September 14, 2026

Workshop Organizers
Workshop on Development Economics and Networks for Early-Career Researchers
National Bureau of Economic Research
1050 Massachusetts Avenue
Cambridge, MA 02138

Dear Professors Breza, Chandrasekhar, and Viviano,

I am writing to apply for the NBER Workshop on Development Economics and Networks for Early-Career Researchers. I am a PhD student in Public Policy at Northeastern University, with a concentration in Economics. My research is in development and applied microeconomics, with a particular interest in how institutions and social and economic environments shape access to educational and labor-market opportunities.

A common question across my current research is whether formal eligibility and institutional access translate into meaningful economic opportunity. In one project, I study India's centralized engineering admissions system and reconstruct the set of college programs available to each applicant to examine how priority rules shape educational opportunities and subsequent allocations. In a second project, I exploit a sharp grace-mark threshold in secondary-school examinations to study how marginal failure affects progression to senior secondary school and students' subsequent fields of study. Both projects have led me to think more broadly about the environments within which individuals respond to institutional opportunities.

This question also motivates my work on local labor markets and networks. In ongoing research on Opportunity Zones, I use commuting flows to study whether place-based investment connects residents of treated neighborhoods to employment opportunities elsewhere in the local economy. The project focuses on the structure of these labor-market connections, including segregation, homophily, and the position of disadvantaged neighborhoods within commuting networks. More broadly, I am interested in how networks mediate the relationship between policies that expand formal access and the economic opportunities individuals can actually reach.

The workshop is therefore particularly well matched to the direction of my research. I hope to deepen my understanding of empirical network methods, particularly the measurement of economically meaningful networks, the identification of network effects, and the distinction between network mechanisms and correlated environments. I would also value the opportunity to receive feedback from researchers working at the intersection of development economics and networks as I develop these questions into a larger part of my dissertation research.

My application includes my curriculum vitae, research statement, and working paper. If I am shortlisted, Professor Nishith Prakash, Professor of Economics and Public Policy at Northeastern University, could provide a recommendation. He can be reached at n.prakash@northeastern.edu.

Thank you for considering my application. I would be grateful for the opportunity to participate in the workshop.

Sincerely,

Kishan Narayan
PhD Student in Public Policy, Economics Concentration
Northeastern University
Boston, MA 02115
Email: narayan.ki@northeastern.edu
Phone: 857-423-5047

# Kishan Narayan

narayan.ki@northeastern.edu | Website | LinkedIn | 857-423-5047 | Boston, MA

## Education

**Northeastern University, Boston, MA**, Ph.D. Student in Public Policy (Economics Concentration) 2023 –
PhD Student Visitor, Stone Inequality Centre (HKS), Harvard University *Fall 2025*
MS Computational Social Science, Northeastern University *2025*

**Jawaharlal Nehru University, India**, M.A. in Applied Economics 2021

**Banaras Hindu University, India**, B.A. (Honors) in Economics 2019

## PhD Coursework Completed

**At Northeastern University:** PhD Microeconomics I–II; PhD Econometrics I–II; Labor Economics I–II; Computational Statistics & Modeling and Analyzing Social Networks
**At MIT:** Behavioral Economics
**At Harvard:** Design and Analysis of Field Experiments in Education

## Work in Progress

- **Priority, Opportunity, and Allocation in Centralized Admissions** (Draft Available on request)

  Reconstructs candidate-specific feasible college-program sets in India's centralized engineering admissions using published closing ranks. Estimates revealed preferences to separate differences in available opportunities from differences in choices, and quantifies how these margins contribute to gaps in program quality and earnings.

- **Second Chances, Progression, and Streams Sorting: Evidence from High-Stakes Board Exams in India**

  Exploits a sharp grace-mark threshold in Rajasthan's Grade 10 board exams to study the consequences of marginal failure. Immediate certification substantially increases Grade 12 progression and eventual completion, while effects on STEM specialization are considerably smaller.

- **Where Do Opportunity Zone Jobs Go? A Social Network Analysis of Employment Flows in Place-Based Tax Policy** with Hitanshu Pandit (Under Revision)

  This paper explains why the Opportunity Zone (OZ) program produced no measurable gains for residents despite large capital inflows. Using ACS and LODES data through 2022, I replicate and extend prior work and find precise null effects on resident employment, poverty, income, and—newly—on workplace employment in OZ tracts. I then show that these null effects arise from the labor-market network structure. Commuting-flow networks reveal that OZ tracts occupy peripheral, highly segregated positions with low in-degree centrality, high structural constraint, and limited connectivity to high-income labor markets. Although OZ tracts are geographically close in network distance to opportunity-rich areas, strong income-based homophily prevents cross-income ties from forming, and policy designation does not alter these patterns.

## Publications

### Peer-Reviewed Publication

"The Frozen Constituency: Delayed Delimitation and Denied Representation of SCs and STs," with Anagha Pradeep. *Economic and Political Weekly*, Vol. 61, Issue No. 23, 06 Jun, 2026. Paper

### Policy Reports and Other Publications

*Rationalisation of Explicit Subsidies at State Level*, with H. K. Amar Nath, S. H. Nayudu, M. Gurdatta, and A. Raj. National Institute of Public Finance and Policy. Study prepared for NITI Aayog, Government of India, 2024. Report

## Research Experience

**Research Assistant** **Northeastern University**
*PIs: Silvia Prina, Nishith Prakash, Daniel Aldrich, Shantanu Khanna* *Sep 2023 – Present*

- Designed and implemented a large-scale field experiment with Global Action for Policy; co-organized NEUDC 2024.
- Estimate causal impacts of India's public housing program on labor-market outcomes using quasi-experimental methods with Shantanu Khanna.
- Model behavioral decision-making under extreme weather uncertainty at the Aldrich Resilience Lab.

**Research Analyst** — **Development Data Lab**
*PIs: Sam Asher, Paul Novosad* — *Feb 2023 – Aug 2023*

- Contributed to SHRUG (Socioeconomic High-resolution Rural-Urban Geographic Platform for India).
- Managed data integration for Indian Parliament datasets on representation, delimitation, and development indicators.

**Research Fellow** — **NIPFP, Ministry of Finance**
*PI: Amarnath H.K.* — *May 2022 – Feb 2023*

- Led ML-based digitization & analysis of historical budget documents to support a subsidy rationalization framework.
- Work presented in Parliament's Budget Session 2023 and adopted by NITI Aayog and the Finance Commission.

## Selected Presentations and Conferences

Southern Economic Association Annual Meeting (2026), APPAM (2026), NetSci (2026) Conference on Network Science; Septses Advance Econometrics Method (2025), NEUDC 2024 (Northeastern), 8th NITI Aayog Governing Council Meeting (2022), Indian Society of Labour Economics (ISLE) (2022)

## Teaching Experience

| | |
|---|---|
| Teaching Assistant – ArcGIS & QGIS for Urban & Regional Policy | Northeastern University, Spring 2025 |
| Teaching Assistant – Economic Analysis for Public Policy | Northeastern University, Fall 2024 |
| Instructor – Introduction to Microeconomics & Statistics | Govt. Sr. Secondary School, Delhi, 2021-2023 |

## Additional Training

| | |
|---|---|
| BREAD–IGC Virtual PhD Course (Social Protection) | Fall 2025 |
| Summer School in Modern Methods in Theoretical and Empirical Economics | Summer 2025 |
| BREAD–IGC Virtual PhD Course (Urban Economics) | Fall 2024 |

## Selected Honors and Awards

| | |
|---|---|
| PhD Network Grant, Northeastern University | 2025 |
| National Fellowship, Ministry of Education (India) | 2022 |
| B.R. Ambedkar Scholarship (Gold Medalist), BHU | 2019 |

## Professional and Public Service

| | |
|---|---|
| **Journal Referee** | Journal of Population Economics, Journal of Economic Behavior & Organization |
| **External-Volunteering** | Mentor & Organizer – Application Abroad Mentorship Program at Bahujan Economists, Delhi government middle school weekly math tutor ('21-'22), JNU prospective student tutor ('19-'20), Ministry of Tourism Volunteer – Tax return Assistance ('17) |

## Skills

**Programming:** Stata, R, Python (pandas, NumPy, statsmodels, scikit-learn, web scraping, statnet)

**Languages:** English (Professional fluency) — Hindi (Native) — Russian (Intermediate)

**Tools:** Qualtrics — SurveyCTO — Gephi – Git & GitHub – ArcGIS – QGIS – oTree

## References

**Professor Nishith Prakash**
Professor of Economics and Public Policy
Northeastern University
Email: n.prakash@northeastern.edu

**Professor Silvia Prina**
Professor of Economics
Northeastern University
Email: s.prina@northeastern.edu

**Research Interests and Motivation for Attending**

My research is in development and applied microeconomics, with a particular interest in how institutions shape access to educational and economic opportunity. Across my current projects, I study settings in which individuals face sharply different opportunity sets because of institutional rules, educational thresholds, or the structure of local labor markets. A broader question increasingly connects this work: when policies or institutions expand formal access, what determines whether individuals can actually take advantage of the opportunities made available to them?

My work on education approaches this question from two directions. In one project, I study India's centralized engineering admissions system and reconstruct the set of college programs available to each applicant. This allows me to separate differences in what students could choose from differences in how they choose among available options, and to study how priority rules translate into differences in educational and labor-market opportunities. In separate work on high-stakes secondary-school examinations in India, I exploit a sharp grace-mark threshold to study what happens to students at the margin of failure. Immediate certification substantially increases progression to senior secondary school and eventual completion, while changes in academic specialization are more limited. Together, these projects have made me interested not only in whether institutions make an opportunity formally available, but also in the environments through which individuals learn about, evaluate, and respond to those opportunities.

This interest has led me increasingly toward network-based questions. Educational and labor-market decisions are embedded in relationships among peers, families, schools, neighborhoods, and workplaces. Information about colleges, occupations, returns to education, and employment opportunities often travels through these connections, while social and spatial segregation can limit which opportunities become visible or reachable. In ongoing work on Opportunity Zones, I use commuting flows to characterize the position of treated neighborhoods within local labor markets. The project asks whether place-based investment connects residents to employment opportunities elsewhere in the metropolitan economy and examines how segregation, homophily, and network position shape those connections. Working on this project has made me especially interested in the distinction between geographic proximity and economic connectedness, and more broadly in how network structure can mediate the effects of policies intended to expand opportunity.

At this stage of my PhD, I want to develop networks from a method I use in one project into a more systematic part of my research agenda. I am particularly interested in how researchers define economically meaningful networks, measure exposure and connectedness, distinguish network mechanisms from correlated environments, and address the endogeneity of network formation. These questions are directly relevant to where I hope to take my work on education and development. For example, I would like to understand when differences in educational choices reflect formal constraints, differences in information transmitted through social networks, or interactions between the two. The NBER Workshop on Development Economics and Networks would give me an opportunity to learn how researchers at the frontier of empirical network research approach these identification and measurement problems, while receiving feedback on research questions I am currently developing. I would especially value the opportunity to engage with other early-career researchers working on related problems and to refine a research agenda centered on how institutions, information, and networks jointly shape access to economic opportunity.

# Priority, Opportunity, and Allocation in Centralized Admissions

Kishan Narayan*

Northeastern University

September 9, 2026

*Preliminary draft. Please do not circulate or cite.*

**Abstract**

We examine how centralized admissions turn differences in priority into differences in earnings. We leverage institutional features of India's Joint Seat Allocation Authority, which assigns seats in the Indian Institutes of Technology by deferred acceptance and rations them through candidate-specific eligibility routes. The allocation publishes closing ranks for every route and program. Using these cutoffs to reconstruct the menu of programs each of 119,000 candidates could have entered, we estimate a random-utility model of program choice and then hold its preference parameters fixed while varying only the menus. Differences in what candidates can attain account for about a quarter of an INR 17.5 lakh top-to-bottom gap in median program compensation; common preferences applied to their own menus reproduce most of the rest. A 2018 reform that added female-only capacity widened access frontiers in the direction the reconstruction implies. Our results suggest that the gap is generated less by how good candidates' options are than by what those options force them to give up.

**Keywords:** Centralized admissions; matching markets; opportunity sets; college choice; priority; higher education; deferred acceptance

**JEL Classification:** C78, D47, I23, I24, J24

*Email: narayan.ki@northeastern.edu. We thank participants in the PhD seminar at Northeastern University and in the Stone Program in Wealth Distribution, Inequality, and Social Policy at the Harvard Kennedy School for helpful comments and suggestions. We used OpenAI and Anthropic tools for assistance with LaTeX coding, grammar, and data management. All substantive decisions, analyses, interpretations, and errors are our own.

# 1 Introduction

Observed assignments in centralized markets reflect both what participants prefer and what the mechanism allows them to obtain. A student assigned to a particular college program may prefer it to every alternative she could realistically reach, yet another student with identical tastes may receive a different assignment because priority, eligibility, or capacity gives her a different set of attainable options. Final allocations therefore combine *choice* with an institutionally generated *opportunity set*.

This distinction is fundamental to matching and market design. Deferred acceptance and related mechanisms map submitted preferences, priorities, and capacities into assignments (Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003). A substantial empirical literature has developed methods for recovering preferences in centralized assignment environments under different reporting and observability conditions. Stability and ex ante priorities can reveal preference information among alternatives for which a participant qualifies ex post (Fack et al., 2019; Arslan, 2021), while observed matches can identify preferences through stability restrictions (Agarwal, 2015). Other approaches explicitly allow submitted rankings to respond strategically to the mechanism (Agarwal and Somaini, 2018), and the structure of coordinated assignment can be used for welfare and mechanism counterfactuals (Abdulkadiroğlu et al., 2017). In Fack et al. (2019) and Arslan (2021) in particular, candidate-specific qualifying or accessible alternatives are important inputs into preference recovery.

We instead make those candidate-specific feasible sets themselves an object of analysis. The distinction between preferences and opportunities is not itself new: observed choices can differ because decision makers value alternatives differently or because they face different opportunities (Zeng and Xie, 2008; Bonhomme et al., 2016). Our contribution is to make the mechanism-generated distribution of individual opportunity sets central to the empirical analysis and to distinguish two economically different dimensions of the constraint created by a centralized priority-rationed market: the *level* of opportunity and the *structure* of opportunity.

Opportunity level asks whether one candidate can attain alternatives with better outcomes on average. Opportunity structure asks which combinations of attributes are jointly available within the feasible menu. Two candidates can face menus with the same average program quality but very different trade-offs. For one candidate, higher-earning programs may require substantially greater geographic distance; for another, similarly high-earning programs may be available nearby. More generally, priority can change the joint configuration of earnings, location, field of study, placement outcomes, and institutional quality over which the candidate chooses. A feasible set is therefore not a lottery over equally weighted alternatives. Participants choose among multidimensional bundles, so the same behavioral rule can generate systematically different assignments when applied to differently structured menus.

We study this problem in centralized undergraduate admissions to the Indian Institutes of Technology (IITs), a highly selective higher-education market in which institution and field are jointly allocated. Allocation among selective institutions and fields is economically consequential:

field and institution choices can have large labor-market consequences, elite-college attendance can affect access to top occupations and incomes, and access to selective colleges is sharply stratified (Kirkeboen et al., 2016; Zimmerman, 2019; Chetty et al., 2020, 2026). These studies motivate why understanding the distribution of access among differentiated and economically consequential educational options matters. IIT-specific evidence also documents that, among top JEE scorers, attendance at the original top five IITs is associated with greater subsequent migration for graduate study, underscoring that allocation across IIT institutions can shape later trajectories (Choudhury et al., 2023). Our estimand is the distribution and use of opportunity within the IIT market rather than the causal return to attending an IIT program.

The IIT setting is particularly useful because the Joint Seat Allocation Authority (JoSAA) generates a rich and observable mapping from institutional priority to individual opportunity. Candidates rank academic programs, while JEE (Advanced) merit lists, reservation categories, gender-specific pools, program capacities, and other eligibility rules determine the routes through which those programs can be obtained. JoSAA processes these objects through a candidate-proposing Multi-Run Multi-Round Deferred Acceptance mechanism. The design and implementation of this centralized system, including its multiple merit lists, virtual program-sections, de-reservation rules, and multiround operation, are documented by Baswana et al. (2019a). We use that institutional architecture for a different empirical purpose: to recover the opportunity set generated for each candidate and study how variation in its level and structure maps into final allocations.

We combine administrative records from nine IIT admission cycles, 2015–2023, with JoSAA program-by-category admission frontiers, academic-program labor-market outcomes, institutional-quality measures, and candidate geography. Using every observed priority route for which a candidate is eligible, we reconstruct the set of IIT academic programs that the candidate could attain. The final analysis contains more than 119,000 candidate choice sets and approximately 18.6 million candidate–program alternatives. The reconstruction is tightly disciplined by realized administrative allocations: for 99.4 percent of candidates, the program actually allocated by JoSAA is contained directly in the opportunity set implied by the published route-specific admission frontiers. For the remaining 0.6 percent, we append only the administratively observed allocated program as revealed evidence that it was attainable under the historical mechanism. We otherwise do not use the candidate's realized choice to expand the feasible menu.

The resulting object is an *institutional opportunity set*, not a behavioral consideration set. We do not require every attainable program to have been actively considered by the candidate, nor do we require candidates to have known the exact closing ranks that would emerge later in the allocation cycle. These distinctions matter because applicants can hold heterogeneous beliefs about admission probabilities (Kapor et al., 2020), can discover preferences during the admissions process (Grenet et al., 2022), and can make dominated or otherwise imperfect choices even under centralized assignment mechanisms (Shorrer and Sóvágó, 2024). More precise information about admission chances can itself alter college choices (Ye, 2024). We therefore distinguish realized institutional attainability from subjective beliefs or consideration and test the principal results using exact

prior-year admission frontiers and lagged program information.

Our empirical strategy is a measurement and decomposition framework. We first reconstruct candidate-specific institutional opportunity sets, characterize how their breadth and average quality vary across predetermined priority, and then estimate revealed sorting among program attributes within those menus. The conditional-choice model supplies a behavioral input to the decomposition; it is not itself the paper's principal empirical contribution.

Our empirical analysis first estimates revealed sorting among program attributes conditional on these candidate-specific feasible sets. Candidates are less likely to select geographically distant IITs and sort systematically toward programs with higher annual median compensation, stronger placement outcomes, and higher institutional-quality scores. In the preferred specification, the distance-equivalent revealed-choice trade-off, which we refer to for convenience as willingness to travel (WTT), is approximately 98 kilometers per additional INR 1 lakh of annual median CTC. Geographic frictions are well established in college choice (Long, 2004; Gibbons and Vignoles, 2012; Suhonen, 2014), and earnings and other pecuniary and nonpecuniary attributes are known to shape field and college choices (Wiswall and Zafar, 2015; Arcidiacono et al., 2012). We therefore use these estimates as behavioral inputs to the opportunity analysis, rather than treating them as a separate methodological contribution or as causal valuations of exogenous changes in program outcomes.

The central analysis instead uses the reconstructed menus themselves to study inequality across the priority distribution. We organize candidates using a predetermined priority measure constructed from merit-list position rather than the category of the seat ultimately allocated. Higher-priority candidates face dramatically broader opportunity sets: the average candidate in the highest priority decile can reach roughly 250 IIT academic programs, compared with fewer than 50 for the lowest-priority decile. Their attainable programs also have substantially higher labor-market outcomes on average.

