her. Conversely, no program that a candidate lists above the seat she is offered is feasible for her, and no candidate without a first-round seat lists a feasible program. Candidates who choose the outside option keep their full feasible sets, and we infer no ranking among the programs in them.

A.4 Outcomes and covariates

A candidate is recorded as accepting a program if she chooses *Float*, *Slide*, or *Freeze* in the first round. Candidates who are offered a seat but do not accept it, and candidates who do not receive a seat, are recorded as choosing the outside option. Later participation decisions are not used as additional observations of preferences.

The reserved-category group combines OBC-NCL, SC, and ST eligibility; a candidate's group does not depend on the type of seat she accepts. Inclusion in the sample and feasible sets use first-round ranks and eligibility. Gender and category, the home state, and Ability use the third-round records. The mean of Ability before centering, among candidates with a valid overall rank, is -9.4500422 . Two ST candidates who accepted a seat have no overall rank on either exam in any of the four regular rounds. Both remain in the sample. They share an indicator for missing Ability in the outside option, which is at its exact limit of negative infinity because both accepted a seat; this leaves the information in their choice of program intact. We do not impute any overall rank. Candidates' origins follow the location rule in Section 2.5; the home state identifies where the candidate passed Class XII, not necessarily where she currently lives.

A.5 Likelihood boundaries and numerical checks

Seventeen programs are feasible for at least one candidate but are never chosen. Their program effects tend to $-\infty$, and in the limit they contribute zero probability to the likelihood. No reserved-category woman accepts a Pharmacy program either, so the corresponding field adjustment is also evaluated at its limit. These limits leave 572 finite program effects, one of which is the reference program, whose effect is zero.

Every group and every unit u (Section 5.2) contains both candidates who accept a seat and candidates who choose the outside option, and all intercepts of the outside option are finite. The two candidates with missing Ability who accepted a seat give one further limit, for the missing-Ability indicator. They still contribute to the likelihood through their choice among programs. No candidate is dropped from the sample, and the estimates use all 29,024 candidates.

At the main estimates, the largest absolute element of the score is 5.75×10^{-8} , and the observed Hessian is positive definite over the 686 finite parameters. No parameter is at a numerical bound. The analytic scores agree with finite differences. When we reestimate the model with 7 quadrature nodes for distance and 41 for the outside option, instead of 15 and 81, every estimated parameter changes by less than 0.13 standard errors and each female-to-male ratio by at most 0.015. These checks support the numerical accuracy of the estimates.

B Coefficient tables

Tables 9–11 report all distance coefficients, the field adjustments by gender and category, and the implied nesting parameters for the correlated main specification. Standard errors are clustered by home state (Section 5.4) and use the joint covariance. Tests of the main comparisons of distance coefficients across groups use the restricted bootstrap p -values reported in Section 6.1. Program effects are included in every specification, and the intercepts of the outside option in every specification that has one, although Table 3 does not display them. Appendix B.1 reports the intercepts of the outside option in the main specification.

Table 9: All distance coefficients: correlated main specification

Coefficient	Estimate	SE
<i>Mean-coefficient parameterization</i>		
Baseline: general-category men	−0.930	0.132
Female $\times$ distance	−0.547	0.109
Reserved $\times$ distance	−0.196	0.068
Female $\times$ Reserved $\times$ distance	−0.260	0.164
<i>Group means</i>		
General-category men	−0.930	0.132
General-category women	−1.477	0.163
Reserved-category men	−1.126	0.106
Reserved-category women	−1.932	0.182
<i>Group SDs</i>		
General-category men	0.655	0.105
General-category women	0.875	0.131
Reserved-category men	0.964	0.138
Reserved-category women	1.512	0.174

Correlated main model. All coefficients use raw utility units, with distance per 1,000 km. Group means combine the four mean-coefficient parameters and their joint covariance. SDs are the absolute values of the group SD parameters, with delta-method SEs. SEs are clustered by home state.