Yet differences in average opportunity quality explain only a minority of the difference in final assignments. For annual median compensation, the observed difference between the programs chosen by candidates in the highest and lowest predetermined-priority deciles is INR 17.50 lakh per annum. The corresponding difference in the mean CTC of their feasible menus is only INR 4.27 lakh. Differences in average opportunity level therefore account for 24.4 percent of the observed top–bottom allocation gap, while the remaining 75.6 percent arises through sorting within feasible opportunity sets.

This first split is an accounting decomposition conditional on the reconstructed menus. It shows that comparing the average quality of accessible programs captures only a limited part of allocation inequality within the observed IIT market. We then ask how much of the remaining within-menu gap can be reproduced by the *structure* of the menus themselves.

To do so, we estimate a common behavioral rule within each admission cycle and apply it to every candidate's own feasible set. Candidates retain their actual combinations of program compensation, geography, placement outcomes, institutional quality, and other included characteristics, but high- and low-priority groups are evaluated using the same estimated choice parameters. Under this

common-choice benchmark, differently structured menus reproduce an additional INR 9.07 lakh of the top–bottom CTC gap. The INR 17.50 lakh observed gap therefore decomposes into INR 4.27 lakh associated with average opportunity level, INR 9.07 lakh attributed to menu structure under the common behavioral rule, and INR 4.17 lakh of residual rank-associated choice. These components represent 24.4, 51.8, and 23.8 percent of the observed gap, respectively. We do not interpret the residual component as pure preference heterogeneity.

Under the preferred common-choice benchmark, the menu-structure component accounts for 68.5 percent of the observed within-menu sorting gap. Propagating uncertainty in the estimated common-choice coefficients yields a 95-percent parametric interval of 62.7–74.5 percent for this share. The result is therefore not that "preferences explain the rest" once differences in average access are removed. Much of the remaining separation can be reproduced by applying the same behavioral rule to opportunity sets containing different configurations of program attributes.

The two stages of the decomposition have different interpretations. The split between average opportunity level and within-menu sorting is an accounting statement conditional on the reconstructed opportunity sets. The subsequent division of within-menu sorting between menu structure and residual rank-associated choice depends on the common-choice operator. Changing the behavioral reference rule can therefore change that second attribution, but it cannot change the observed allocation gap, the opportunity-level component, or the within-menu sorting gap. Our central conclusion is correspondingly robust to this distinction: differences in average opportunity quality explain only a minority of the CTC allocation gap, while differently structured opportunity sets reproduce a large share of what remains under a common behavioral rule.

The distinction is not specific to compensation. Repeating the framework for placement outcomes and NIRF scores yields different combinations of opportunity level, menu structure, and residual rank-associated choice. Priority therefore does not generate a single scalar inequality in "college quality." It changes a multidimensional opportunity environment, and the importance of average access, the configuration of feasible trade-offs, and remaining observed choice differences varies across outcomes.

Three sets of exercises address the main threats to this interpretation. First, because contemporaneous closing ranks are themselves realized market outcomes, we reconstruct opportunity sets using exact prior-year program cutoffs and replace contemporaneous program attributes with lagged information. This addresses the distinction between realized institutional attainability and the information candidates could plausibly possess when submitting preferences, which is central in work on heterogeneous admission beliefs and preference discovery (Kapor et al., 2020; Grenet et al., 2022). The decomposition remains close to the preferred result. Second, we test sensitivity to the composition of the market and the scalar priority measure using alternative sample periods, leave-one-cycle-out exercises, and alternative predetermined priority definitions. Third, we vary the common-choice operator itself. The accounting split between opportunity level and within-menu sorting is unchanged by these behavioral reference rules, while the subsequent attribution between menu structure and residual choice moves as expected.

We also find suggestive evidence that revealed choice behavior varies across the priority distribution: higher-priority candidates exhibit a larger CTC–distance trade-off than lower-priority candidates in the preferred model. The difference is less precisely estimated and more sensitive to information, support, and substitution assumptions than the decomposition results. Because merit-list position is correlated with preparation, background, information, geography, and other potential determinants of program choice, we describe this pattern as *priority-associated choice heterogeneity*, not as a causal effect of priority on preferences.

The 2018 introduction of female supernumerary seats provides a complementary policy application in which the institution deliberately changes opportunity architecture. A large mechanism-design literature shows that affirmative-action outcomes depend not only on the amount of protected capacity but also on how reserves, priorities, and distributional constraints are implemented (Abdulkadiroğlu, 2005; Hafalir et al., 2013; Dur et al., 2018; Kamada and Kojima, 2015; Sönmez and Yenmez, 2022). Empirical work on affirmative action in Indian engineering admissions studies effects on access, college quality, and academic outcomes (Bertrand et al., 2010; Bagde et al., 2016), while Aygün and Turhan (2017) study the IIT reservation architecture from a matching-design perspective and Mello (2022) separates mechanical and behavioral channels under centralized affirmative action in Brazil. Recent work directly studies the IIT female-supernumerary reform: Sengupta et al. (2025) examine its implementation, effects on female representation, and heterogeneity across IITs and academic programs, while Gupta (2026) examines whether exposure to the reform affects girls' earlier educational choices and participation in the STEM pipeline. Our analysis addresses a distinct institutional margin: how genuinely additional program-level capacity changes the female-specific admission frontier and how that expanded frontier is used in realized allocations.

The 2018 JoSAA rules created Female-Only and Gender-Neutral program sections and linked female-specific capacity to prior program-level female representation (Baswana et al., 2019b). We reconstruct the genuinely additional program-level capacity generated by this seat-construction rule and trace that capacity into the Female-Only/Gender-Neutral admission frontier and, ultimately, into realized Female-Only allocations beyond the corresponding Gender-Neutral frontier.

To estimate the effect of incremental capacity on these outcomes, we exploit the discontinuous relationship between predetermined 2017 female representation and the probability of receiving positive additional female-supernumerary capacity. In the preferred symmetric ±4-percentage-point window, being below the 14-percent threshold raises the probability of receiving positive additional capacity by 0.482. The corresponding reduced-form increase in the predetermined-capacity-weighted share of regular reservation-category routes for which the Female-Only frontier extends beyond the Gender-Neutral frontier is 10.8 percentage points. The fuzzy-IV estimate is 22.5 percentage points, and weak-instrument-robust Anderson–Rubin inference rejects a zero effect for this extensive-margin outcome ($p = 0.017$). Estimates of the magnitude of the frontier extension and of realized beyond-GN recipient counts are considerably less precise. We therefore interpret the reform as local policy-induced evidence that additional capacity can move the opportunity frontier, rather than as a complete estimate of the reform's overall causal effect.

The paper makes three contributions. First, we build directly on empirical work recovering preferences in centralized assignment markets. Fack, Grenet, and He (2019) and Arslan (2021) are particularly close because candidate-specific attainable or accessible alternatives help identify preference information. Agarwal and Somaini (2018) shows how strategic reporting can be incorporated into demand estimation, while Abdulkadiroğlu et al. (2017) uses the structure of coordinated assignment for preference and welfare analysis. We take a different next step: the attainable set is not only a conditioning object required to estimate choice. We study its distribution as an economic outcome and ask how much allocation inequality is already embedded in the opportunities generated by the mechanism.

Second, we connect this literature to work separating preferences from heterogeneous opportunities (Zeng and Xie, 2008; Bonhomme et al., 2016). Our contribution is not the generic observation that unequal opportunities can produce unequal choices. We operationalize this distinction in a centralized priority-rationed market by measuring mechanism-generated individual opportunity sets along two dimensions: their average level and the multidimensional structure of the trade-offs they contain. Applying a common behavioral operator then provides a transparent measure of how much allocation separation differently structured menus can reproduce. Because this second-stage attribution depends on the behavioral benchmark, we make that dependence explicit rather than presenting the resulting components as invariant structural parameters.

Third, we connect the opportunity-set framework to research on selective higher education, geographic frictions, and affirmative action. Most closely, Kirkeboen et al. (2016) use applicant rankings and admission cutoffs in centralized Norwegian admissions to identify causal returns to fields and institutions. We use route-specific admission cutoffs for a different empirical purpose: to reconstruct candidate-specific attainable programs and study how inequality in those opportunities maps into observed allocations. Geography shapes which colleges students apply to and attend (Long, 2004; Gibbons and Vignoles, 2012), while affirmative-action rules alter access to selective educational opportunities (Bertrand et al., 2010; Bagde et al., 2016; Mello, 2022). We add a distinct margin: priority and access policies can change not only the average quality of attainable alternatives but also the *configuration of trade-offs* over which participants subsequently choose. The female-seat application provides a policy-induced counterpart to the main opportunity-set analysis by tracing targeted program-level capacity through the assignment mechanism into a female-specific admission frontier and the realized use of the portion of that frontier beyond the corresponding Gender-Neutral cutoff.

Two features define the scope of the evidence. First, our analysis is conditional on candidates who ultimately receive an IIT allocation. We do not observe the complete outside-option set, including NITs, private institutions, foreign universities, or non-enrollment. Off-platform alternatives can be economically consequential in centralized assignment systems (Kapor et al., 2024), and without them we do not calculate a complete candidate-surplus or social-welfare counterfactual. Second, predetermined priority is not randomly assigned. The priority gradients and the main decomposition characterize how opportunity and observed sorting differ across the priority distribution; they are

not causal effects of examination rank on preferences.

The central empirical message is therefore about the anatomy of allocation inequality rather than a causal effect of rank. In a priority-rationed market, priority and eligibility generate the feasible opportunity set; candidates choose within that set; and the assignment mechanism translates those choices into final allocations. The mechanism determines more than how good an alternative a participant can reach. It determines the set of trade-offs over which preferences can be expressed. Empirical analyses that observe final assignments without recovering that constraint risk attributing to preferences what is embedded in opportunity; analyses that summarize the constraint only by average quality can miss the structure through which those opportunities translate into allocations.

Sections 2–4 describe the JoSAA mechanism, the data and reconstruction of candidate-specific feasible menus, and the empirical framework. Section 4.3 and Section 5 present revealed program trade-offs and the main opportunity–structure–choice decomposition. Section 6 studies the female-supernumerary-seat reform, Section 7 discusses interpretation, policy implications, and the scope of the evidence, and Section 8 concludes.

## 2 Institutional Setting and the JoSAA Assignment Mechanism

Undergraduate admission to the Indian Institutes of Technology (IITs) operates through a large centralized priority-based assignment market. Candidates first compete for positions on one or more JEE (Advanced) merit lists and then submit an ordered list of IIT academic programs. The Joint Seat Allocation Authority (JoSAA) combines these submitted rankings with program capacities, category-specific eligibility, merit-list priorities, and other institutional rules to determine assignments through a candidate-proposing Multi-Run Multi-Round Deferred Acceptance mechanism (Baswana et al., 2019a).

The institutional feature most important for our analysis is that a candidate's allocation is jointly determined by two conceptually distinct objects: the candidate's ordering of academic programs and the set of programs that her priority and eligibility permit her to obtain. Candidates facing the same program characteristics can therefore receive different assignments because they face different institutional opportunity sets; conversely, candidates with similar opportunities can obtain different programs because they rank the available trade-offs differently.

This section develops the institutional foundation for separating those two margins. We first describe JEE (Advanced), program choice submission, and the multiple merit lists through which candidates compete. We then explain how JoSAA represents reservation and eligibility rules through virtual program-sections, how candidate-proposing deferred acceptance operates over those objects, and how repeated rounds and de-reservation modify the static allocation environment. We describe the female-supernumerary-seat reform and the principal institutional changes during 2015–2023. The section concludes by formalizing the mapping from route-specific priority to candidate-specific institutional opportunity.

## 2.1 JEE Advanced and centralized IIT admissions

Admission to undergraduate IIT programs is based on performance in the Joint Entrance Examination (Advanced), subject to the eligibility requirements in force in the relevant admission cycle. Qualified candidates can appear on one or more merit lists. The Common Rank List (CRL) provides the priority ordering used for OPEN seats. Eligible candidates can additionally appear on category-specific merit lists, including OBC-NCL, SC, and ST lists and their PwD counterparts. Beginning in 2019, GEN-EWS and GEN-EWS-PwD merit lists and program-sections entered the allocation architecture (Joint Seat Allocation Authority, 2019).

A lower numerical rank represents stronger priority. Importantly, the merit lists constitute distinct priority orders rather than transformations of a single universal rank. A reserved-category candidate can, for example, possess both a CRL rank and an own-category rank and can compete for the same physical academic program through more than one institutional route. The rank governing one route need not therefore order candidates in the same way as the rank governing another. This multiplicity of merit lists is a central feature of the JoSAA design (Baswana et al., 2019a).

Since 2015, IIT allocation has been incorporated into a common JoSAA process that also coordinates admissions to the NIT+ system. Before this reform, IITs and the other centrally funded technical institutions operated separate allocation processes. Candidates could consequently hold multiple offers, generating vacancies when one was subsequently rejected and potentially preventing seats from reaching candidates who would have accepted them. The 2015 reform created a common allocation window and a unified mechanism that assigns at most one program to each participating candidate (Baswana et al., 2019a, 2015).

Our empirical analysis focuses on the IIT side of this centralized market. This restriction removes several institutional dimensions that are central to NIT+ allocation but not to the IIT market, most notably Home-State and Other-State quotas. IIT academic programs operate under an All-India allocation structure while retaining category-specific reservations, PwD provisions, preparatory-course rules, female seat pools after 2018, and other special eligibility channels.

Throughout the paper, an *academic program* is an institute–degree-program combination, such as B.Tech. in Mechanical Engineering at IIT Kanpur. This definition follows the JoSAA Business Rules, which treat the institute–program combination as the substantive alternative ranked by the candidate (Joint Seat Allocation Authority, 2023). The assignment algorithm operates at a more disaggregated level: as described below, a single physical academic program can be represented internally by multiple *program-sections* or *virtual programs* that implement the applicable reservation, gender-pool, and other eligibility rules.

## 2.2 Choice submission

Candidates do not submit separate applications to individual IIT programs. Instead, each participating candidate submits one ordered list of academic programs through the JoSAA portal.

During the choice-filling period, candidates can add eligible programs and rearrange their relative ordering. JoSAA also conducts mock seat allocations before the final deadline, allowing candidates to observe the implications of tentative submissions and revise their choices before the list is locked. The JoSAA rules do not impose a marginal monetary fee for listing additional program choices, although search, information, and cognitive costs may still affect the submitted list.

At the deadline, the submitted list becomes fixed. A candidate who does not actively lock the list has the last saved version locked automatically, and after the deadline candidates cannot change the ordering, add programs, or remove programs. The program ranking that enters the allocation process is therefore determined before the realized closing ranks of that admission cycle are observed (Joint Seat Allocation Authority, 2023).

For candidate $i$ in admission cycle $t$, let

$$P_{it} = \left(j_{i1t}, j_{i2t}, \ldots, j_{iL_{it}t}\right) \tag{2.1}$$

denote the ordered list of the $L_{it}$ physical academic programs submitted by the candidate, with $j_{i1t}$ the most-preferred listed program. JoSAA combines this ordering with the candidate's eligibility and priority on the relevant merit lists to seek the highest-ranked listed program that can be assigned under the institutional rules.

The incentive interpretation of the submitted list requires care. Under the standard candidate-proposing deferred-acceptance environment, truthful ordering of the alternatives a candidate chooses to list is strategy-proof under the usual assumptions: strategically lowering a preferred attainable alternative cannot improve the candidate's assignment (Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003). The JoSAA algorithmic analysis establishes the corresponding incentive property for the regular virtual-program construction used by the mechanism (Baswana et al., 2015, 2019a).

This property does not imply that the submitted list is a complete behavioral ranking over every IIT academic program. Candidates can omit programs they would not attend, possess alternatives outside JoSAA, update information during the admission process, reject an allocation, or later withdraw from the centralized process. Accordingly, the mechanism gives a revealed-ordering interpretation to programs that are both present on the relevant submitted list and institutionally attainable. Our empirical analysis subsequently uses a broader reconstructed institutional opportunity set $F_{it}$; Section 4 states the additional maintained choice-set interpretation required for the conditional-choice model.

### 2.3 Reservation categories, multiple merit lists, and virtual programs

A direct application of textbook deferred acceptance is insufficient for JoSAA because seats within a physical academic program are governed by multiple eligibility and priority rules. The mechanism implements these rules by representing each physical academic program as a collection of virtual programs, referred to in later Business Rules as program-sections (Baswana et al., 2015, 2019a).

Consider an IIT academic program $j$ before the introduction of GEN-EWS and female seat pools. Its capacity is partitioned across OPEN, OBC-NCL, SC, and ST program-sections and their applicable PwD counterparts. Each section carries the capacity and merit list associated with that route. An OPEN program-section uses the Common Rank List; an OBC-NCL program-section uses the OBC-NCL merit list; an SC program-section uses the SC merit list; and so forth.

Candidate eligibility is translated into a corresponding virtual preference sequence. A reserved-category candidate who also qualifies on the CRL is not sent directly to a reserved seat. The mechanism first considers the candidate through the applicable OPEN route and then, if necessary, through the candidate's eligible reserved-category route. For a non-PwD OBC-NCL candidate, for example, physical academic program $j$ can be represented internally as

$$j^{OPEN} \succ j^{OBC},$$

where the two virtual objects lead to the same physical academic program but use different capacities and merit lists. Analogous virtual refinements implement SC, ST, PwD, preparatory-course, and other eligibility provisions.

The virtual representation serves two distinct purposes. First, it preserves the candidate's ordering over *physical academic programs*: an OPEN seat and an OBC-NCL seat in the same institute–degree program are not treated as two substantively different college alternatives. Second, it embeds reservation and eligibility rules into the priority system used by deferred acceptance.

This distinction is central to our empirical construction. JoSAA allocates over virtual program-sections, whereas the choice alternatives in our empirical analysis are physical academic programs. We therefore evaluate attainability at the route level and then collapse all qualifying routes associated with the same institute–program combination into one candidate–program alternative.

JoSAA's use of ordered eligibility routes also places the institution within the broader matching literature on affirmative action and reserve design. Aygün and Turhan (2017) study large-scale affirmative action in IIT admissions, while Sönmez and Yenmez (2022) analyze the interaction of vertical, horizontal, and overlapping reservations in India. We take the implemented JoSAA route architecture as an institutional input rather than proposing an alternative reserve mechanism.

### 2.4 Candidate-proposing deferred acceptance

Conditional on a set of capacities, virtual program-sections, candidate rankings, eligibility conditions, and route-specific merit lists, a regular run of the JoSAA allocation algorithm follows the logic of candidate-proposing deferred acceptance (Baswana et al., 2019a).

Abstracting temporarily from the additional business rules, every candidate begins unassigned and proposes to the highest-ranked eligible virtual route associated with her most-preferred remaining physical academic program. Each program-section tentatively retains the highest-priority candidates up to its capacity and rejects lower-priority proposals. Rejected candidates proceed to the next eligible route implied by their submitted ordering and virtual preference

sequence. The process continues until no rejection generates a further proposal.

The candidate's submitted ordering and institutional priority therefore play different roles. Preferences determine the sequence in which alternatives are sought. Eligibility and route-specific priority determine whether a candidate can be retained when a particular program-section is contested.

For fixed capacities and responsive program priorities, candidate-proposing DA produces the candidate-optimal stable allocation in the standard college-admissions environment (Gale and Shapley, 1962; Abdulkadiroğlu and Sönmez, 2003). Baswana et al. adapt this structure to JoSAA's multiple merit lists, virtual programs, and institutional business rules (Baswana et al., 2019a). For the regular nested route architecture, the construction implements the corresponding fairness, non-wastefulness, and candidate-optimality properties.