Table 10: Field adjustments by gender and category

Field	Female	Reserved	Female $\times$ Reserved
Aerospace Engineering	-0.089 (0.308)	0.091 (0.143)	-0.094 (0.630)
Bio Sciences	1.224 (0.515)	-0.861 (0.317)	-1.278 (0.964)
Bioengineering and Biotechnology	0.938 (0.147)	-0.025 (0.272)	0.029 (0.242)
Chemical Engineering	0.250 (0.152)	0.130 (0.099)	-0.056 (0.162)
Chemistry and Chemical Sciences	0.478 (0.376)	-0.372 (0.326)	0.479 (0.464)
Civil Engineering	-0.460 (0.160)	1.068 (0.142)	0.002 (0.147)
Information Technology	-0.176 (0.154)	-0.685 (0.149)	0.456 (0.214)
Mathematics	0.066 (0.327)	-1.706 (0.163)	-0.646 (0.480)
Physics and Engineering Physics	-0.465 (0.414)	-0.973 (0.186)	0.933 (0.395)
Design and Product Design	-0.804 (1.008)	-0.732 (0.248)	1.356 (1.380)
Earth, Mining and Petroleum	-0.456 (0.401)	0.388 (0.155)	-0.627 (0.453)
Economics	0.275 (0.603)	-1.316 (0.183)	0.735 (0.620)
Electrical, Electronics and Instrumentation	0.144 (0.093)	0.122 (0.146)	0.178 (0.090)
Food and Agricultural Engineering	-0.655 (0.816)	0.430 (0.404)	1.002 (1.141)
Materials and Metallurgy	-0.109 (0.171)	-0.181 (0.135)	0.248 (0.205)
Mechanical, Industrial and Manufacturing	-2.038 (0.100)	0.450 (0.117)	0.496 (0.140)
Naval Architecture and Ocean Engineering	-0.692 (0.530)	-0.405 (0.224)	0.183 (0.846)
Pharmacy	0.501 (0.744)	-0.196 (0.440)	$-\infty$

Correlated main model; raw utility shifts relative to Computer Science and Engineering. Parentheses contain SEs clustered by home state. The Pharmacy interaction for reserved-category women has a limiting value of $-\infty$ because that group never chooses Pharmacy.

Table 11: Nesting parameters for all fields, by group

Field	General		Reserved	
	Men	Women	Men	Women
Aerospace Engineering	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Bio Sciences	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Bioengineering and Biotechnology	0.309 (0.038)	0.317 (0.041)	0.411 (0.040)	0.425 (0.051)
Chemical Engineering	0.296 (0.023)	0.304 (0.028)	0.396 (0.034)	0.409 (0.046)
Chemistry and Chemical Sciences	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Civil Engineering	0.270 (0.018)	0.278 (0.027)	0.366 (0.026)	0.379 (0.033)
Computer Science and Engineering	0.429 (0.033)	0.439 (0.035)	0.540 (0.037)	0.554 (0.048)
Information Technology	0.428 (0.028)	0.437 (0.035)	0.538 (0.034)	0.553 (0.042)
Mathematics	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Physics and Engineering Physics	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Design and Product Design	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Earth, Mining and Petroleum	0.437 (0.045)	0.446 (0.055)	0.547 (0.056)	0.561 (0.057)
Economics	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Electrical, Electronics and Instrumentation	0.419 (0.025)	0.429 (0.030)	0.530 (0.029)	0.544 (0.039)
Food and Agricultural Engineering	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Materials and Metallurgy	0.291 (0.026)	0.299 (0.032)	0.390 (0.027)	0.404 (0.043)
Mechanical, Industrial and Manufacturing	0.401 (0.023)	0.410 (0.028)	0.511 (0.028)	0.525 (0.042)
Naval Architecture and Ocean Engineering	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)
Pharmacy	0.515 (0.039)	0.525 (0.048)	0.623 (0.040)	0.637 (0.049)

Implied nesting parameters λ_{gk} in the correlated main model, with delta-method SEs clustered by home state in parentheses. Ten smaller fields share a base parameter, so their nesting parameters are equal within each group. These are the implied parameters rather than the underlying logit coefficients; the outside option is a nest on its own and has no separately estimated parameter.

B.1 Intercepts of the outside option

Table 12 reports the intercepts of the outside option in the main specification (Section 5.2). Except for general-category men, the levels of the group terms κ_g depend on the choice of Computer Science and Engineering as the reference field (Section 5.1); the home-state terms γ_u do not. State codes follow the Class XII records.

Table 12: Intercepts of the outside option

	General		Reserved	
	Men	Women	Men	Women
Group term (κ_g)	−2.365 (0.153)	−2.905 (0.195)	−2.487 (0.179)	−2.453 (0.236)

Home state	γ_u	Home state	γ_u	Home state	γ_u
AP	−0.335 (0.112)	JK	−0.122 (0.447)	PO	0.982 (0.949)
BH	−0.873 (0.149)	JR	−0.717 (0.186)	RJ	Reference
CG	−0.183 (0.293)	KL	1.434 (0.161)	TL	0.507 (0.111)
CH	2.213 (0.272)	KR	2.344 (0.176)	TN	0.617 (0.186)
DL	3.289 (0.159)	MH	1.981 (0.126)	UP	−0.026 (0.111)
GJ	1.904 (0.163)	MP	0.017 (0.156)	UR	−0.116 (0.293)
GO	1.763 (0.695)	Northeast	−0.285 (0.242)	Union territories	−0.794 (0.849)
HP	−1.297 (0.441)	OR	0.119 (0.213)	WB	1.074 (0.165)
HR	0.307 (0.173)	PB	0.657 (0.223)		

For a candidate in group g whose home state belongs to unit u , the intercept of the outside option is $\kappa_g + \gamma_u$, evaluated at a centered Ability of zero. Rajasthan (RJ) is the reference unit, with $\gamma_u = 0$. The pooled northeastern states are Arunachal Pradesh, Assam, Manipur, Meghalaya, Mizoram, Nagaland, Sikkim, and Tripura. The pooled union territories are the Andaman and Nicobar Islands, Dadra and Nagar Haveli, and Daman and Diu; Lakshadweep has no candidates in the sample. Parentheses give model-based SEs from the inverse Hessian. SEs clustered by home state are not informative for these parameters, because each home-state term depends on the candidates of one unit, and the level of the group terms depends on those of Rajasthan. Less familiar state codes are BH Bihar, CG Chhattisgarh, GO Goa, JR Jharkhand, KR Karnataka, OR Odisha, PO Puducherry, TL Telangana, and UR Uttarakhand.

C Constructing candidate location proxies

This appendix documents the location rule in Section 2.5, combining JEE-Main examination cities with candidates' home states.

C.1 City codes and coordinates

Appendix 2 of the JEE-Main 2015 information bulletin lists examination-city codes from 101 through 383 ([Central Board of Secondary Education, 2015](#)). The first three digits of the anonymized eight-digit JoSAA roll number fall in this range after subtracting 500 from prefixes of 500 or more. We use the resulting code to link candidates to the official examination-city list. The mapping relies on the anonymization preserving the city-code component of the roll number.

City, state, and institute coordinates come mainly from OpenStreetMap. All programs at an institute share its destination coordinate. In our sample, every candidate who took the exam in India has matched city coordinates.

C.2 Location rule and limitations

We harmonize state-code aliases and apply the same origin to every program in a candidate's feasible set. Of 29,024 candidates, 24,902 retain in-state city coordinates and 4,122 use home-state centroids: 4,072 Indian out-of-state examinees and 50 overseas examinees. Compared with the robustness check below, only the 4,072 out-of-state candidates' distances change; the sample, feasible sets, and choices are unchanged.

Of the 1,070 Rajasthan examinees with a different home state, 313 have Bihar as their home state and 276 Uttar Pradesh: together, 55.0 percent. The next largest origins are Jharkhand (110), Madhya Pradesh (105), and Haryana (43). This pattern is consistent with Kota's coaching centers serving students from other states ([Government of Rajasthan, 2015](#)). The main rule uses these candidates' home state as the broader origin proxy.

For 86 percent of candidates, the decoded examination city lies in their home state, which is consistent with both the decoding and the use of the home state as a proxy for state of residence.

The robustness check in Section 7.3 uses examination-city coordinates for all 28,974 candidates who took the exam in India and home-state centroids for the 50 overseas examinees. A second check there uses home-state centroids for all candidates. None of these specifications identifies permanent home addresses; all measure distance from a pre-college location proxy.

Confidentiality. Decoded cities remain internal. Public replication files contain only aggregate cells or derived distances; they exclude individual candidate locations, direct identifiers, and unrestricted text fields.