These statements should not be read as asserting that every historical JoSAA rule reduces to an ordinary fixed-capacity DA market. In particular, the Defence Service (DS) rule introduces a nonnested common quota for which a stable matching need not exist under the usual formulation. JoSAA handles that case using a separate rejection-chain procedure and, where necessary, supernumerary allocation (Baswana et al., 2019a). Preparatory-course allocations and several other special channels also involve additional rules. Our formal result below is therefore stated for the regular route-level environment to which the standard priority logic applies.

The actual historical mechanism is also richer than a single static run. Candidate continuation decisions, document verification, de-reservation, Min-Cutoff protections, credential changes, and the female-supernumerary architecture modify the effective input to subsequent runs and rounds. We do not invoke a generic textbook stability result as if the complete 2015–2023 process were one unchanging static market. Where these institutional extensions matter, we rely on the JoSAA-specific algorithmic construction and the Business Rules in force in the relevant admission cycle.

### 2.5 Multiple rounds: freeze, float, slide, and withdrawal

A JoSAA admission cycle contains multiple allocation rounds because candidates can accept, reject, or withdraw from seats and can remain eligible for higher-ranked alternatives after accepting a provisional allocation. Candidates who accept a seat select among continuation options defined by the JoSAA Business Rules (JoSAA, 2023).

Under *Freeze*, the candidate accepts the current academic program and leaves further upgrading. Under *Float*, the candidate accepts the current allocation but remains eligible for programs appearing above it on the locked choice list, including programs at other institutes. Under *Slide*, the candidate accepts the current allocation but remains eligible for a more-preferred academic program at the same institute.

These actions do not permit candidates to rewrite the original preference ordering after the choice-filling deadline. Rather, they determine which portion of the locked ordering remains active in subsequent rounds. A frozen candidate exits further upgrading; a candidate choosing Float or Slide remains active only for the subset of higher-ranked programs consistent with the selected

continuation rule.

A candidate can also reject an allocation and leave the subsequent process. Beginning in 2016, candidates who had previously accepted a seat were additionally permitted to *withdraw* before the applicable deadline, returning that seat for reallocation. The inaugural 2015 process did not contain this withdrawal option (Baswana et al., 2019b).

The distinction between a *round* and a *run* is important. A round is a stage of the market separated by candidate actions and administrative updates. A run is an execution of the deferred-acceptance algorithm for a fixed set of current inputs. JoSAA can therefore execute multiple DA runs within a single allocation round, which is why the implementation is described as Multi-Run Multi-Round Deferred Acceptance (MRDA) (Baswana et al., 2015, 2019b).

The multi-round implementation also contains protections for candidates who retain an earlier allocation while seeking an upgrade. In particular, Min-Cutoff information carried from the preceding round is used so that participation in a later round does not mechanically cause a continuing candidate to lose an earlier protected allocation merely because the market is rerun. Document verification and other credential changes can nevertheless alter eligibility, and rejection, withdrawal, or other explicitly specified candidate actions can remove a prior assignment (Baswana et al., 2015, 2019b).

For our purposes, the key implication is that historical attainability is generated by a sequence of route-specific priority frontiers rather than by one arbitrarily selected round. This motivates the any-round reconstruction in Section 3.2 and helps explain why a small number of valid administrative allocations are not reproduced by a mechanical comparison with one published cutoff record.

### 2.6 De-reservation and multiple DA runs

Reservation rules create an additional dynamic margin within an allocation round. Under the applicable Business Rules, specified unfilled program-sections can be reallocated to designated parent sections. The precise mapping depends on the category, PwD and preparatory-course rules, and the business rules in force in the relevant admission cycle (Baswana et al., 2019a).

JoSAA implements these reallocations through repeated runs of deferred acceptance. After one run terminates, seats eligible for de-reservation are transferred according to the applicable route hierarchy, program-section capacities are updated, and DA is run again. The procedure continues until no further eligible de-reservation remains.

The JoSAA algorithmic analysis establishes a useful monotonicity property across these de-reservation runs: the process expands rather than contracts the options made available by the de-reservation sequence to candidates who remain active. Seats flow toward the specified parent program-sections, and subsequent runs can leave a candidate with the same or a more-preferred allocation rather than worsening the allocation solely because another de-reservation run occurs (Baswana et al., 2019b).

De-reservation is another reason why a candidate cannot generally be represented by one scalar rank and one program cutoff. The same physical academic program can be reached through

different program-sections governed by different merit lists, and the relevant capacity can change across runs. Our empirical reconstruction therefore evaluates qualification at the route level and takes the union of the routes through which a candidate can reach the same physical academic program.

### 2.7 Female supernumerary seats

A major change to IIT allocation during our sample is the introduction of female supernumerary seats in 2018. The reform modified both program capacities and the route structure through which female candidates compete.

Beginning in 2018, applicable academic programs were divided into Gender-Neutral and Female-Only pools. Non-female candidates could compete only through Gender-Neutral program-sections. Female candidates could compete through both pools. Within a given category route, a female candidate is considered first through the Female-Only section and, if she is not retained there, through the corresponding Gender-Neutral section. The mechanism was designed so that the creation of female-specific capacity did not eliminate the ability of a sufficiently high-priority female candidate to compete for the Gender-Neutral pool (Baswana et al., 2019b).

For a female OBC-NCL candidate who also appears on the CRL, suppressing PwD routes, the virtual sequence associated with physical academic program $j$ can therefore be represented schematically as

$$j^{OPEN,FO} \succ j^{OPEN,GN} \succ j^{OBC,FO} \succ j^{OBC,GN}.$$

The OPEN route is considered before the candidate's applicable reserved-category route, while within each category route the Female-Only section precedes the corresponding Gender-Neutral section. The physical academic program remains the same throughout this virtual refinement.

The initial 2018 seat construction was linked to prior female representation at the academic-program level. Let $C$ denote the program's 2017 capacity, $f$ the number of female candidates allocated to that program in 2017, and $x$ the additional capacity associated with reaching the baseline 14-percent female target. For programs governed by the below-target construction, $x$ satisfies

$$f + x = 0.14(C + x), \tag{2.2}$$

subject to the rounding and institute-specific target provisions in the implementation. Programs whose prior female representation lay between 14 and 20 percent, programs above 20 percent, and newly introduced programs were handled under the corresponding cases specified by the 2018 algorithmic rules (Baswana et al., 2019b).

This construction is distinct from a conventional reservation that merely reallocates a fixed stock of existing seats. Female-Only program-sections can contain genuinely supernumerary capacity, while female candidates retain access to the Gender-Neutral pool through the prescribed virtual

sequence. The designers explicitly distinguish this architecture from an ordinary female reservation (Baswana et al., 2019b).

The female target subsequently increased. The 2019 JoSAA Business Rules incorporated GEN-EWS program-sections and raised the female representation target to at least 17 percent, while the target reached at least 20 percent in 2020 (JoSAA, 2018, 2020). Later rules continued to operate with distinct Female-Only and Gender-Neutral pools. Section 6 returns to the 2018 introduction of this architecture and exploits the program-level seat-construction rule to study how incremental capacity changes female-specific access frontiers and realized allocations.

### 2.8 Institutional changes during 2015–2023

The candidate-proposing MRDA architecture remains the organizing mechanism throughout our 2015–2023 sample, but the institutional environment evolves in ways that affect candidate eligibility routes and program-section capacities. The inaugural 2015 process introduced joint allocation and multi-run de-reservation; the Withdraw option entered in 2016; Female-Only and Gender-Neutral program-sections entered in 2018; GEN-EWS routes entered in 2019, when the female target rose to at least 17 percent; and the female target reached at least 20 percent in 2020 (Baswana et al., 2019b; JoSAA, 2020).

Other implementation details, including the number and timing of rounds, document-verification rules, and special-route provisions, also vary across cycles. We therefore do not impose one year's allocation rules on another. Eligibility routes, merit lists, seat pools, and observed admission frontiers are reconstructed separately for each admission cycle using the corresponding year-specific JoSAA rules.

### 2.9 From route-level priority to institutional opportunity

The preceding institutional description highlights the distinction underlying our empirical analysis. Candidates rank *physical academic programs*, but eligibility and priority operate through the virtual program-sections associated with those programs. We therefore distinguish the object over which the candidate submits an ordering from the institutional routes through which that object can be obtained.

Fix admission cycle $t$. Let $\mathcal{I}_t$ denote the set of candidates, $\mathcal{J}_t$ the set of physical IIT academic programs, and $\mathcal{R}_{jt}$ the collection of regular virtual routes associated with program $j$. A route can encode, for example, OPEN, OBC-NCL, SC, ST, or GEN-EWS eligibility and, after 2018, the relevant Female-Only or Gender-Neutral pool.

For candidate $i$, let

$$e_{irt} \in \{0, 1\} \tag{2.3}$$

indicate whether the candidate is eligible for route $r$, and let $q_{irt}$ denote the candidate's position on the merit list governing that route. Smaller numerical ranks represent stronger priority. Because

a candidate can appear on more than one merit list, $q_{irt}$ is intrinsically route specific.

For a fixed regular allocation environment, let $\kappa_{jrt}$ denote the closing priority for route $r$ to physical academic program $j$: the weakest priority admitted through that program–route combination. When the applicable route remains unfilled, the cutoff can be represented as $+\infty$. JoSAA's treatment of ties at the closing rank implies weak qualification,

$$q_{irt} \leq \kappa_{jrt}. \tag{2.4}$$

Let

$$\boldsymbol{\kappa}_t = \left\{\kappa_{jrt}\right\}_{j,r}$$

denote the complete vector of program–route cutoffs in admission cycle $t$. Candidate $i$'s cutoff-qualified institutional opportunity set is

$$Q_{it}(\boldsymbol{\kappa}_t) = \left\{j \in \mathcal{J}_t : \exists r \in \mathcal{R}_{jt} \text{ such that } e_{irt} = 1 \text{ and } q_{irt} \leq \kappa_{jrt}\right\}. \tag{2.5}$$

The object $Q_{it}(\boldsymbol{\kappa}_t)$ is institutional. It records the physical academic programs for which the candidate clears at least one eligible route-specific priority frontier. It does not imply that every such program appears on the candidate's submitted list, was actively considered, or was regarded as acceptable.

To state the revealed-ordering implication precisely, let $\mathcal{L}_{it} \subseteq \mathcal{J}_t$ denote the set of physical academic programs remaining on candidate $i$'s active submitted list in the regular allocation environment under consideration, and let $\succeq_{it}$ denote the weak ordering induced by that list. Let $a_{it}$ denote the physical academic program assigned to candidate $i$ in that environment.

**Proposition 1** (Cutoff qualification in the regular route-level market)**.** *Consider a candidate-proposing DA allocation environment in admission cycle $t$ with fixed route capacities, eligibility, route-specific merit lists, and program–route cutoffs $\boldsymbol{\kappa}_t$. Suppose that:*

*(i) the institutional virtual-route refinement preserves the candidate's ordering over physical academic programs;*

*(ii) candidate $i$ remains active in the allocation environment and her route eligibility is correctly specified;*

*(iii) route priorities are implemented according to the applicable merit lists, including the institutional treatment of candidates tied at the closing priority; and*

*(iv) allocation terminates under the regular candidate-proposing DA rules.*

*If candidate $i$ receives physical academic program $a_{it}$, then*

$$a_{it} \succeq_{it} j \qquad \forall j \in Q_{it}(\boldsymbol{\kappa}_t) \cap \mathcal{L}_{it}. \tag{2.6}$$

*Equivalently, among the programs on the candidate's active submitted list whose route-specific priority*

*frontiers she clears, the allocation gives candidate $i$ her most-preferred physical academic program.*

A proof adapted to JoSAA's virtual-route representation is provided in Appendix A.2.

The proposition is an institution-specific application of standard stability and priority logic rather than a new matching theorem. Its revealed-ordering interpretation is closely related to the ex-post qualification arguments in Fack et al. (2019) and Arslan (2021): observed admission frontiers and ex ante priorities can identify alternatives for which an applicant qualifies, permitting preference information to be recovered among relevant qualified alternatives. The JoSAA application is more institutionally elaborate because qualification can operate through multiple merit lists and multiple virtual routes to the same physical academic program.

Proposition 1 also clarifies what route-specific closing ranks do *not* establish. They do not imply that every cutoff-qualified program was listed or actively considered. Nor do they imply that one scalar rank globally orders candidates' opportunity sets. Different candidates can compete through different merit lists, reservation routes, and gender pools.

The result is deliberately local to the regular route-level environment. The complete historical JoSAA process contains repeated runs, multiple rounds, de-reservation, Min-Cutoff protections, credential changes, preparatory-course allocations, and special routes such as the DS common quota. We therefore do not infer a single global stability theorem for every business rule used during 2015–2023 from Proposition 1. Appendix A.1 maps these institutional qualifications to the analytical sample and the empirical reconstruction.

For admission cycle $t$, let

$$\mathcal{R}_t = \bigcup_{j \in \mathcal{J}_t} \mathcal{R}_{jt} \tag{2.7}$$

denote the collection of regular institutional route types represented in the market. For candidate $i$, define the set of routes for which the candidate is eligible as

$$\mathcal{E}_{it} = \{r \in \mathcal{R}_t : e_{irt} = 1\}. \tag{2.8}$$

Because JoSAA priority is route specific, comparing two candidates requires more than comparing one rank. We say that candidate $i$ *routewise dominates* candidate $k$ in admission cycle $t$ if every route for which candidate $k$ is eligible is also available to candidate $i$,

$$\mathcal{E}_{kt} \subseteq \mathcal{E}_{it}, \tag{2.9}$$

and candidate $i$ has weakly stronger priority on every route available to candidate $k$,

$$q_{irt} \leq q_{krt} \qquad \forall r \in \mathcal{E}_{kt}. \tag{2.10}$$

**Lemma 1** (Routewise priority monotonicity)**.** *Fix admission cycle $t$ and a common program–route cutoff environment $\boldsymbol{\kappa}_t$. If candidate $i$ routewise dominates candidate $k$ according to (2.9)–(2.10), then*

$$Q_{kt}(\boldsymbol{\kappa}_t) \subseteq Q_{it}(\boldsymbol{\kappa}_t). \tag{2.11}$$

*Thus, conditional on a common set of program–route admission frontiers, weakly stronger priority on every route available to another candidate cannot reduce the candidate's cutoff-qualified institutional opportunity set.*

A proof is provided in Appendix A.3.

Lemma 1 formalizes the institutional link from route-specific priority to opportunity. Conditional on a fixed cutoff environment, a candidate who possesses every eligibility route available to another candidate and has weakly stronger priority on each such route can attain every physical academic program available to the latter.

The result is deliberately conditional on $\boldsymbol{\kappa}_t$. It is a monotonicity property of cutoff-qualified opportunity sets, not a causal comparative-static exercise in which one candidate's examination performance is changed and the entire JoSAA market is re-solved. Such a change could alter rejection chains, assignments, and the admission frontiers themselves.

Nor does the lemma imply that candidates' opportunities are generally nested according to one scalar rank. JoSAA priority is multidimensional: candidates can differ in merit-list membership, reservation-category eligibility, gender-pool eligibility, and other route-specific qualifications. When neither candidate routewise dominates the other, their institutional opportunity sets need not be nested.

This distinction is important for the empirical analysis. The predetermined P1 measure introduced in Section 4.5 is a scalar summary used to organize candidates for comparisons across the priority distribution; it is not a sufficient statistic for the complete route-specific priority vector. The empirical relationship between P1 and reconstructed opportunities is therefore an object to be measured rather than a mechanical implication of Lemma 1.

The institutional results establish the mapping needed for the empirical analysis. Priority and eligibility operate at the route level and generate candidate-specific constraints over physical academic programs. Among cutoff-qualified programs that remain on a candidate's active submitted list, the regular candidate-proposing allocation gives the observed assignment a revealed-ordering interpretation. Neither result requires the institutional opportunity set to coincide with the candidate's behavioral consideration set.

Section 3 therefore takes the route-specific priority architecture as an institutional input rather than attempting to infer opportunity from the candidate's final assignment alone. Using the historical opening- and closing-rank records, we construct the direct empirical counterpart $Q_{it}^{D}$ of the formal cutoff-qualified set $Q_{it}(\boldsymbol{\kappa}_t)$, and then form the final empirical menu $F_{it}$ after the minimal revealed-feasible augmentation described there. The resulting object measures institutional attainability; the behavioral interpretation of choice conditional on that reconstructed menu is developed separately in Section 4.

# 3 Data and Reconstruction of Candidate Opportunity Sets

Our empirical analysis combines administrative records from the IIT admissions process with program-level labor-market outcomes, institutional-quality measures, and geographic information. The central measurement problem is not simply to observe where candidates are allocated, but to recover the set of IIT academic programs lying within each candidate's observed institutional priority frontier under the eligibility and allocation rules of the centralized mechanism. We use candidate rank and allocation records together with published opening- and closing-rank data to reconstruct these candidate-specific opportunity sets.

The resulting data have a candidate–program structure. For each admitted candidate, we observe the academic program ultimately allocated through JoSAA and construct a set of alternative academic programs lying within at least one eligible priority-route frontier. We then attach to every candidate–program pair measures of geographic distance, program compensation, placement outcomes, and institutional quality. This produces the long candidate–program panel used in the empirical analysis.

Two features of the construction are important at the outset. First, the analysis is conditional on candidates who ultimately receive an IIT allocation. We therefore study sorting *within the IIT market* rather than the extensive margin of IIT admission, and we do not observe a complete outside option containing NITs, private institutions, foreign universities, or non-enrollment. Second, our preferred opportunity sets are reconstructed from admission frontiers realized during the same allocation cycle. They should therefore be understood as measures of *realized institutional opportunity*, not as exact representations of candidates' ex ante beliefs about which programs they expected to obtain. We address this distinction separately using lagged program information and prior-year admission frontiers.

## 3.1 Administrative candidate and allocation records

The candidate records cover nine IIT admission cycles, from 2015 through 2023. For each candidate, the administrative files contain the available JEE rank information, social category, examination-center information, and the academic program ultimately allocated through the centralized process. The underlying records distinguish the candidate's own social category from the seat category through which the realized allocation occurs. We preserve that distinction throughout the analysis.

This distinction matters because JoSAA priority is route specific. A candidate from a reserved category can potentially qualify for an OPEN program-section using the Common Rank List or All India Rank and can also qualify through the corresponding reserved-category merit list. Beginning in 2018, academic programs are additionally partitioned into Gender-Neutral and Female-Only seat pools, and GEN-EWS enters the allocation architecture beginning in 2019. Neither the candidate's own category nor the category of the ultimately allocated seat is therefore sufficient to characterize the candidate's attainable programs.

We restrict the baseline analysis to the regular IIT allocation market. Preparatory-course, PwD,

Defence Service, malformed or unresolved seat routes, and other special allocation channels for which the regular rank–frontier mapping cannot be implemented comparably are excluded from the baseline reconstruction. From 2018 onward, we additionally require an administratively identified Gender-Neutral or Female-Only route where the pool distinction is relevant. These exclusions do not treat the corresponding institutional channels as unimportant; they preserve a common interpretation of the rank–eligibility–frontier relationship used to construct the baseline opportunity sets. Appendix A.1 documents the institutional rationale, and Table A6 summarizes the analytical treatment of each regular and special route.

The final reconstructed market contains 23 IITs and 353 distinct institute–academic-program combinations. An academic program is defined as an IIT–degree-program or IIT–discipline combination. Different reservation categories and gender pools associated with the same physical institute–program combination are treated as different institutional *routes* to the same academic program rather than as different choice alternatives.

Our principal sample contains 119,285 candidate–cycle observations for whom the final reconstructed opportunity set contains at least two academic programs. The number of candidates grows from 9,684 in 2015 to 16,986 in 2023, alongside expansion in IIT capacity and changes in the academic programs represented in the market. Appendix Table A2 reports sample composition by admission cycle and candidate category.

The restriction to IIT recipients is central to interpretation. We do not claim that the final sample represents all students who sit for JEE Advanced, all candidates who qualify for IIT admission, or all students choosing among engineering colleges in India. It represents the population for whom we observe a realized IIT allocation and can construct a meaningful within-IIT opportunity set. The choice analysis therefore characterizes how these candidates sort among reconstructed IIT opportunities. It does not identify the value of IIT admission relative to an outside option and, by itself, cannot support welfare calculations requiring the complete applicant market.

### 3.2 Opening and closing ranks

The second administrative source is the JoSAA opening- and closing-rank data. These records report, by admission cycle and allocation round, the merit-list frontier associated with an academic program and the relevant seat category and, where applicable, gender pool. Throughout the paper, we refer to these records as the OR–CR data.

Closing ranks provide the empirical link between institutional priority and program attainability. Ranks are ordered so that smaller numerical values represent stronger priority. For a given physical academic program and eligible route, a candidate whose relevant merit-list rank is weakly better than the published closing rank lies within the observed priority frontier for that program–route combination. Because a candidate can be eligible through multiple routes, the comparison is performed separately for each applicable route.

For OPEN routes, the relevant ranking is the AIR/CRL ranking. For GEN-EWS, OBC-NCL, SC, or ST routes, we use the corresponding category-specific merit-list rank. When gender-specific

program-sections apply, the route is further distinguished by the relevant Gender-Neutral or Female-Only pool. A candidate need clear the observed frontier through only one eligible route for the underlying physical academic program to enter the direct reconstructed opportunity set.

The OR–CR data are round specific. Let $\mathcal{S}_{jrt}$ denote the set of allocation rounds in admission cycle $t$ for which a valid published closing rank is observed for physical academic program $j$ and regular route $r$, and let $\kappa_{jrts}$ denote that closing rank in round $s$. We define the direct historical frontier for the program–route pair as

$$\kappa^{D}_{jrt} = \max_{s \in \mathcal{S}_{jrt}} \kappa_{jrts}, \tag{3.1}$$

where a larger numerical closing rank corresponds to a more permissive frontier because smaller numerical merit-list ranks indicate stronger priority. Thus, $\kappa^{D}_{jrt}$ records the farthest observed extension of the published program–route frontier during admission cycle $t$.

The direct empirical frontier is defined only for regular program–route combinations for which a valid published frontier can be matched to the relevant merit list. We do not mechanically equate absence of a numeric OR–CR record with the theoretical $+\infty$ convention for an unfilled route. Special, malformed, or otherwise noncomparable records are treated according to the analytical rules documented in Appendix A.1.

The use of the most permissive observed frontier is intentional. JoSAA is a multi-round mechanism in which seats can be accepted, rejected, frozen, floated, slid, withdrawn, and reallocated. A single arbitrarily selected round therefore need not summarize the full admission frontier realized during the historical process.

The opening rank describes the strongest rank among candidates assigned through a particular route and is useful for characterizing the realized distribution of allocations. It is not required for determining whether a candidate lies inside the relevant closing-rank frontier. The opportunity-set reconstruction therefore relies on closing ranks.

### 3.3 Reconstructing candidate-specific feasible menus

Section 2.9 defines $Q_{it}(\boldsymbol{\kappa}_t)$, the cutoff-qualified institutional opportunity set for candidate $i$ in admission cycle $t$, and distinguishes that object from the candidate's submitted or behavioral consideration set. Our empirical reconstruction implements the corresponding route-specific qualification logic using the historical admission frontiers observed in the JoSAA OR–CR records.

Let

$$\boldsymbol{\kappa}^{D}_{t} = \left\{ \kappa^{D}_{jrt} \right\}_{j,r}$$

denote the collection of direct empirical program–route frontiers in admission cycle $t$. We define candidate $i$'s direct reconstructed institutional opportunity set as

$$Q_{it}^{D} = \left\{ j \in \mathcal{J}_t : \exists r \in \mathcal{R}_{jt} \text{ such that } e_{irt} = 1 \text{ and } q_{irt} \leq \kappa_{jrt}^{D} \right\}. \tag{3.2}$$

On the observed support of regular routes, this is the direct empirical counterpart of the formal object developed in Section 2:

$$Q_{it}^{D} = Q_{it}\left(\boldsymbol{\kappa}_t^{D}\right). \tag{3.3}$$

Thus, a physical academic program enters $Q_{it}^{D}$ when candidate $i$ lies inside the observed admission frontier through at least one eligible route during admission cycle $t$. Multiple qualifying routes to the same physical academic program are collapsed after route-level qualification is evaluated, so each attainable academic program appears only once in $Q_{it}^{D}$.

Taking the union across eligible routes is essential. An OBC-NCL candidate, for example, can qualify for the same physical academic program through an OPEN route using AIR/CRL or through an OBC-NCL route using the corresponding category merit-list rank. Beginning in 2018, the relevant route can additionally depend on the Gender-Neutral or Female-Only pool. The construction therefore uses the full multidimensional priority information available in the administrative data rather than attempting to represent attainability with one rank.

The any-round construction should be interpreted carefully. $\boldsymbol{\kappa}_t^{D}$ is an ex-post envelope of the route-specific admission frontiers realized during the historical JoSAA process. It is not a candidate-specific history of the offers literally presented to candidate $i$. In particular, a program can lie inside $Q_{it}^{D}$ because its eligible route frontier reached the candidate during some round even if the candidate's realized Freeze, Float, Slide, withdrawal, or earlier allocation path did not place her in position to receive that program in that particular round. We therefore interpret $Q_{it}^{D}$ as a frontier-based measure of realized institutional attainability rather than as a sequence of candidate-specific offers.

The object $Q_{it}^{D}$ is also not an observed behavioral consideration set. A program can lie inside the candidate's institutional priority frontier even if the candidate did not list it, actively consider it, or regard it as an acceptable alternative.[1] Our reconstruction addresses institutional availability. The maintained behavioral interpretation required for estimation is stated separately in Section 4.

The cutoff-based qualification logic is closely related to the use of applicant-specific qualifying or accessible alternatives for preference recovery in centralized admissions; see Section 2.9 and Fack et al. (2019); Arslan (2021). Our empirical use differs in that the resulting heterogeneous opportunity sets are themselves outcomes of interest rather than only conditioning objects for estimating preferences.

The routewise monotonicity result in Lemma 1 carries directly to this empirical construction conditional on the observed frontier environment. If candidate $i$ is eligible for every route available to candidate $k$ and has weakly stronger priority on each such route, then

[1] Related demand literatures distinguish institutional availability from behavioral consideration or choice-set formation; see, for example, Horowitz (1991) and Conlon and Mortimer (2013).

$$Q_{kt}^{D} \subseteq Q_{it}^{D}.$$

This inheritance follows because both candidates are evaluated against the same realized program–route frontier $\boldsymbol{\kappa}_{t}^{D}$.

The reconstruction is conceptually distinct from the scalar predetermined priority measure introduced in Section 4.5. That measure is used later to organize candidates along a common empirical priority distribution; it is not used to determine route-level qualification or to construct $Q_{it}^{D}$.

The reconstruction also respects the fact that candidates can possess different combinations of merit-list ranks. A reserved-category candidate with both AIR and an own-category merit-list rank can qualify through either eligible route. When the common-list rank is unavailable but the relevant category rank is observed, the category-specific route can still generate attainable programs. We do not impute a missing common-list rank from a category rank or a missing category rank from the common list.

The direct set $Q_{it}^{D}$ is therefore the empirical analogue of the cutoff-qualified institutional opportunity object developed in Section 2.9, implemented using the regular priority routes and realized historical frontiers observed in the JoSAA data. The next subsection addresses the small number of cases in which the administratively observed allocation is not recovered by this direct frontier-based reconstruction.

### 3.4 Revealed-feasible augmentation

Published OR–CR records do not encode every state variable and administrative guarantee used by the full Multi-Run Multi-Round Deferred Acceptance process. Features such as Min-Cutoff protections, supernumerary treatment, credential changes, de-reservation, ties, and other round- or route-specific business rules can therefore make the published frontier an incomplete representation of the historical mechanism's full state. We do not attempt to assign individual reconstruction exceptions to a particular institutional rule.

Let $a_{it}$ denote candidate $i$'s administratively observed final physical academic-program allocation. The final reconstructed opportunity set is

$$F_{it} = \begin{cases} Q_{it}^{D}, & a_{it} \in Q_{it}^{D}, \\ Q_{it}^{D} \cup \{a_{it}\}, & a_{it} \notin Q_{it}^{D}. \end{cases} \tag{3.4}$$

Thus, whenever the published frontier fails to recover the realized allocation, we add exactly one program: the program the historical administrative mechanism demonstrably assigned. No other alternative is introduced through revealed augmentation.

This operation makes a deliberately weak claim. It does *not* assert that the appended program satisfies a published closing-rank rule that we have failed to reconstruct, nor does it infer additional unobserved opportunities from the realized assignment. It asserts only that the documented

allocation is direct evidence that the candidate could obtain that physical academic program under the historical mechanism.

The magnitude of the augmentation is small. Among 119,285 candidate menus in the principal sample, the observed allocated program is recovered directly in $Q_{it}^{D}$ for 118,567 candidates, or 99.40 percent. Only 718 candidates, 0.60 percent, require the allocated program to be appended. Every candidate in the final sample has a nonempty direct frontier before augmentation.

The 99.40-percent direct recovery rate provides a strong mechanism-consistency diagnostic for the route-level reconstruction: for almost all candidates, the administratively observed allocation lies inside $Q_{it}^{D}$ without using that allocation to construct the direct set. We do not require 100-percent recovery because published OR–CR records are not a complete representation of every state variable and guarantee embedded in the historical MRDA implementation.

The exceptions are concentrated in 2019: 531 of the 718 augmented menus occur in that admission cycle. We do not attribute this concentration to a particular institutional cause. Appendix Table A3 reports the complete year-by-year distribution and the corresponding reconstruction audit.

Additional integrity checks verify that each final choice set contains exactly one realized allocated program, contains no duplicate candidate–program alternatives, and has a stored menu size equal to the number of physical academic-program alternatives represented in the long data. All retained principal-sample menus contain at least two academic programs. Appendix Table A3 reports these checks.

The final analytical menu $F_{it}$ is therefore generated almost entirely by the observed route-specific frontier reconstruction. Revealed augmentation is a minimal administrative correction, not a procedure for expanding menus using assumptions about unobserved opportunities.

### 3.5 Program labor-market outcomes and institutional quality

We merge the reconstructed candidate–program panel with two measures of program labor-market performance and one institute-level proxy for institutional quality.

Our primary program outcome is reported annual median cost to company (CTC), measured in lakh rupees per annum. CTC summarizes the annual compensation associated with the median reported placement outcome for the relevant IIT academic program; it is not a measure of realized individual earnings. We convert nominal CTC to constant 2023 rupees using the Consumer Price Index so that monetary outcomes are comparable across admission cycles. Throughout the paper, "CTC" without qualification refers to this real 2023 measure.

The second labor-market measure is the program placement rate, defined as the percentage of the relevant graduating cohort placed through the institute's placement process. CTC and placement capture related but distinct dimensions: reported compensation among placement outcomes and the fraction of the cohort receiving a reported placement.

A practical feature of both measures is that newly created academic programs may not yet have produced a graduating cohort. For such program-years, CTC and placement are unavailable outcomes rather than substantively observed zeros. Candidate-level descriptive statistics involving

CTC or placement are therefore calculated over programs for which the relevant outcome is observed.

In the choice-model estimation, alternatives lacking an observed graduating batch are retained using an explicit missing-outcome indicator together with the numerical normalization implemented in the estimation code. This separates the absence of an observed graduating cohort from a genuinely low observed program outcome. Appendix Table A1 reports the exact source, unit, timing, and estimation treatment of these variables.

Our third attribute is the National Institutional Ranking Framework (NIRF) Engineering score, which we use as an institute-level proxy for institutional quality. Unlike CTC and placement, NIRF varies at the IIT rather than the academic-program level, so programs offered by the same IIT in a given admission cycle share the same score. We use the underlying NIRF score rather than the published ordinal rank.

NIRF begins during our sample period, so the score is unavailable for some institute–cycle observations.

The choice-model implementation retains such observations using the corresponding missing-information treatment documented in Appendix Table A1.

Because CTC, placement, and NIRF differ in their level of aggregation and economic interpretation, we do not treat them as interchangeable measurements of one latent program-quality index. The empirical specifications allow the attributes to enter separately.

### 3.6 Candidate–IIT distance

Geographic distance provides the principal spatial trade-off in the choice model. We geocode the JEE examination center associated with each candidate and the location of each IIT. For every candidate–program alternative, we calculate the Haversine distance from the candidate's examination-center location to the IIT offering the program.

The examination center is a geographic proxy rather than a literal measure of the candidate's household residence. Its principal advantage is that it is observed consistently in the JEE administrative records and is recorded before the college-allocation process. We therefore refer to this measure as candidate–IIT distance and do not interpret it as exact residential distance.

Distance varies across IITs within a candidate's opportunity set but, by construction, is identical across physical academic programs offered by the same IIT. This feature becomes relevant in the IIT-nested-logit robustness analysis, where geographic distance operates at the institute margin while CTC and placement can vary across programs within the same institution.

The preferred model sample contains valid distance information for 119,259 candidate choice sets. Relative to the 119,285 reconstructed menus, the distance requirement therefore removes only a small number of candidate choice sets. The corresponding estimation panel contains 18,629,219 candidate–program alternatives.

### 3.7 Information timing

The empirical objects used in the paper differ in their information timing. Merit-list positions and examination-center information are observed before JoSAA preference submission. Candidates also declare category status before choice filling, although final route eligibility can subsequently be affected by document verification and credential changes under the applicable Business Rules. NIRF scores are publicly released before the JoSAA choice-filling period in the sample years for which the relevant score is available. By contrast, contemporaneous CTC and placement information can be released close to, or overlap with, the choice-filling period in normal admission cycles.

Most importantly, current-cycle closing ranks are themselves equilibrium outcomes of the allocation process. The preferred $F_{it}$ should therefore not be interpreted as a perfectly known ex ante choice set. It is a reconstruction of realized institutional opportunity using the admission frontiers ultimately generated by the historical market. Objective institutional attainability can differ from applicants' subjective beliefs about admission probabilities, and such beliefs can affect behavior in centralized admissions (Kapor et al., 2020). More precise information about admission chances can itself change college choices (Ye, 2024).

We address these timing concerns through two conceptually distinct robustness exercises.

First, a *lagged-program-information specification* replaces contemporaneous program information with exact immediately preceding-year values and re-estimates the principal choice specifications. This exercise asks whether the estimated sorting patterns depend on program information measured during the current admission cycle rather than on information already available before that cycle.

Second, a *prior-year-frontier reconstruction* replaces the current-cycle program–route frontier $\kappa^{D}_{jrt}$ with the corresponding frontier observed in admission cycle $t-1$. For candidate $i$ observed in cycle $t$, qualification is therefore evaluated against the previous JoSAA cycle's program–category–pool frontier rather than against the frontier generated by the current market. This is a predetermined historical-frontier proxy, not a model of the candidate's subjective beliefs.

The implemented prior-year construction does not treat the absence of an observed previous-cycle frontier as evidence about a candidate's subjective expectations. Its purpose is to test whether the main opportunity results are robust to replacing contemporaneous equilibrium frontiers with predetermined historical admission information.

Among 109,601 candidates in the 2016–2023 universe for which a previous JoSAA cycle is available, 107,695, or 98.3 percent, have a nonempty strict prior-year-cutoff frontier. The revealed-augmented prior-year construction retains 107,685 candidates with at least two programs. The resulting opportunity gradients and the central decomposition remain close to those obtained using contemporaneous reconstructed menus, as reported in Table 4.

These exercises address different concerns. Lagged program attributes restrict the information used to describe alternatives, whereas prior-year frontiers replace the contemporaneous institutional opportunity proxy. Neither exercise requires us to claim that candidates knew or used the exact historical values in forming their choices. Together, they test whether the main results depend on contemporaneously realized program information or current-cycle admission frontiers.

### 3.8 The reconstructed IIT opportunity market

Table 1 summarizes the final reconstructed market. The 119,285 candidate menus contain 18,632,124 candidate–program alternatives. Opportunity sets are broad on average but highly dispersed. The mean candidate has 156.2 attainable academic programs, with a standard deviation of 71.8. The tenth percentile contains 51 programs, the median contains 165, and the ninetieth percentile contains 245. The mean candidate can access programs at 20.8 IITs, and the median candidate has at least one attainable academic program at all 23 IITs represented in the sample.

Coverage of reported program compensation within these menus is substantial. The average candidate has 132.7 attainable programs with observed CTC, representing 83.3 percent of the reconstructed menu. The average of candidate-level mean feasible CTC is INR 14.03 lakh per annum in constant 2023 prices. Importantly for the choice analysis, the opportunity sets contain substantial variation in compensation rather than simply many programs with similar reported outcomes: the average within-menu standard deviation of observed CTC is INR 3.97 lakh.

Geographic opportunity also varies substantially across candidates. The average distance to the closest IIT represented in the opportunity set is 154.9 kilometers, while the median is 130.1 kilometers. These statistics describe the geographic reach of institutional opportunity and should not be confused with the distance candidates ultimately choose to travel.

Panel C of Table 1 reports characteristics of the realized allocations. Among candidates whose allocated program has observed CTC, mean annual median CTC is INR 17.89 lakh in constant 2023 prices. The corresponding mean placement rate is 80.6 percent, and the mean NIRF Engineering score is 72.4. Candidates' allocated IITs are substantially farther away than the closest IIT represented in their reconstructed opportunity sets: mean distance to the allocated IIT is 750.6 kilometers, with a median of 676.7 kilometers.

These descriptive statistics illustrate why the empirical analysis distinguishes the *breadth*, *level*, and *structure* of opportunity. Candidates generally have many physical academic programs and multiple institutions within the reconstructed frontier, but the number, average outcomes, geographic distribution, and combinations of attributes represented in those menus vary substantially. Final allocations therefore cannot be understood solely from the average quality of attainable programs. They also reflect how compensation, placement outcomes, institutional quality, and geographic distance are bundled within each candidate's opportunity set and how candidates sort over those bundles.

The next section formalizes these objects. It defines opportunity breadth and opportunity level, specifies the revealed-choice model used to characterize sorting within reconstructed menus, introduces the predetermined priority measure, and develops the opportunity–structure–choice decomposition used in the remainder of the paper.