D Academic preparation and the admissions funnel

Table 13 places the conditional preference estimates in the wider admissions population. Among JEE-Main rank holders with valid reported marks, women average higher Class XII marks; their share is lower among JEE-Advanced rank holders and candidates assigned an IIT seat. These cross-sectional shares do not isolate the roles of examination performance, participation, or program choice.

Table 13: Academic preparation and female representation

<i>A. Female representation across admissions populations</i>		
Population	Candidates	Women (%)
Administrative candidate roster	1,304,217	26.5
JEE-Main engineering Common Rank List	1,212,950	25.8
Ordinary JEE-Advanced rank holders	25,189	11.3
Ordinary first-round IIT assignees	9,820	8.6
<i>B. Board marks among JEE-Main engineering rank holders</i>		
Group	Valid marks	Mean mark
Men	899,598	71.2
Women	312,743	74.9

Notes: First-round administrative records. Female shares use men and women as the denominator; nine other-coded records enter the first two candidate counts only. Ordinary JEE-Advanced rank holders have any positive ordinary overall, category or PwD rank, excluding preparatory-only ranks. Ordinary assignments exclude preparatory seats, defence-supernumerary seats, and seats assigned to foreign nationals, and they include PwD candidates. Including these seats gives 10,033 assignees and 8.7% women. Board marks are the reported engineering-admissions mark on a 100-point scale; 600 zero entries are excluded. These populations are distinct from the estimation sample of 29,024 candidates.

In our estimation sample, excluding the nine candidates without a positive recorded mark, women’s mean board mark is 93.8 versus 90.9 for men. The 2015 IIT admission criterion allowed either the board- and category-specific top 20 percent or an aggregate of at least 75 percent for general-category and OBC-NCL candidates and 70 percent for SC, ST, and PwD candidates ([Joint Seat Allocation Authority, 2015](#), Annexure 3).

E Descriptive assignment regressions

Table 14 reports OLS regressions of first-round assigned distance on gender and progressively richer controls. All four columns use the same 25,340 candidates with a first-round seat. In columns (2)–(4), the Female coefficient is the gender gap among general-category candidates; adding the Female $\times$ Reserved coefficient gives the gender gap among reserved-category candidates.

Table 14: Gender and first-round assigned distance: regression estimates

Variable	Dependent variable: Assigned distance (km)			
	(1)	(2)	(3)	(4)
Female	-120.3 (17.5)	-128.0 (21.4)	-124.0 (22.3)	-152.7 (24.4)
Reserved	—	3.8 (16.9)	28.0 (30.5)	2.5 (28.3)
Female $\times$ Reserved	—	16.0 (17.7)	14.3 (16.8)	14.8 (16.3)
Ability	—	—	13.1 (14.9)	5.3 (14.7)
Constant	656.0 (38.0)	653.9 (38.1)	639.1 (48.7)	
Home-state FE	No	No	No	Yes
<i>N</i>	25,340	25,340	25,340	25,340

Notes: Home-state-clustered standard errors (35 clusters) are in parentheses. Reserved pools OBC-NCL, SC, and ST candidates. Ability is defined in Section 5.2: minus the log of the smaller of the candidate's valid JEE-Main engineering and JEE-Advanced overall ranks, centered at its sample mean and not standardized. Higher values indicate better rank. The two candidates without an overall rank have Ability set to zero; columns (3) and (4) include an indicator for them, whose coefficient is omitted. Column (4) includes home-state fixed effects; their coefficients and the constant are omitted. Distance uses the origin defined in Section 2.5.

F Repeat attempts and the scale of program values

This appendix uses changes in admission ranks between consecutive years to interpret the program-value standard deviation (the unit in which we express candidates’ distance sensitivity) in Section 5.4. A candidate who spends another year preparing and accepts a newly available undergraduate seat starts the first year of the new program one year later. Repeat attempts are common: 3,393 of the 9,820 candidates offered ordinary IIT seats in the first round of 2015 (more than a third) link to the 2014 Advanced qualifying list or to a recorded 2014 Main rank, using the unique name–date-of-birth links described below. This count includes all categories and both sexes, counting each candidate once. We do not observe preparation effort or prior college enrollment in the matched records.