**Table 1.** Sample Construction and Reconstructed Opportunity Sets, 2015–2023

*Panel A. Sample construction*

| Sample feature | Count |
|---|---|
| Admission cycles, 2015–2023 | 9 |
| Participating IITs | 23 |
| Distinct IIT–program combinations | 353 |
| Reconstructed candidate choice sets, $K_i \geq 2$ | 119,285 |
| Reconstructed candidate–program alternatives | 18,632,124 |
| Preferred model: candidate choice sets | 119,259 |
| Preferred model: candidate–program alternatives | 18,629,219 |

*Panel B. Reconstructed opportunity sets*

| Measure | *N* | Mean | SD | P10 | Median | P90 |
|---|---|---|---|---|---|---|
| Feasible academic programs, $K_i$ | 119,285 | 156.2 | 71.8 | 51 | 165 | 245 |
| Feasible IITs | 119,285 | 20.8 | 4.1 | 15 | 23 | 23 |
| Feasible programs with observed CTC | 119,285 | 132.7 | 64.4 | 39 | 139 | 214 |
| Observed CTC coverage within feasible menu (%) | 119,285 | 83.3 | 6.9 | 74.0 | 85.2 | 89.7 |
| Mean feasible CTC (real 2023, LPA) | 119,278 | 14.03 | 1.82 | 11.66 | 13.86 | 16.56 |
| Within-menu CTC SD (real 2023, LPA) | 119,212 | 3.97 | 1.16 | 2.68 | 3.82 | 5.65 |
| Distance to closest feasible IIT (km) | 119,259 | 154.9 | 147.5 | 11.4 | 130.1 | 334.3 |

*Panel C. Chosen-program characteristics*

| Measure | *N* | Mean | SD | P10 | Median | P90 |
|---|---|---|---|---|---|---|
| Program median CTC (real 2023, LPA) | 110,747 | 17.89 | 7.10 | 10.64 | 16.35 | 28.00 |
| Placement rate (%) | 110,747 | 80.6 | 14.9 | 60.0 | 83.7 | 97.0 |
| NIRF score | 102,330 | 72.40 | 12.42 | 55.11 | 74.90 | 88.08 |
| Distance to chosen IIT (km) | 119,259 | 750.6 | 500.1 | 119.6 | 676.7 | 1,459.2 |

*Notes:* Panel A summarizes the reconstructed IIT admission market. Panels B and C report candidate-level statistics with equal candidate weights. Feasible menus contain academic programs attainable through at least one eligible priority route under the paper's reconstruction of category, seat-pool, and admission-frontier rules. CTC denotes annual median cost to company in lakh rupees per annum (LPA), deflated to 2023 prices; CTC and placement statistics use observed program-year outcomes. Distances are Haversine distances from the candidate's JEE examination-centre location to the IIT. NIRF denotes the institute's Engineering-category score. The preferred model sample excludes candidate–program observations with missing variables required for estimation.

# 4 Empirical Framework

This section develops the empirical framework used to separate differences in candidates' opportunities from differences in their choices conditional on those opportunities. Sections 2 and 3 establish the institutional and empirical objects underlying the analysis: JoSAA generates candidate-specific opportunity sets through route-specific priority and eligibility, and we reconstruct those opportunities as $F_{it}$. We take those reconstructed menus as given here and characterize them along three dimensions: their breadth, their average opportunity level, and the measured structure of the trade-offs among the alternatives they contain.

We then estimate how candidates sort among programs within their reconstructed menus and construct a common-choice benchmark that applies the same estimated behavioral rule to candidates facing different opportunity sets. This permits us to distinguish inequality in the average level of opportunity from allocation separation reproduced by differently structured menus under a common behavioral benchmark.

The analysis is deliberately organized in two stages. The first is an accounting decomposition that separates differences in average opportunity level from sorting within reconstructed opportunities. This stage does not depend on an estimated common-choice model, although it is conditional on the chosen scalar measure of opportunity level. The second stage uses an estimated common-choice operator to divide the within-menu sorting gap between a menu-structure component and residual rank-associated choice. That second attribution is necessarily benchmark dependent.

The broader distinction between preferences and heterogeneous opportunities has important antecedents (Zeng and Xie, 2008; Bonhomme et al., 2016). We operationalize that distinction in a centralized priority-rationed market by measuring the mechanism-generated opportunity set itself before applying a common behavioral rule to the alternatives it contains.

## 4.1 Candidates, programs, and reconstructed opportunities

Consider admission cycle $t$, with candidates $i \in \mathcal{I}_t$ and physical academic programs $j \in \mathcal{J}_t$. Section 2 defines the route-specific priority environment through which candidates can reach those programs, and Section 3 uses the historical JoSAA admission frontiers to construct candidate $i$'s final empirical opportunity set $F_{it}$.

The distinction between the institutional priority system and the empirical menu is important. JoSAA priority is fundamentally route specific and need not admit a single scalar representation. Construction of $F_{it}$ therefore uses the complete observed route-level priority information available for the candidate. The scalar predetermined priority measure introduced in Section 4.5 has a different purpose: it organizes candidates for empirical comparisons across the priority distribution and is not used to construct their menus.

Throughout this section, $a_{it} \in F_{it}$ denotes candidate $i$'s observed final physical academic-program allocation. We take $F_{it}$ as the institutional opportunity object delivered by the reconstruction in Section 3. The behavioral model below is conditional on that object; it does not reconstruct priority

routes or admission frontiers again.

### 4.2 Breadth, level, and measured structure of opportunities

A reconstructed opportunity set contains more information than whether a candidate can reach a particular high-outcome program. We characterize $F_{it}$ along three dimensions.

The first is *opportunity breadth*. Define

$$K_{it} = |F_{it}|, \tag{4.1}$$

the number of distinct physical academic programs in candidate $i$'s reconstructed opportunity set. Related statistics, such as the number of distinct IITs represented in $F_{it}$, measure other aspects of breadth.

The second dimension is *opportunity level*. Let $y_{jt}$ denote a program outcome of interest, such as annual median CTC, placement rate, or the NIRF Engineering score. Because outcome coverage differs across program-years, define

$$F_{it}^{y} = \left\{ j \in F_{it} : y_{jt} \text{ is observed} \right\}. \tag{4.2}$$

For candidates with $F_{it}^{y} \neq \varnothing$, define the equal-weight opportunity-level benchmark

$$B_y(F_{it}) = \frac{1}{|F_{it}^{y}|} \sum_{j \in F_{it}^{y}} y_{jt}. \tag{4.3}$$

The equal weighting in (4.3) is intentional. It summarizes the average outcome level among the candidate's attainable programs without using an estimated behavioral rule to weight those alternatives. $B_y(F_{it})$ is therefore an opportunity measure, not a predicted choice outcome and not a welfare statistic.

The equal-weight mean is also a measurement choice rather than a uniquely privileged welfare object. One could instead summarize a menu using its maximum, median, or another distributional statistic. Accordingly, the first accounting decomposition below is behavior-model free conditional on $B_y$, rather than invariant to every possible definition of opportunity level. We use the mean because it preserves the natural units of the focal outcome and admits a direct comparison with the common-choice weighted outcome defined below.

The third dimension is *measured opportunity structure*. Let $X_{ijt}$ collect the alternative characteristics represented in the preferred behavioral model as experienced by candidate $i$. These include candidate–IIT distance, program CTC and placement outcomes, the IIT's NIRF score, and the corresponding missing-information controls. We define the indexed configuration

$$\mathcal{X}_{it} = \left(X_{ijt}\right)_{j \in F_{it}} \tag{4.4}$$

as the measured structure of candidate $i$'s opportunity set.

The indexed notation in (4.4) is deliberate. Opportunity structure depends not only on which attribute vectors occur but also on the number and composition of alternatives carrying those combinations. Two distinct programs with the same observed characteristics remain two distinct alternatives in the behavioral choice denominator. Thus, measured menu structure incorporates the configuration and multiplicity of alternatives in $F_{it}$.

For outcome $y$, the corresponding outcome-supported structure is

$$\mathcal{X}_{it}^{y} = \left(X_{ijt}\right)_{j \in F_{it}^{y}}. \tag{4.5}$$

The empirical menu-structure component developed below therefore captures structure in the dimensions represented by $X_{ijt}$ and on the observed support of the focal outcome. It does not separately attribute unobserved program characteristics to menu structure.

Breadth and structure are related but conceptually distinct descriptive objects. We report $K_{it}$ directly because the number of attainable programs is economically meaningful. In the decomposition, however, breadth is not isolated as a fourth additive component. Because the common-choice operator acts on the complete indexed menu, changes in the number and composition of alternatives can affect the measured menu-structure component together with changes in their observed attribute configuration.

### 4.3 Effective choice within reconstructed institutional opportunities

Let

$$D_{ijt} = \mathbf{1}\{a_{it} = j\} \tag{4.6}$$

indicate that candidate $i$'s observed final IIT allocation in admission cycle $t$ is academic program $j$.

Our use of centralized-assignment restrictions to interpret observed choices builds on empirical work recovering preference information from priorities, reported rankings, and applicant-specific qualifying or accessible alternatives (Fack et al., 2019; Arslan, 2021). As established in Section 2.9, the institutional mechanism gives a revealed-ordering interpretation to priority-qualified alternatives appearing on the candidate's active submitted list.

Our econometric model makes one additional maintained specification: it treats the broader reconstructed institutional opportunity set $F_{it}$ as the relevant availability set for effective program choice. This does not follow mechanically from Proposition 1. A program can lie inside $F_{it}$ without having been literally submitted or actively considered by the candidate. We therefore interpret the estimated model as describing *effective sorting over reconstructed institutional opportunities*, rather than as recovering a complete, context-invariant latent preference ordering over every IIT academic program.[2]

[2]The distinction between institutional availability and behavioral consideration is related to the broader discrete-choice literature on availability and choice-set formation; see, for example, Conlon and Mortimer (2013) and Horowitz (1991).

We characterize this sorting using a conditional random-utility model (McFadden, 1972), an approach widely used in empirical college-choice applications; see, for example, Long (2004). For candidate $i$ and program $j \in F_{it}$, write latent utility as

$$U_{ijt} = \alpha_{it} + X'_{ijt}\beta + \varepsilon_{ijt}, \qquad j \in F_{it}, \tag{4.7}$$

where $\alpha_{it}$ is constant across alternatives in candidate $i$'s choice set and $X_{ijt}$ contains the alternative-specific attributes described above. Under the Type-I extreme-value specification for $\varepsilon_{ijt}$, the conditional probability of selecting program $j$ is

$$P_{ijt}(\beta) = \frac{\exp\left(X'_{ijt}\beta\right)}{\sum_{k \in F_{it}} \exp\left(X'_{ikt}\beta\right)}. \tag{4.8}$$

The candidate-specific additive term $\alpha_{it}$ cancels from the conditional probabilities.

Our preferred specification includes candidate–IIT distance, annual median CTC, placement rate, NIRF score, and the missing-information indicators described in Section 3.5. The scaling used in estimation is distance in hundreds of kilometers, CTC in units of INR 10 lakh, placement in units of 10 percentage points, and NIRF in raw score points.

We estimate the conditional-logit slope parameters using Poisson pseudo-maximum likelihood with candidate choice-set fixed effects. The corresponding conditional-mean representation is

$$\mathrm{E}\left[D_{ijt} \mid X_{ijt}, F_{it}\right] = \exp\left(\alpha_{it} + X'_{ijt}\beta\right). \tag{4.9}$$

With exactly one realized choice per candidate menu, concentrating out the candidate fixed effect yields the same slope first-order conditions and normalized conditional choice probabilities as the corresponding multinomial conditional-logit likelihood (GuimarÃ£es et al., 2003; Schmidheiny and Brülhart, 2011). The PPML representation is computationally convenient for the approximately 18.6-million-row candidate–program panel; Appendix B reports specification, support, and behavioral robustness checks.

Identification of the slope coefficients comes from variation across alternatives within the same candidate's reconstructed menu. Candidate-level characteristics and other factors constant across all alternatives in $F_{it}$ are absorbed by the candidate choice-set fixed effect. Thus, for example, the CTC coefficient is not identified from cross-candidate differences in average feasible CTC or from the fact that candidates at different points of the priority distribution face systematically different menus. It reflects how observed choice varies with CTC across alternatives available to the same candidate, conditional on the other included program characteristics.

A structural taste interpretation additionally requires the unobserved alternative-specific utility component to satisfy the maintained random-utility restrictions conditional on $X_{ijt}$ and $F_{it}$. Observed program attributes can be correlated with omitted valued characteristics such as program reputation, peer composition, employer networks, or other dimensions of program desirability. We

therefore interpret the coefficients as revealed sorting relationships conditional on the included attributes rather than as causal effects of changing those attributes.

Our preferred inference allows residual dependence both within candidate choice sets and across candidates considering the same academic program–admission-year alternative. We therefore use two-way clustered covariance estimation by candidate choice set and academic-program-by-admission-year (Cameron et al., 2011). The first dimension accommodates arbitrary within-menu dependence, while the second allows candidate observations to share common shocks associated with the same program-year.

The Type-I extreme-value specification also implies the familiar independence of irrelevant alternatives restriction. For any two feasible programs $j, k \in F_{it}$,

$$\frac{P_{ijt}(\beta)}{P_{ikt}(\beta)} = \exp\left[\left(X_{ijt} - X_{ikt}\right)' \beta\right], \tag{4.10}$$

so their relative odds do not depend on the characteristics or availability of other alternatives in $F_{it}$.

This restriction can be strong in the IIT setting. Academic programs within the same IIT share location, campus amenities, institutional reputation, and other potentially unobserved characteristics, and can therefore be closer substitutes than similarly characterized programs at different IITs. We accordingly report an IIT-nested-logit robustness exercise that permits stronger substitution among academic programs within the same institution. We use that exercise to evaluate whether the principal priority-associated trade-offs survive a richer substitution structure, rather than to choose the preferred model ex post.

### 4.4 Distance-equivalent program trade-offs

To express the estimated program-attribute coefficients on an economically interpretable common scale, we convert them into distance-equivalent trade-offs. Let $\beta_d < 0$ denote the coefficient on distance measured in hundreds of kilometers, $\beta_c$ the coefficient on annual median CTC measured in units of INR 10 lakh, $\beta_p$ the coefficient on placement measured in units of 10 percentage points, and $\beta_n$ the coefficient on the NIRF score measured in raw score points.

The distance equivalent of a INR 1 lakh difference in annual median CTC is

$$WTT^{CTC} = 10\frac{\beta_c}{-\beta_d}, \tag{4.11}$$

measured in kilometers per INR 1 lakh of annual CTC.

Because placement enters the model in units of 10 percentage points, the corresponding distance equivalent of a one-percentage-point difference in placement is

$$WTT^{Placement} = 10\frac{\beta_p}{-\beta_d}, \tag{4.12}$$

measured in kilometers per percentage point.

Because NIRF enters in raw score points, its distance equivalent is

$$WTT^{NIRF} = 100\frac{\beta_n}{-\beta_d}, \tag{4.13}$$

measured in kilometers per NIRF score point.

We refer to these quantities as willingness to travel for compactness, but their interpretation is specifically a utility-equivalent revealed-choice trade-off. Conditional on the reconstructed opportunity set and the included attributes, the estimated utility difference associated with a program attribute is equivalent to the utility difference associated with the reported distance.

Standard errors for these nonlinear transformations are calculated using the delta method applied to the preferred two-way-clustered covariance matrix.

These ratios are not causal effects of changing a program's CTC, placement rate, or NIRF score. Program attributes can proxy for other valued characteristics that are not completely observed. A robustness specification with academic-program fixed effects absorbs persistent program-specific heterogeneity and identifies the corresponding attribute coefficients from within-program changes over time, but it does not by itself convert those relationships into causal effects.

## 4.5 Predetermined priority

The institutional priority environment is multidimensional because candidates can compete through different merit lists and eligibility routes. For comparisons across the priority distribution, however, we require a predetermined scalar measure that is defined before the candidate's realized allocation.

Our preferred measure, P1, uses the candidate's own relevant merit list. Let $m_{it}^{P1}$ denote the raw priority rank used for this purpose. For OPEN candidates, $m_{it}^{P1}$ is the AIR/CRL rank. For GEN-EWS, OBC-NCL, SC, and ST candidates, it is the corresponding own-category merit-list rank.

Let $g$ denote candidate category and let $N_{gt}$ denote the number of candidates in admission cycle $t$ and category $g$ for whom $m_{it}^{P1}$ is observed. Within each admission-cycle-by-category stratum, let

$$R_{it}^{P1} = \text{midrank}_{g,t}\left(m_{it}^{P1}\right) \tag{4.14}$$

denote the empirical rank of the raw merit-list position among retained candidates in that stratum, with ties assigned their average empirical rank.

We convert this empirical midrank into the percentile

$$p_{it} = \frac{R_{it}^{P1} - \frac{1}{2}}{N_{gt}}. \tag{4.15}$$

The P1 decile is then

$$d_{it}^{P1} = \min\left\{10, \max\left[1, \left\lfloor 10p_{it} \right\rfloor + 1\right]\right\}. \tag{4.16}$$

Thus, decile 1 contains the highest-priority candidates and decile 10 the lowest-priority candi-

dates within the corresponding admission-cycle-by-category stratum.

This construction has two important advantages. First, it uses no information about the candidate's realized program, realized seat category, or realized gender pool. Second, it permits reserved-category candidates to be located along a meaningful own-list priority distribution even when AIR is unavailable.

P1 is nevertheless not a sufficient statistic for complete JoSAA priority. The institutional reconstruction in Sections 2 and 3 uses the full route-specific priority information available to the candidate. P1 is instead a scalar empirical summary used to organize comparisons across candidates. As emphasized by Lemma 1, opportunity sets need not be globally nested by this scalar measure.

To characterize priority-associated choice heterogeneity, we estimate flexible specifications that allow the coefficients on distance, CTC, placement, and NIRF to vary across P1 deciles. Differences in the resulting distance-equivalent trade-offs are descriptive priority gradients, not causal effects of obtaining a better examination rank. Merit-list position is correlated with preparation, socioeconomic background, information, geography, and other candidate characteristics that can themselves be related to program choice.

### 4.6 A common-choice benchmark

The equal-weight opportunity benchmark in (4.3) treats all outcome-observed alternatives symmetrically. To quantify the consequences of measured menu structure, we instead ask what outcomes would be predicted if candidates retained their own opportunity sets but were evaluated using the same estimated behavioral rule within an admission cycle.

For each admission cycle $t$, let $\widehat{\beta}_t^c$ denote the slope vector from the preferred full-attribute choice specification estimated using all candidates in that cycle. Let $P_t^c$ denote the resulting cycle-specific common-choice operator. For candidate $i$,

$$P_{ijt}^c \equiv P_t^c\left(j \mid F_{it}\right) = \frac{\exp\left(X_{ijt}'\widehat{\beta}_t^c\right)}{\sum_{k \in F_{it}} \exp\left(X_{ikt}'\widehat{\beta}_t^c\right)}, \qquad j \in F_{it}. \tag{4.17}$$

The common-choice operator uses the estimated slope coefficients only. Candidate choice-set fixed effects are constant across the alternatives in $F_{it}$ and therefore cancel when probabilities are normalized.

Within an admission cycle, the same slope vector is applied to candidates at every point of the P1 distribution. Their predicted probabilities can nevertheless differ because they face different programs, because those programs contain different configurations and multiplicities of observed attributes, and because candidate-specific attributes of the alternatives, most notably geographic distance, differ across candidates.

For outcome $y$, recall that $F_{it}^y$ denotes the subset of candidate $i$'s menu for which $y_{jt}$ is observed.

Conditional on predicted choice falling on this outcome-supported subset, define

$$P_{ijt}^{c,y} = \frac{P_{ijt}^{c}}{\sum_{k \in F_{it}^{y}} P_{ikt}^{c}}, \qquad j \in F_{it}^{y}, \tag{4.18}$$

provided $F_{it}^{y} \neq \varnothing$. The common-choice predicted outcome is

$$C_y(F_{it}; P_t^c) = \sum_{j \in F_{it}^{y}} P_{ijt}^{c,y} y_{jt}. \tag{4.19}$$

Equations (4.18) and (4.19) are equivalent to weighting observed values of $y$ by $P_{ijt}^{c}$ and renormalizing by the predicted probability mass on alternatives for which $y$ is observed. This is the same outcome-support convention used in the empirical implementation for CTC, placement, and NIRF.

The phrase *common choice* is deliberate. The operator does not assert that candidates literally possess identical preferences. It applies a common estimated behavioral rule as a measurement benchmark. Nor is $C_y(F_{it}; P_t^c)$ an equilibrium policy counterfactual. Candidate menus, other candidates' allocations, program capacities, and admission frontiers are held fixed.

Finally, the operator captures opportunity structure only through the alternative characteristics represented in the preferred behavioral model and their configuration within the reconstructed menu. Differences in latent program attributes omitted from $X_{ijt}$ are not separately identified as menu structure.

## 4.7 Opportunity–structure–choice decomposition

We now combine the preceding objects to separate the high–low priority allocation gap into differences in average opportunity level, additional separation reproduced by differently structured menus under a common-choice benchmark, and residual rank-associated choice.

Let

$$H = \left\{(i,t) : d_{it}^{p1} = 1\right\}$$

denote candidate–cycle observations in the highest predetermined-priority decile and

$$L = \left\{(i,t) : d_{it}^{p1} = 10\right\}$$

the corresponding lowest-priority group.

The decomposition is outcome specific because CTC, placement, and NIRF have different observed support. For outcome $y$, define the common candidate sample

$$\mathcal{S}_y = \left\{(i,t) : Y_{it}^{y} \text{ is observed},\ F_{it}^{y} \neq \varnothing,\ C_y(F_{it}; P_t^c) \text{ is defined}\right\}, \tag{4.20}$$

where $Y_{it}^{y}$ is defined below. Let

$$H_y = H \cap \mathcal{S}_y, \qquad L_y = L \cap \mathcal{S}_y. \tag{4.21}$$

Every component of the decomposition for outcome $y$ is evaluated on these same high- and low-priority candidate samples. This prevents differences among components from arising mechanically because of changing candidate composition across the observed allocation, feasible-menu, and common-choice objects.

In the empirical implementation, the conditional expectations below denote sample means over the indicated retained candidate–cycle observations. Candidate–cycle observations are pooled across admission years with equal candidate weight; we do not first average within admission cycles or categories and then weight those strata equally.

For any candidate-level scalar $Z_{it}$, define the outcome-specific high–low contrast

$$\Delta_y Z = \mathrm{E}\big[Z_{it} \mid (i,t) \in H_y\big] - \mathrm{E}\big[Z_{it} \mid (i,t) \in L_y\big]. \tag{4.22}$$

For program outcome $y$, define the observed allocated-program outcome as

$$Y_{it}^{y} = y_{a_{it}t}. \tag{4.23}$$

The observed high–low allocation gap is

$$\Delta_Y^{y} = \mathrm{E}\big[Y_{it}^{y} \mid (i,t) \in H_y\big] - \mathrm{E}\big[Y_{it}^{y} \mid (i,t) \in L_y\big]. \tag{4.24}$$

The *opportunity-level component* is

$$\Delta_O^{y} = \mathrm{E}\big[B_y(F_{it}) \mid (i,t) \in H_y\big] - \mathrm{E}\big[B_y(F_{it}) \mid (i,t) \in L_y\big]. \tag{4.25}$$

Define the remaining difference as the *within-menu sorting gap*,

$$\Delta_W^{y} = \Delta_Y^{y} - \Delta_O^{y}. \tag{4.26}$$

By construction,

$$\Delta_Y^{y} = \Delta_O^{y} + \Delta_W^{y}. \tag{4.27}$$

This first decomposition is an exact sample accounting identity conditional on the reconstructed opportunity sets, the common outcome-support sample, and the equal-weight opportunity-level measure $B_y$. It does not depend on an estimated common-choice operator. It asks how much of the observed high–low allocation gap is already present in the average level of attainable opportunities and how much remains because observed allocations differ from those equal-weight menu benchmarks.

We next introduce the model-dependent attribution of the fixed within-menu sorting gap. Let

$$P^c = \left\{P_t^c\right\}_t \tag{4.28}$$

denote the collection of admission-cycle-specific common-choice operators. For outcome $y$, the common-choice high–low outcome gap is

$$\Delta_C^y(P^c) = \mathrm{E}\left[C_y(F_{it}; P_t^c) \mid (i,t) \in H_y\right] - \mathrm{E}\left[C_y(F_{it}; P_t^c) \mid (i,t) \in L_y\right]. \tag{4.29}$$

We define the *menu-structure component* as

$$\Delta_S^y(P^c) = \Delta_C^y(P^c) - \Delta_O^y. \tag{4.30}$$

This quantity measures the additional high–low separation reproduced when the common-choice operator is applied to candidates' differently configured menus relative to the equal-weight opportunity benchmark. It therefore attributes to measured menu structure the difference between what the common behavioral rule predicts from the observed configuration and multiplicity of alternatives and what would arise from assigning equal weight to those outcome-supported opportunities.

The structure component is not a pure covariance statistic and does not separately identify breadth. Because the common-choice rule operates on the indexed collection of alternatives, both the composition and multiplicity of observed program configurations can contribute to $\Delta_S^y(P^c)$.

Finally, define the *residual rank-associated choice component* as

$$\Delta_R^y(P^c) = \Delta_Y^y - \Delta_C^y(P^c). \tag{4.31}$$

The residual measures the difference between the observed high–low allocation gap and the gap reproduced when the common behavioral rule is applied to candidates' reconstructed menus. It should not be interpreted as pure preference heterogeneity. It can reflect systematic differences in effective preferences, information, omitted program attributes, candidate backgrounds, alternative decision rules, or misspecification of the behavioral model.

*Remark* 1 (Decomposition identity and benchmark dependence). For any program outcome $y$, comparison groups $H_y$ and $L_y$ defined on a common outcome-specific sample for which the relevant means exist, and any admissible collection of common-choice operators $P^c$,

$$\Delta_Y^y = \Delta_O^y + \Delta_S^y(P^c) + \Delta_R^y(P^c). \tag{4.32}$$

Equivalently,

$$\Delta_W^y = \Delta_S^y(P^c) + \Delta_R^y(P^c). \tag{4.33}$$

The observed allocation gap $\Delta_Y^y$, the opportunity-level component $\Delta_O^y$, and the within-menu sorting gap $\Delta_W^y$ are invariant to the choice of common-choice operator. Only the attribution of the

fixed within-menu gap between $\Delta_S^y(P^c)$ and $\Delta_R^y(P^c)$ depends on that benchmark.

More precisely, for any two admissible common-choice collections $P^c$ and $\widetilde{P}^c$,

$$\Delta_S^y(P^c) - \Delta_S^y(\widetilde{P}^c) = \Delta_C^y(P^c) - \Delta_C^y(\widetilde{P}^c), \tag{4.34}$$

and

$$\Delta_R^y(P^c) - \Delta_R^y(\widetilde{P}^c) = -\left[\Delta_S^y(P^c) - \Delta_S^y(\widetilde{P}^c)\right]. \tag{4.35}$$

A proof is provided in Appendix A.4.

the decomposition identity in Remark 1 makes explicit the hierarchy of the empirical exercise. The first split,

$$\Delta_Y^y = \Delta_O^y + \Delta_W^y,$$

is determined by observed allocations, reconstructed opportunity sets, the chosen opportunity-level functional $B_y$, and the focal outcome's observed support. It does not depend on a common-choice model.

The subsequent split,

$$\Delta_W^y = \Delta_S^y(P^c) + \Delta_R^y(P^c),$$

answers a different question: how the fixed within-menu sorting gap is attributed under a specified common behavioral benchmark. Changing $P^c$ can therefore move variation between menu structure and residual choice but cannot alter the observed allocation gap, the opportunity-level component, or the within-menu sorting gap. The adjustment is one-for-one.

the decomposition identity in Remark 1 is algebraically an exact accounting identity. Its purpose is not to claim a new matching-theory result, but to make precise which empirical objects are defined without a behavioral benchmark and exactly where model dependence enters the opportunity–structure–choice measurement framework. We therefore place greater interpretive weight on the first split than on any single numerical attribution of the within-menu gap. The common-choice operator provides a disciplined benchmark for measuring how much separation differently configured menus can reproduce; it does not turn the resulting components into invariant structural parameters.

*Remark* 2 (Common-choice reproduction shares). Under the conditions of the decomposition identity in Remark 1, suppose $\Delta_Y^y \neq 0$. The share of the observed high–low allocation gap reproduced when the common-choice collection $P^c$ is applied to candidates' own reconstructed opportunity sets is

$$\mathcal{A}_y(P^c) = \frac{\Delta_C^y(P^c)}{\Delta_Y^y} = \frac{\Delta_O^y + \Delta_S^y(P^c)}{\Delta_Y^y} = 1 - \frac{\Delta_R^y(P^c)}{\Delta_Y^y}. \tag{4.36}$$

If, in addition, $\Delta_W^y \neq 0$, the share of the within-menu sorting gap attributed to measured menu

structure under the same common-choice benchmark is

$$\mathcal{M}_y(P^c) = \frac{\Delta_S^y(P^c)}{\Delta_W^y} = 1 - \frac{\Delta_R^y(P^c)}{\Delta_W^y}. \tag{4.37}$$

The two shares answer different questions. $\mathcal{A}_y(P^c)$ measures how much of the *total observed allocation gap* is reproduced when candidates retain their own reconstructed opportunity sets but are evaluated using common admission-cycle-specific behavioral rules. It therefore includes both the high–low difference in average opportunity level and the additional difference attributed to measured menu structure under those rules.

By contrast, $\mathcal{M}_y(P^c)$ conditions on the portion of the observed allocation gap remaining after differences in average opportunity level have been removed. It measures the share of that *within-menu sorting gap* attributed to measured menu structure under the common-choice benchmark.

Both $\mathcal{A}_y(P^c)$ and $\mathcal{M}_y(P^c)$ inherit the benchmark dependence of $\Delta_S^y(P^c)$ and $\Delta_R^y(P^c)$. They are therefore measurement statistics under a specified common behavioral rule rather than causal or welfare parameters. Because they are signed ratios of empirical contrasts, they are not mechanically restricted to the unit interval in every possible application, although they lie between zero and one for the principal decompositions reported below.

### 4.8 Uncertainty and interpretation

The opportunity-level and within-menu components are sample accounting objects conditional on the reconstructed menus, the focal outcome's observed support, and the P1 comparison groups. The common-choice components additionally depend on estimated behavioral parameters.

We quantify uncertainty in that model-dependent attribution using parametric simulation. For admission cycle $t$ and simulation draw $b = 1, \ldots, 2000$, we draw

$$\beta_t^{(b)} \sim N\left(\widehat{\beta}_t^c, \widehat{V}_t\right), \tag{4.38}$$

where $\widehat{V}_t$ is the preferred two-way-clustered covariance matrix for the year-specific common-choice model. Because the year-specific models are estimated on disjoint admission-cycle samples, the coefficient vectors are drawn independently across admission cycles within each simulation draw.

For each draw, we recompute the common-choice probabilities and therefore $\Delta_C^y(P^c)$, $\Delta_S^y(P^c)$, $\Delta_R^y(P^c)$, and their benchmark-dependent shares. The observed candidate sample, P1 grouping, reconstructed opportunity sets, outcome-support convention, $\Delta_Y^y$, $\Delta_O^y$, and $\Delta_W^y$ remain fixed.

The resulting intervals are therefore *parametric simulation intervals reflecting estimation uncertainty in the common-choice coefficients conditional on the observed sample and reconstructed opportunity sets*. They are not comprehensive confidence intervals incorporating uncertainty in menu reconstruction, the fixed accounting components, or every earlier stage of data construction.

Finally, none of the objects in (4.27)–(4.32) should be interpreted as welfare. CTC, placement, and

NIRF are characteristics of allocated academic programs rather than cardinal utility measures. The choice model supplies revealed trade-offs among observed attributes conditional on participation and the maintained choice-set specification, but the data do not contain the full outside-option set required for a complete welfare comparison.

The decomposition instead has a different purpose: to characterize the anatomy of allocation inequality within the observed IIT market by distinguishing differences in average attainable outcomes, separation reproduced by the measured configuration of opportunities under a common behavioral rule, and remaining rank-associated differences in observed choice.

This separation between opportunity and behavior organizes the empirical results that follow.

**Table 2.** Program Attributes and Revealed College Choice

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Distance only | + CTC | Full attributes | + Program FE |
| *Panel A. Utility coefficients* | | | | |
| Distance (100 km) | −0.126*** | −0.134*** | −0.133*** | −0.135*** |
| | (0.003) | (0.003) | (0.003) | (0.002) |
| Median CTC (Rs. 10 lakh, 2023 prices) | | 1.499*** | 1.303*** | 0.498*** |
| | | (0.058) | (0.053) | (0.048) |
| Placement rate (10 pp) | | | 0.261*** | 0.010 |
| | | | (0.019) | (0.010) |
| NIRF score | | | 0.022*** | −0.002 |
| | | | (0.002) | (0.003) |
| *Panel B. Implied willingness to travel* | | | | |
| CTC (km per 1 lakh rupees) | | 111.7 | 97.7 | 36.8 |
| | | [102.0, 121.4] | [88.7, 106.7] | [29.7, 43.8] |
| Placement (km per 1 percentage point) | | | 19.6 | 0.8 |
| | | | [16.6, 22.6] | [−0.8, 2.3] |
| NIRF (km per 1 score point) | | | 16.2 | −1.6 |
| | | | [12.5, 19.9] | [−6.0, 2.9] |
| *Panel C. Specification and sample* | | | | |
| No-graduated-batch indicator | No | Yes | Yes | Yes |
| NIRF-missing indicator | No | No | Yes | Yes |
| Candidate choice-set fixed effects | Yes | Yes | Yes | Yes |
| Academic-program fixed effects | No | No | No | Yes |
| Two-way clustered standard errors | Yes | Yes | Yes | Yes |
| Candidate–program alternatives | 18,629,219 | 18,629,219 | 18,629,219 | 18,629,219 |
| Candidate choice sets | 119,259 | 119,259 | 119,259 | 119,259 |
| Program–year clusters | 2,291 | 2,291 | 2,291 | 2,291 |

*Notes:* All columns estimate conditional-choice models by Poisson pseudo-maximum likelihood with candidate choice-set fixed effects on the same final analysis sample. The dependent variable equals one for the allocated academic program within each candidate's reconstructed feasible menu. Distance is measured in 100 km. CTC is annual program median compensation measured in units of Rs. 10 lakh and expressed in 2023 prices. Placement is measured in 10-percentage-point units. NIRF is the institute's Engineering-category score. The no-graduated-batch indicator is included from column (2), and the NIRF-missing indicator from column (3). Column (4) additionally includes academic-program fixed effects. Standard errors in parentheses are two-way clustered by candidate choice set and academic-program-by-admission-year. Panel B reports willingness-to-travel estimates obtained as coefficient ratios relative to distance; brackets contain 95% delta-method confidence intervals. These trade-offs summarize revealed choice associations conditional on the included program attributes and should not be interpreted as causal effects of program characteristics. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

# 5 Priority, Opportunity Structure, and Allocation

Section 4.3 establishes how candidates sort, on average, over compensation, geography, placement outcomes, and measured institutional quality within their reconstructed IIT opportunity sets. We

now turn to differences across the predetermined priority distribution. The central question is whether the large differences in final program allocations between higher- and lower-priority candidates can be understood primarily through differences in the average level of the opportunities they can attain, or whether the configuration of those opportunities also plays an important role.

We proceed in four steps. First, we document how opportunity breadth and opportunity level vary across predetermined priority. Second, we examine whether revealed CTC–distance trade-offs differ across the priority distribution. Third, we compare observed allocations with the equal-weight opportunity benchmark and with the predictions generated by a common estimated choice rule applied to candidates' own menus. Finally, we quantify these differences using the opportunity–structure–choice decomposition developed in Section 4.7.

Throughout this section, priority is measured using the predetermined P1 measure defined in Section 4.5. Decile 1 contains the highest-priority candidates and decile 10 the lowest-priority candidates within the corresponding admission-year-by-category stratum. Unless otherwise indicated, the figures and high–low contrasts pool candidate–cycle observations across admission years and candidate categories after P1 has been constructed within those strata, so each retained candidate–cycle observation receives equal weight. The resulting gradients therefore compare candidates at different positions in their predetermined within-stratum priority distributions rather than on a single national scalar priority scale.

These comparisons are descriptive rather than causal. P1 is predetermined with respect to the realized JoSAA allocation, but examination rank is not randomly assigned and is correlated with candidate characteristics that can themselves shape educational opportunities and program choice.

### 5.1 Priority, opportunity breadth, and opportunity level

Lemma 1 establishes that, conditional on a common vector of admission frontiers, complete routewise priority dominance weakly expands the cutoff-qualified institutional opportunity set. P1 does not impose that dominance relation: it is a scalar summary of a multidimensional priority system in which candidates can differ in merit-list membership, reservation eligibility, gender-pool eligibility, and other route-specific qualifications. Figure 1 therefore asks whether a corresponding opportunity gradient is visible empirically when candidates are ordered by the predetermined P1 measure.

Panel A shows a pronounced gradient in opportunity breadth. Candidates in the highest-priority P1 decile have, on average, approximately 250 attainable IIT academic programs. The corresponding mean declines steadily across the priority distribution and falls below 50 programs in the lowest-priority decile. The P1 gradient is therefore not a tautological nesting result: the scalar measure used to organize candidates is distinct from the complete route-level information used to construct $F_{it}$. It nevertheless reveals a strong empirical relationship between candidates' predetermined positions in the priority system and the breadth of their reconstructed IIT opportunities.

Panel B shows that the gradient extends beyond menu size. Mean feasible CTC also declines across P1 deciles. Candidates in the highest-priority decile face opportunity sets with average

annual median CTC of approximately INR 16.7 lakh, whereas the corresponding mean is roughly INR 12.4 lakh for candidates in the lowest-priority decile. Higher-priority candidates therefore face opportunity sets that are not only substantially broader but also contain programs with higher reported compensation on average.

The distinction between breadth and opportunity level is economically important. Adding programs whose characteristics closely resemble those already attainable can substantially increase menu size without materially changing average feasible outcomes or the trade-offs available to the candidate. Conversely, adding a smaller number of programs with different combinations of compensation, distance, placement outcomes, and the NIRF-based institute-quality measure can alter the decision environment considerably.

For this reason, neither menu size nor average feasible CTC is a sufficient description of opportunity. The analysis below asks how the observed configuration of attributes within these heterogeneous menus maps into allocations when candidates are evaluated using the same behavioral rule.

### 5.2 Priority-associated heterogeneity in revealed choice

Differences in reconstructed opportunities need not be the only source of allocation differences across the priority distribution. Candidates at different positions in that distribution can also sort differently among the alternatives available to them.

To examine this margin, we allow the principal program-attribute coefficients in the conditional-choice model to vary flexibly across P1 deciles. Figure 2 reports the resulting CTC–distance trade-off, expressed in kilometers per INR 1 lakh of annual median CTC using equation (4.11).