F.1 Matched candidates and rank changes

We link the 2014 JEE-Advanced qualifying rank list and the 2014 JEE-Main candidate database separately to the 2015 JoSAA first-round candidate roster. Links require an exact match on normalized name and date of birth that is unique in both source files. We retain general-category candidates in both years, include men and women, and require a recorded overall rank in both years. These are supplementary samples, distinct from the preference-estimation sample.

The JEE-Advanced sample contains 1,337 candidates: 1,106 improve their rank, 230 worsen, and one is unchanged. A further 933 linked general-category candidates with a recorded 2014 Advanced rank have no recorded 2015 Advanced rank and cannot enter the log-rank calculation. We do not classify these missing outcomes as unchanged or as improvements. For JEE-Main, we restrict the 2014 rank to at most 30,000 and exclude candidates with a recorded 2014 Advanced rank better than 10,000, as motivated in Appendix F.3. We retain all recorded 2015 Main ranks, including those above 30,000. This gives 1,889 pairs, of which 1,089 improve and 800 worsen; another 245 links meeting the 2014 restrictions lack a recorded 2015 Main rank. The reported gains are therefore conditional on linkage and a recorded rank in both years.

Figure 4 shows all recorded rank pairs for both exams. For each exam and sample, we fit by unweighted least squares

$$\ln r_{i,2015} = a + b \ln r_{i,2014} + \varepsilon_i, \quad (11)$$

both with a free slope and with $b = 1$. In the latter case,

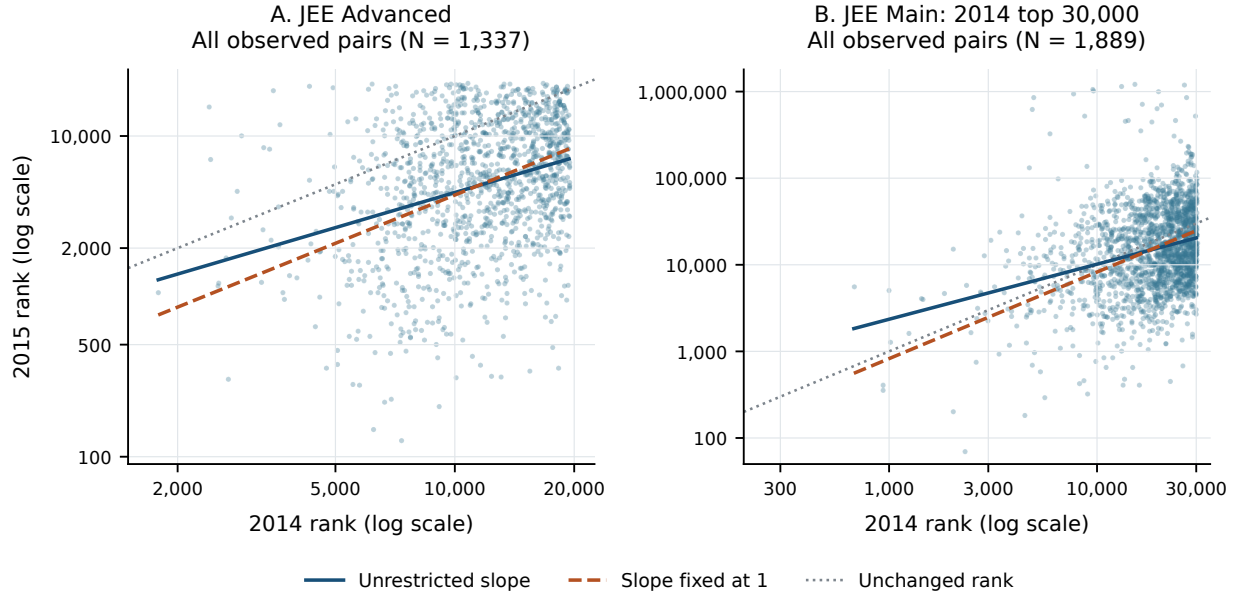
$$-\hat{a} = \frac{1}{N} \sum_i (\ln r_{i,2014} - \ln r_{i,2015}) \equiv \overline{G}. \quad (12)$$

Thus, the slope-one intercept summarizes the average log-rank gain without requiring a tight linear relationship between ranks across attempts. Its associated geometric improvement factor is $\exp(\overline{G})$. The improvers sample additionally requires $r_{i,2015} < r_{i,2014}$.