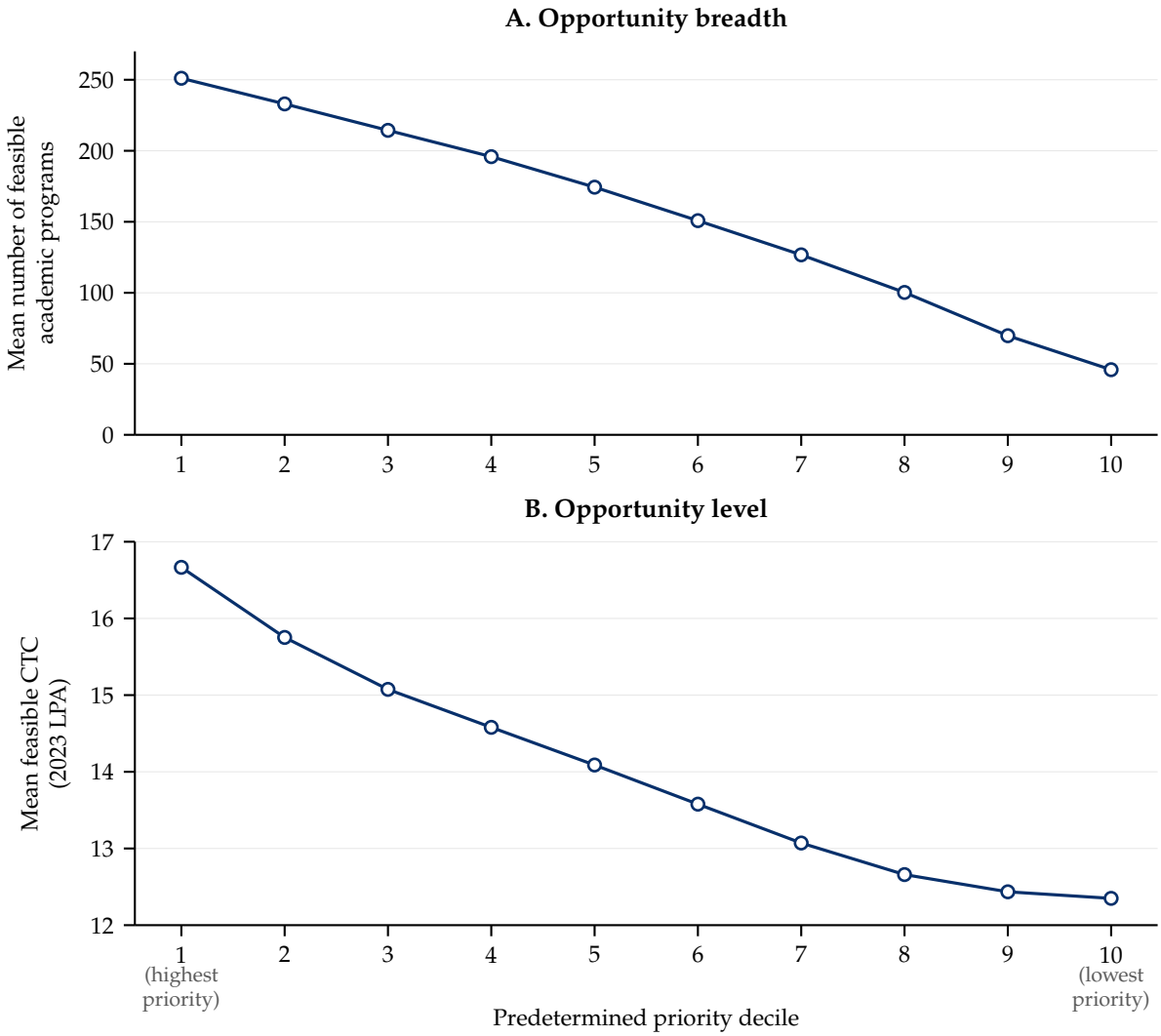
The point estimates decline across the priority distribution. Candidates in P1 decile 1 have an estimated CTC–distance trade-off of 103.64 kilometers per INR 1 lakh, compared with 78.67 kilometers in decile 10. The estimated top–bottom difference is therefore 24.97 kilometers per INR 1 lakh, with a 95-percent confidence interval of $[-4.22,\ 54.16]$ and a two-sided $p$-value of 0.094.

The point estimates therefore suggest economically meaningful priority-associated heterogeneity in revealed CTC–distance trade-offs, but the endpoint difference is not precisely estimated at conventional levels. This pattern is supporting evidence rather than the central empirical result of the paper.

The interpretation is also deliberately noncausal. P1 is predetermined with respect to the realized allocation but is correlated with academic preparation, family background, information, geography, and other characteristics that can themselves be related to college-program choice. Figure 2 therefore describes *priority-associated choice heterogeneity*; it does not identify a causal effect of obtaining stronger priority on willingness to trade distance for compensation.

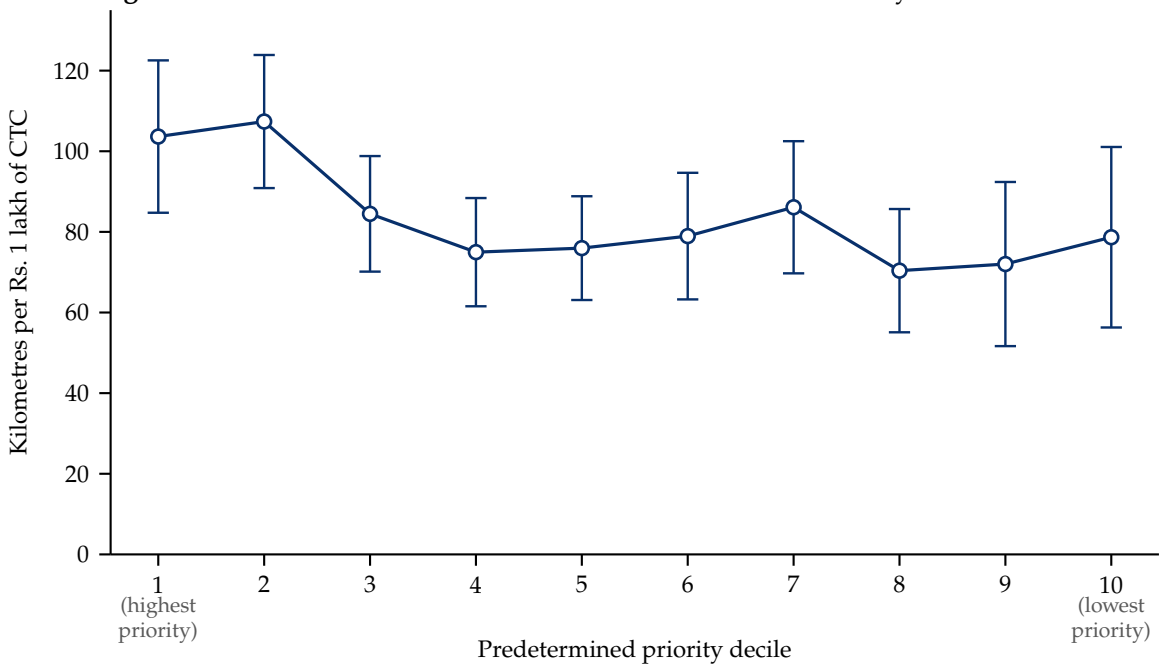
The decomposition below does not require such a causal interpretation. Its main exercise instead asks how much of the observed allocation gradient can be reproduced when candidates at different points of the priority distribution retain their own reconstructed opportunity sets but are evaluated using the same estimated behavioral rule.

**Figure 1.** Predetermined Priority and Reconstructed Opportunities




*Notes:* Priority is measured using the predetermined own-list measure (P1). For OPEN candidates, priority is based on AIR/CRL rank; for EWS, OBC-NCL, SC, and ST candidates, it is based on the corresponding own-category merit-list rank. Deciles are constructed within admission year and candidate category, with decile 1 denoting the highest priority and decile 10 the lowest. Feasible programs are reconstructed from candidates' eligibility, priority, and observed admission frontiers. Panel A reports the candidate-weighted mean number of feasible academic programs. Panel B reports the candidate-weighted mean CTC across programs in each candidate's reconstructed feasible menu. CTC is annual program median compensation in lakh rupees per annum, expressed in 2023 prices. The figure is descriptive and does not interpret differences across priority deciles causally.

**Figure 2.** CTC–Distance Trade-offs across the Predetermined Priority Distribution




*Notes:* The figure reports CTC-related willingness to travel from the flexible P1-decile version of the preferred choice model. Priority is measured using the predetermined own-list measure (P1). For OPEN candidates, P1 is based on AIR/CRL rank; for EWS, OBC-NCL, SC, and ST candidates, it is based on the corresponding own-category merit-list rank. Deciles are constructed within admission year and candidate category, with decile 1 denoting the highest priority and decile 10 the lowest. For each decile, the CTC and distance coefficients are allowed to vary, and willingness to travel is calculated as the utility-equivalent distance in kilometres associated with Rs. 1 lakh of real annual median CTC. Vertical bars show 95% confidence intervals from two-way clustered covariance. The top-minus-bottom-decile difference is 24.97 km per Rs. 1 lakh of CTC (95% CI: [−4.22, 54.16]; $p = 0.094$). The pattern is interpreted as priority-associated heterogeneity in revealed choice behaviour, not as a causal effect of priority.

### 5.3 Observed allocations and the common-choice benchmark

Figure 3 brings together the three objects that motivate the CTC decomposition. The figure is calculated on the common CTC candidate sample $\mathcal{S}_{CTC}$ defined in Section 4.7, so the observed allocation, feasible-menu benchmark, and common-choice prediction are compared using the same candidate–cycle observations within each P1 decile.

The first series is mean *observed allocated-program CTC*: the average CTC of the physical academic programs actually allocated to candidates in each P1 decile. This series exhibits a steep priority gradient. On the common outcome-supported sample, the difference between the highest- and lowest-priority deciles is

$$\Delta_Y^{CTC} = 17.5023 \quad \text{lakh rupees per annum.}$$

The second series is the equal-weight opportunity benchmark, $B_{CTC}(F_{it})$. Its priority gradient is substantially flatter. The corresponding high–low difference is

$$\Delta_O^{CTC} = 4.2677 \quad \text{lakh.}$$

Thus, only part of the observed allocation separation is already present in the average compensation level of the programs lying inside candidates' reconstructed opportunity sets.

The third series is common-choice predicted CTC, $C_{CTC}(F_{it}; P_t^c)$. Candidates retain their own reconstructed menus, including their candidate-specific distances and the actual configuration and multiplicity of observed program attributes they face, but all candidates in the same admission cycle are evaluated using the same estimated slope parameters.

The common-choice prediction lies much closer to the observed allocation series than does the equal-weight feasible-menu benchmark. Its high–low gap is

$$\Delta_C^{CTC}(P^c) = 13.3331 \quad \text{lakh.}$$

Relative to the observed allocation gap, the common-choice exercise therefore reproduces

$$\mathcal{A}_{CTC}(P^c) = \frac{13.3331}{17.5023} = 0.762 \tag{5.1}$$

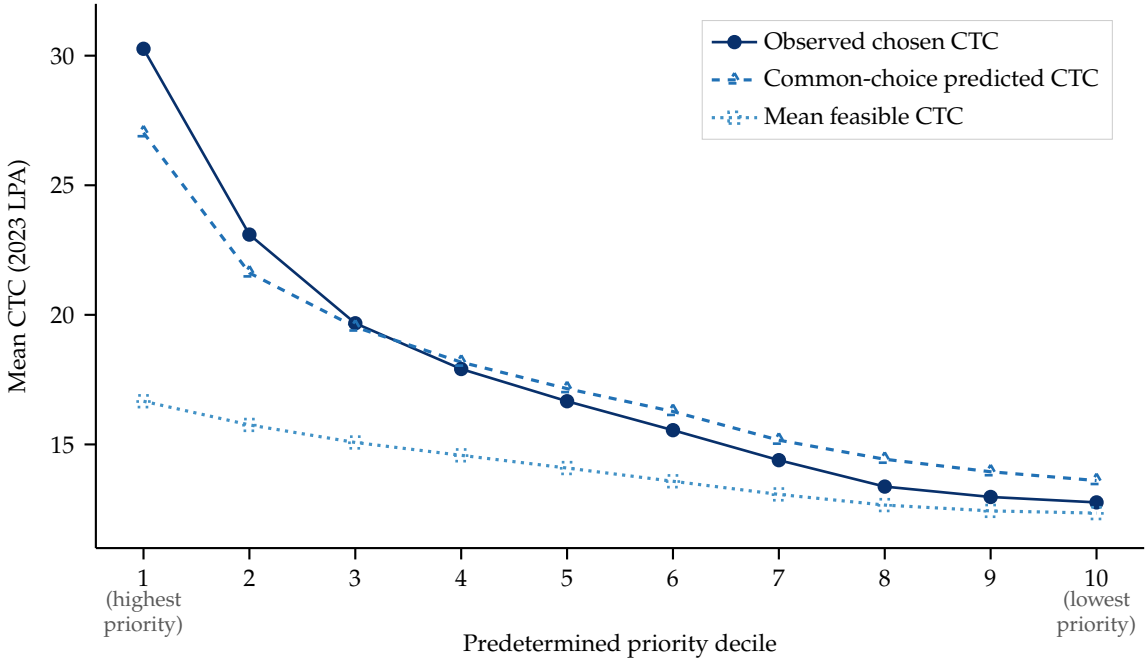
or approximately 76.2 percent of the observed high–low allocated-program CTC gap.

This result provides an intuitive summary of the exercise. Candidates at the top and bottom of the priority distribution are being evaluated using the same year-specific behavioral rule. The predicted allocation gradient nevertheless remains large because those candidates face different opportunity sets: different programs are available, candidate-specific distances differ, and compensation, placement, the NIRF score, and the other measured dimensions of the preferred choice model are configured differently across their menus.

The common-choice comparison does not establish that these menu differences caused the historical allocation gap. It measures how much separation is *reproduced* when a common estimated

behavioral operator is applied to the observed institutional opportunity sets.

**Figure 3.** Observed, Feasible, and Common-Choice CTC across the Predetermined Priority Distribution




*Notes:* The figure reports candidate-level CTC outcomes by predetermined priority decile. Priority is measured using the predetermined own-list measure (P1). For OPEN candidates, P1 is based on AIR/CRL rank; for EWS, OBC-NCL, SC, and ST candidates, it is based on the corresponding own-category merit-list rank. Deciles are constructed within admission year and candidate category, with decile 1 denoting the highest priority and decile 10 the lowest. "Observed chosen CTC" is the mean CTC of candidates' assigned programs. "Mean feasible CTC" is the mean CTC across programs in each candidate's reconstructed feasible menu. "Common-choice predicted CTC" applies the same year-specific pooled estimated choice rule to each candidate's own reconstructed menu and averages the resulting predicted CTC. CTC is annual program median compensation in lakh rupees per annum, expressed in 2023 prices. The common-choice series is model-based and is not interpreted as a causal counterfactual.

## 5.4 Decomposing the allocated-program CTC gap

Table 3 applies the framework in Section 4.7 to annual median CTC.

The first result is the behavior-model-free accounting split, conditional on the reconstructed opportunity sets, common outcome support, and the equal-weight opportunity benchmark $B_{CTC}$. The observed high–low allocated-program CTC gap satisfies

$$17.5023 = 4.2677 + 13.2346, \tag{5.2}$$

where INR 4.2677 lakh is the difference in average feasible CTC and INR 13.2346 lakh is the within-menu sorting gap.

Expressed relative to the observed allocation gap,

$$\frac{4.2677}{17.5023} = 24.4\%,$$

so differences in average feasible CTC account for only 24.4 percent of the high–low allocation gap. The remaining 75.6% is a within-menu sorting gap.

This is the least model-dependent conclusion of the decomposition. Candidates with stronger predetermined priority do face substantially higher-CTC opportunities on average, but inequality in average opportunity level accounts for only about one-quarter of the observed allocated-program CTC gap. Roughly three-quarters of the gap remains after that difference has been removed.

We next use the preferred year-specific common-choice operator to characterize this fixed within-menu gap. The common-choice prediction implies 13.2346 = 9.0654 + 4.1692, where INR 9.0654 lakh is the menu-structure component and INR 4.1692 lakh is residual rank-associated choice. Combining the two stages gives 17.5023 = 4.2677 + 9.0654 + 4.1692. As shares of the observed high–low allocated-program CTC gap, these components are 24.4 percent opportunity level, 51.8 percent measured menu structure, and 23.8 percent residual rank-associated choice.

The first two components together reproduce 76.2 percent of the total observed gap under common choice. Of this, 24.4 percentage points are already present in average feasible CTC, while an additional 51.8 percentage points are reproduced by the configuration and multiplicity of observed alternatives when the common behavioral rule is applied.

A second useful normalization conditions on the fixed within-menu sorting gap. Under the preferred operator,

$$\mathcal{M}_{CTC}(P^c) = \frac{9.0654}{13.2346} = 0.685, \tag{5.3}$$

so 68.5 percent of the observed within-menu CTC gap is attributed to measured menu structure under the preferred common-choice benchmark. The remaining 31.5 percent of the within-menu gap is residual rank-associated choice.

The hierarchy between these findings is important. The conclusion that only 24.4 percent of the observed allocation gap is present in average feasible CTC does not depend on the common-choice model. The subsequent statement that 68.5 percent of the remaining within-menu gap is attributed to measured menu structure is benchmark dependent. It answers what happens when the preferred common estimated behavioral rule is applied to candidates' actual opportunity sets.

The residual component should not be interpreted as pure preference heterogeneity. It is the difference between observed allocations and the outcomes reproduced under the specified common rule. It can therefore contain systematic differences in effective preferences, information, candidate backgrounds, omitted program attributes, alternative decision rules, or misspecification of the behavioral model.

Panel C of Table 3 quantifies parameter uncertainty in this second-stage attribution. The 95-percent parametric simulation interval for the menu-structure component is [8.30, 9.86] lakh. Expressed as a share of the total allocation gap, the corresponding interval is [47.4%, 56.3%]. The residual component has a 95-percent parametric interval of [3.38, 4.93] lakh. Most directly, the menu-structure share of the within-menu gap is 68.5 percent, with a 95-percent parametric simulation interval of [62.7%, 74.5%].

Conditional on the observed sample, reconstructed menus, outcome-support convention, and preferred common-choice specification, uncertainty in the estimated common-choice coefficients therefore does not alter the conclusion that the menu-structure component constitutes a majority of the within-menu CTC gap. These intervals propagate uncertainty in the estimated behavioral parameters only; they are not comprehensive confidence intervals for every stage of menu reconstruction or for the accounting components themselves.