Table 15: Rank changes between 2014 and 2015

Exam	Sample	N	Free a	Free b	a ($b = 1$)	Gain factor
Advanced	All pairs	1,337	1.699	0.727	-0.848	2.33
Advanced	Improvers	1,106	-2.263	1.124	-1.103	3.01
Main ($\leq 30,000$)	All pairs	1,889	3.371	0.636	-0.193	1.21
Main ($\leq 30,000$)	Improvers	1,089	-1.065	1.013	-0.938	2.56

Notes: General-category candidates in both years, including both sexes, with unique name–date-of-birth links and a recorded overall rank in both years. “All pairs” includes observed deteriorations and unchanged ranks. Both Main samples require a 2014 Main rank at most 30,000 and exclude candidates with a recorded 2014 Advanced rank better than 10,000. Coefficients refer to natural logarithms in equation (11); the last column is $\exp(-\hat{a})$ from the slope-one fit.

**Figure 4:** Entrance-exam ranks across consecutive attempts

Notes: Matched general-category candidates of both sexes with a recorded rank in both years, including candidates whose ranks worsen. Panel A shows JEE-Advanced. Panel B shows JEE-Main, restricting the 2014 Main rank to at most 30,000 and excluding candidates with a recorded 2014 Advanced rank better than 10,000; all recorded 2015 Main ranks are retained. Both axes show raw ranks on logarithmic scales; their ranges differ across panels. The solid and dashed lines are the unrestricted and slope-one fits from Table 15; the dotted line denotes an unchanged rank. The unrestricted fits have $R^2 = 0.095$ for Advanced and 0.082 for Main.

F.2 Conversion to program-value units

Figure 3 relates the common program effects from our preferred specification to the logarithm of the 2015 first-round ordinary open closing rank, c_j . Its 493 programs have positive first-round acceptances by general-category men: 213 IIT and 280 non-IIT programs. We use these acceptances, n_{gj} , both as regression weights and to size the bubbles. We fit IIT and non-IIT programs separately because their cutoffs come from different exams. With program effects divided by $s_{\text{prog}} = 3.42642$, the scale used throughout the paper, the acceptance-weighted fits are

$$\frac{\delta_j}{s_{\text{prog}}} = \begin{cases} 7.4623 - 0.8601 \ln c_j + u_j, & \text{IIT,} \\ 5.9412 - 0.7185 \ln c_j + u_j, & \text{non-IIT.} \end{cases} \quad (13)$$

The weighted R^2 values are 0.924 and 0.888, respectively. Because general-category men are the reference group, δ_j is their value of program j . Mapping the mean Advanced log-rank gain $\bar{G}$ through the IIT relationship gives $0.8601 \bar{G}$ SD of program value at the admission cutoff:

$$\begin{aligned} \text{All recorded rank pairs: } & 0.8601 \times 0.8476 \simeq 0.729 \text{ SD,} \\ \text{Improvers only: } & 0.8601 \times 1.1027 \simeq 0.948 \text{ SD.} \end{aligned}$$

The corresponding geometric rank-improvement factors are 2.33 and 3.01. Even among improvers, the mean gain is therefore about one SD of program value.

These descriptive translations do not measure candidate-specific changes in the best feasible program or realized assignment. Selecting improvers gives an optimistic benchmark for observed gains. The exercise identifies neither a causal upper bound on the effect of preparation nor willingness to pay for a fixed one-SD gain.

F.3 JEE-Main comparison

The JEE-Main comparison checks the Advanced results using the other exam. Candidates aiming for highly preferred IIT seats may focus their preparation on JEE-Advanced at the expense of JEE-Main: those with a 2014 Advanced rank better than 10,000 may aim for a rank better than 5,000 to reach more attractive IIT programs. We therefore exclude them from both Main samples, so that Main rank changes are less likely to understate gains from preparation directed toward Advanced. The exclusion uses the linked 2014 Advanced qualifying list; candidates without a recorded 2014 Advanced rank are retained. We do not observe preparation focus directly.

Panel B of Figure 4 shows the resulting Main sample, retaining all recorded 2015 ranks. Of these candidates, 57.6 percent improve. Applying the separate non-IIT calibration to the slope-one fits in Table 15 gives 0.138 SD overall and 0.674 SD among improvers.