**Table 3.** Decomposing the High–Low Priority Gap in Chosen-Program CTC

| | CTC component (2023 LPA) | Share of total gap | Share of within-menu gap |
|---|---|---|---|
| *Panel A. Accounting decomposition* | | | |
| Total chosen-program CTC gap | 17.50 | 100.0% | |
| Opportunity-level component | 4.27 | 24.4% | |
| Within-menu sorting | 13.23 | 75.6% | 100.0% |
| *Panel B. Decomposition of within-menu sorting under common choice* | | | |
| Menu structure under common choice | 9.07 | 51.8% | 68.5% |
| Residual rank-associated choice | 4.17 | 23.8% | 31.5% |

*Panel C. Parametric uncertainty for model-dependent components*

| Measure | Estimate | 95% parametric simulation interval |
|---|---|---|
| Menu-structure component | 9.07 | [8.30, 9.86] |
| Residual rank-associated choice | 4.17 | [3.38, 4.93] |
| Menu-structure share of total gap | 51.8% | [47.4%, 56.3%] |
| Residual share of total gap | 23.8% | [19.3%, 28.2%] |
| Menu-structure share of within-menu gap | 68.5% | [62.7%, 74.5%] |

*Panel D. Analysis sample*

| Priority group | Candidates |
|---|---|
| Top predetermined-priority decile | 11,806 |
| Bottom predetermined-priority decile | 9,788 |

*Notes:* The table decomposes the difference in chosen-program CTC between candidates in the highest and lowest deciles of the predetermined own-list priority measure (P1). For OPEN candidates, P1 is based on AIR/CRL rank; for EWS, OBC-NCL, SC, and ST candidates, it is based on the corresponding own-category merit-list rank. Priority deciles are constructed within admission year and candidate category.

Panel A reports the accounting decomposition $\Delta_Y = \Delta_O + \Delta_W$, where $\Delta_Y$ is the observed top-minus-bottom chosen-CTC gap, $\Delta_O$ is the difference in mean feasible-menu CTC, and $\Delta_W$ is the remaining within-menu sorting gap. Panel B further decomposes the within-menu gap as $\Delta_W = \Delta_S(P^c) + \Delta_R(P^c)$, where $\Delta_S(P^c)$ is the menu-structure component obtained by applying the same year-specific pooled estimated choice rule to candidates' own reconstructed feasible menus, and $\Delta_R(P^c)$ is residual rank-associated choice. The latter should not be interpreted as identified pure taste heterogeneity.

Panel C reports 95% parametric simulation intervals from 2,000 draws of the year-specific common-choice coefficients using covariance matrices two-way clustered by candidate choice set and academic-program-by-admission-year. The observed sample, reconstructed menus, total chosen-CTC gap, opportunity-level component, and within-menu gap are held fixed. Accordingly, these intervals reflect estimation uncertainty in the common-choice parameters conditional on the observed sample and reconstructed menus; they are not comprehensive confidence intervals for the accounting components.

CTC is annual program median compensation in lakh rupees per annum (LPA), expressed in 2023 prices. The decomposition is accounting- and model-based and is not interpreted causally.

### 5.5 Alternative outcomes and robustness

The decomposition is not mechanically specific to CTC. Appendix Table A11 applies the same P1 grouping and common-choice framework to placement rates and NIRF scores.

For placement, 46.1 percent of the observed high–low allocation gap is associated with average opportunity level, 31.7 percent is attributed to measured menu structure under common choice, and 22.2 percent remains as residual rank-associated choice. For NIRF, the corresponding shares are 37.9, 20.5, and 41.7 percent. The decomposition therefore differs materially across outcomes: average feasible opportunity is relatively more important for placement, whereas residual rank-associated choice is relatively more important for the NIRF score. The menu-structure component remains positive and economically meaningful for both outcomes.

Table 4 and the corresponding appendix exercises examine five additional dimensions of the CTC result.

First, the decomposition is stable to the information vintage used for program attributes (Table A14, panel A). On the common 2016–2023 period, the contemporaneous-information specification yields shares of approximately 24.4%, 48.6%, 27.0%, for opportunity level, menu structure, and residual choice. Replacing current program information with exact immediately preceding-year information produces 25.0%, 52.8%, 22.2%, respectively. The corresponding menu-structure share of the within-menu gap is approximately 70.4 percent. Table A16 reports coverage of exact prior-year program information, which for CTC is 78.0 percent of feasible alternatives and 89.1 percent of chosen programs.

Second, we replace contemporaneous admission frontiers with prior-year frontiers (Table A14, panel B). This exercise changes the historical frontier used to reconstruct attainable programs; unless otherwise indicated, target-year candidate geography and program characteristics retain their baseline measurement. Under the strict prior-year-frontier construction, the decomposition shares are 23.3%, 50.8%, 25.8%, while the revealed-augmented version gives 23.9%, 50.9%, 25.2%.

The close correspondence shows that the central decomposition is not highly sensitive to defining institutional opportunity using contemporaneous rather than predetermined historical closing ranks. We do not interpret the prior-year frontier as candidates' literal subjective opportunity set; it is a predetermined historical proxy for the admission frontier. Table A17 reports the frontier construction, sample retention, and menu sizes for both ex-ante designs. The strict construction retains 60.5 percent of the 2016–2023 candidate universe, because a candidate's chosen program must itself be predicted feasible from prior-year cutoffs, while the revealed-augmented design retains 98.3 percent.

Third, we vary the empirical definition of priority while holding the preferred year-specific pooled common-choice rule fixed (Table A12, panel A). Across the legacy year-by-allocated-seat-category grouping and the two predetermined alternatives to P1, the menu-structure share of the within-menu gap lies between 66.8 and 68.5 percent. The AIR-only common-list subsample yields 57.2 percent on a smaller candidate sample with a total gap of INR 13.44 lakh rather than INR 17.50 lakh. Because both the priority grouping and the candidate sample change in that row, its total

and opportunity-level components are not comparable to the preferred specification. We therefore report it as the most adverse priority definition among those we consider rather than as a competing benchmark.

Fourth, we vary the common-choice operator (Table A12, panel B). As established in the decomposition identity in Remark 1, changing the operator leaves $\Delta_Y^{CTC}$, $\Delta_O^{CTC}$, and $\Delta_W^{CTC}$ unchanged and affects only the allocation of the fixed within-menu gap between menu structure and residual choice. A full-period pooled rule produces a menu-structure share of the within-menu gap close to the preferred estimate, while a symmetric top–bottom reference rule produces a smaller attribution. We additionally report top- and bottom-priority reference rules as deliberately extreme diagnostics of benchmark dependence rather than as competing preferred specifications; they bracket the menu-structure share of the within-menu gap at 87.2 and 26.4 percent.

The year-specific pooled operator remains preferred for a design-based reason: it applies the same contemporaneously estimated slope vector to every priority group within an admission cycle while allowing aggregate choice relationships to evolve across cycles. We therefore interpret the preferred 68.5-percent share as a transparent benchmark-dependent attribution rather than as an institution-invariant structural parameter.

Fifth, we relax the baseline conditional-logit substitution restriction using the IIT-nested specification (Table A10). The estimated inclusive-value parameter is approximately $\widehat{\lambda} = 0.430$, indicating substantial within-IIT substitution. The positive top–bottom WTT pattern survives this relaxation: the nested specification produces a top–bottom point-estimate difference of approximately 41.5 kilometers per INR 1 lakh of CTC. Because the sequential nested specification does not recover the cross-stage covariance required for inference on that combined top–bottom difference, we treat this as a point-estimate robustness exercise rather than a formal significance test.

Other support restrictions likewise leave the central decomposition much more stable than the exact magnitude or precision of the WTT gradient (Table A8, panel A). In particular, restrictions to alternatives with observed CTC can produce larger and more precisely estimated priority differences in WTT, but these exercises change the candidate–alternative support and hence the estimand. We therefore treat them as sensitivity analyses rather than selecting specifications because they produce larger behavioral differences.

The results are also stable across admission cycles (Table A15, panels B and C). Across leave-one-cycle-out samples, the opportunity-level share remains between approximately 24.1 and 24.7 percent, while the share of the within-menu gap attributed to menu structure under common choice ranges from approximately 64.3 to 71.6 percent.

Across the full set of exercises, and setting aside the two deliberately extreme reference diagnostics, the menu-structure share of the within-menu gap lies between 57.2 and 71.6 percent. Table A13 reports parametric uncertainty for the model-dependent components, Table A7 the choice-model specification ladder and alternative estimator, and Table A9 the sensitivity of average trade-off inference to the clustering convention.

Taken together, the robustness exercises reinforce the evidence hierarchy established by the

main decomposition. The precise magnitude of priority-associated behavioral heterogeneity is less stable across specifications. By contrast, the central opportunity result is remarkably persistent: differences in average feasible CTC explain only a minority of the observed allocation gap, and differently configured opportunity sets reproduce a substantial share of the remaining within-menu separation under common behavior.

The results in this section characterize inequality across an existing priority-rationed assignment environment. Higher-priority candidates face substantially broader and higher-CTC opportunity sets, yet differences in average opportunity level account for only about one-quarter of the observed high–low allocated-program CTC gap. Under the preferred common-choice benchmark, differently structured menus reproduce most of the remaining within-menu separation.

The female-supernumerary-seat reform introduced in 2018 provides a complementary policy setting in which the institution deliberately changes the capacity and route architecture facing one group of candidates. Section 6 uses that reform to trace an explicit opportunity expansion from additional program-level capacity into the female-specific admission frontier and, ultimately, into realized allocations beyond the corresponding Gender-Neutral frontier.

# 6 Policy-Induced Opportunity Expansion: Female Supernumerary Seats

The preceding analysis shows that institutional priority affects allocations through the opportunity sets within which candidates choose. Differences across the priority distribution extend beyond the average quality of attainable programs: candidates face different combinations of institutes, fields, earnings, and geographic trade-offs, and a common estimated choice rule operating on those different menus reproduces a substantial share of the observed allocation gap.

The 2018 female-supernumerary reform provides a complementary policy application. Rather than comparing opportunity sets that arise from persistent differences in candidate priority, we study a reform that deliberately changed the capacity architecture of the market. The policy created Female-Only (FO) and Gender-Neutral (GN) seat pools and, for programs with sufficiently low historical female representation, added genuinely new capacity for female candidates. This setting allows us to examine a narrower mechanism:

$$\text{additional capacity} \longrightarrow \text{female-specific access} \longrightarrow \text{realized allocation}.$$

Our objective is not to estimate the overall welfare or enrollment effect of female affirmative action. Instead, we ask whether an institutional rule that changes capacity also changes the opportunity frontier that is central to the rest of the paper. This distinction is important because an FO seat is not necessarily a newly created seat, and an observed FO allocation is not necessarily a program that the candidate could not reach through another eligible GN route. We therefore separate the descriptive architecture of the 2018 reform from the local variation generated by its seat-construction rule.

**Table 4.** Robustness of the Decomposition and Priority-Associated CTC Trade-offs

*Panel A. Robustness of the decomposition*

| Specification | Opportunity level | Menu structure | Residual choice | Menu-structure share of within-menu gap |
|---|---|---|---|---|
| Preferred P1 | 24.4% | 51.8% | 23.8% | 68.5% |
| Exact $t-1$ CTC | 25.0% | 52.8% | 22.2% | 70.4% |
| Strict prior-year-cutoff menu | 23.3% | 50.8% | 25.8% | 66.3% |
| Revealed-augmented prior-year menu | 23.9% | 50.9% | 25.2% | 66.9% |
| Full-period pooled common choice | 24.4% | 52.5% | 23.1% | 69.4% |
| Symmetric top/bottom common choice | 24.4% | 44.0% | 31.6% | 58.2% |

*Panel B. Priority-associated CTC–distance trade-off*

| Specification | Top P1 decile | Bottom P1 decile | Difference | $p$-value |
|---|---|---|---|---|
| Preferred P1 | 103.64 | 78.67 | 24.97 | 0.094 |
| Exact $t-1$ information | 104.34 | 82.83 | 21.51 | 0.182 |
| Strict prior-year-cutoff menu | 127.02 | 72.58 | 54.43 | $< 0.001$ |
| Revealed-augmented prior-year menu | 131.24 | 85.21 | 46.03 | 0.018 |
| IIT-nested logit | 95.52 | 53.98 | 41.54 | – |

*Notes:* Panel A reports the opportunity–structure–choice decomposition of the top-minus-bottom chosen-program CTC gap under alternative information, feasible-menu, and common-choice constructions. The first three columns are shares of the total gap; the final column reports the menu-structure share of the within-menu gap. The preferred P1 specification uses contemporaneous reconstructed menus and year-specific pooled common-choice models. Exact $t-1$ CTC and prior-year-cutoff menus vary the information and feasible-set construction, while the final two rows vary the common-choice benchmark holding P1 groups and the candidate sample fixed.
Panel B reports CTC–distance trade-offs in kilometers per Rs. 1 lakh of annual program CTC for the top and bottom P1 deciles and their difference. Reported $p$-values test equality of the two trade-offs. For the IIT-nested-logit specification, inference for the combined difference is unavailable because the sequential estimator does not provide the required cross-stage covariance. CTC is annual program median compensation in lakh rupees per annum (LPA), expressed in 2023 prices. The decomposition is accounting- and model-based, and the priority-associated differences describe revealed-choice heterogeneity; neither is interpreted causally.

### 6.1 Policy architecture and observed opportunity expansion

Before 2018, IIT programs did not operate separate gender pools. The reform divided each relevant program into a Gender-Neutral pool and a Female-Only pool. Non-female candidates competed through the GN pool. Female candidates were eligible for both pools, but were considered first for the relevant FO route and entered the GN route only after failing to obtain the corresponding FO seat. The mechanism was designed to raise female representation while leaving the number of seats available to non-female candidates nearly unchanged (Baswana et al., 2019b,a).

The distinction between the FO pool and genuinely additional capacity is central to our analysis. The 2018 IIT seat matrix contained 12,079 seats: 10,227 in GN pools and 1,852 in FO pools. We reconstruct only 800 seats as genuinely additional female-supernumerary capacity. The remaining FO seats are part of the gender-pool decomposition of the underlying seat matrix and should not themselves be interpreted as net-new capacity. Table A21 reports the aggregate and reservation-category accounting.

For a continuing academic program $p$, let $C_{p,2017}$ denote its 2017 capacity and $f_{p,2017}$ its number of female allotments. Under the baseline 14-percent construction, a program with female representation below 14 percent received additional capacity $x_p^*$ satisfying

$$f_{p,2017} + x_p^* = 0.14 \left( C_{p,2017} + x_p^* \right), \tag{6.1}$$

or, before integer rounding,

$$x_p^* = \frac{0.14 C_{p,2017} - f_{p,2017}}{0.86}. \tag{6.2}$$

The implementation also accommodated integer seat construction, institute-specific targets between 14 and 20 percent, separate rules for programs already above the target, and programs introduced in 2018 (Baswana et al., 2019b). These provisions are one reason the 14-percent rule generates strong but imperfect treatment assignment rather than a sharp discontinuity.

Before exploiting that rule, Figure 4 documents what the 2018 architecture did to observed opportunities. Panel A compares the most permissive finite FO and GN closing-rank frontiers observed across the 2018 counseling rounds. Comparisons are made within the same regular reservation category, so both ranks are measured on the same merit list. Among 825 program-category cells with finite FO and GN frontiers, 759, or 92.0 percent, have

$$CR_{pc}^{FO} > CR_{pc}^{GN}.$$

The pattern is strongest in OPEN and OBC-NCL routes and weaker for ST routes; Table A23 reports the category-specific levels and relative frontier gaps.

Panel B translates these route-level frontiers into the candidate-specific physical-program opportunity sets used elsewhere in the paper. For each of the 1,712 point-identified regular FO recipients, we evaluate every GN route for which the candidate is eligible and every observed

eligible FO route. A physical program is female-specific incremental access only when the candidate can reach it through at least one FO route but through none of her eligible observed GN routes. We then union routes at the physical-program level.

This reconstruction shows that the policy architecture changed much more than the label attached to an allocated seat. The median FO recipient has 56 female-specific incremental physical programs. These programs account for 46.4 percent of the median candidate's reconstructed IIT opportunity set. Table A24 reports the corresponding distribution by candidate social category.

These quantities are descriptive of the observed 2018 market. In particular, a program lying beyond every observed eligible GN frontier is not necessarily a program the candidate would have lost in a counterfactual market without Female-Only capacity. Removing the reform would change capacities, rejection chains, assignments, and equilibrium closing ranks throughout JoSAA. Figure 4 therefore establishes that the implemented architecture substantially expanded observed female-specific opportunities; it does not re-solve the allocation system under a no-policy counterfactual.

Recent work studies other margins of the same reform. Sengupta et al. (2025) examine the implementation of the IIT female-supernumerary scheme and heterogeneity in female representation, while Gupta (2026) studies educational responses to exposure to the policy. Our exercise focuses on a different institutional margin: we isolate genuinely net-new program capacity and ask whether variation induced by the seat-construction rule shifts the observed FO/GN access frontier.

### 6.2 A fuzzy local-threshold design

We exploit the relationship between completed 2017 female representation and the creation of additional capacity in 2018. For every academic program observed in both years, define

$$s_p = \frac{f_{p,2017}}{C_{p,2017}}, \tag{6.3}$$

and express distance from the baseline threshold in percentage points as

$$r_p = 100(s_p - 0.14). \tag{6.4}$$

The policy assignment indicator is

$$Z_p = \mathbf{1}\{r_p < 0\}, \tag{6.5}$$

while the endogenous treatment is

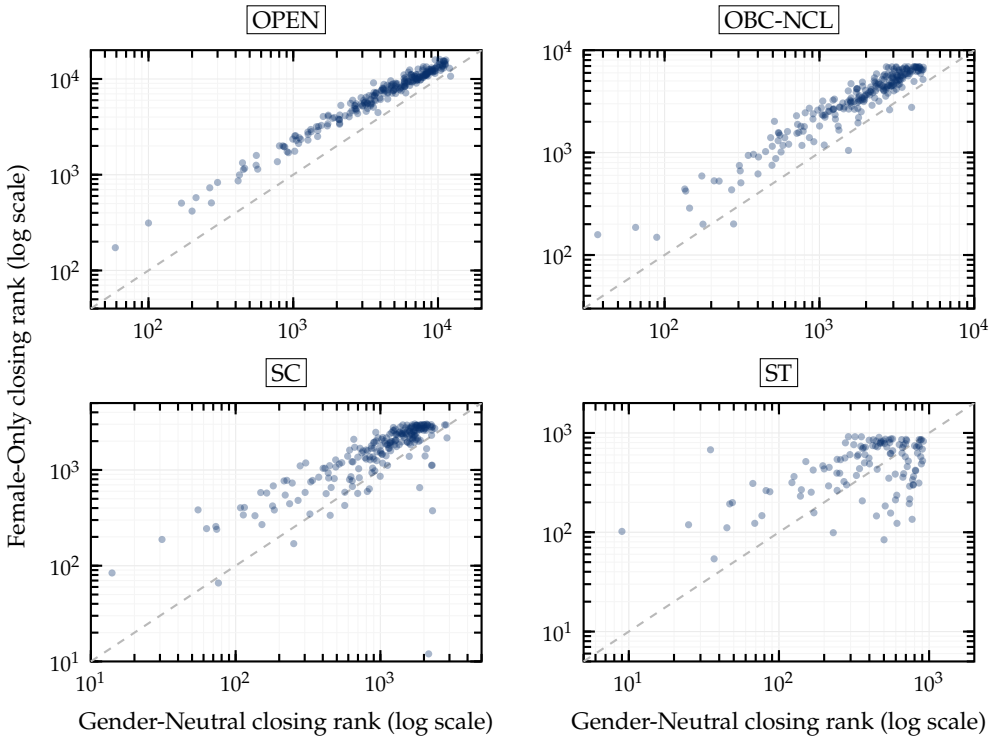
$$D_p = \mathbf{1}\{x_p > 0\}, \tag{6.6}$$

where $x_p$ is the genuinely additional capacity observed in the 2018 seat matrix.

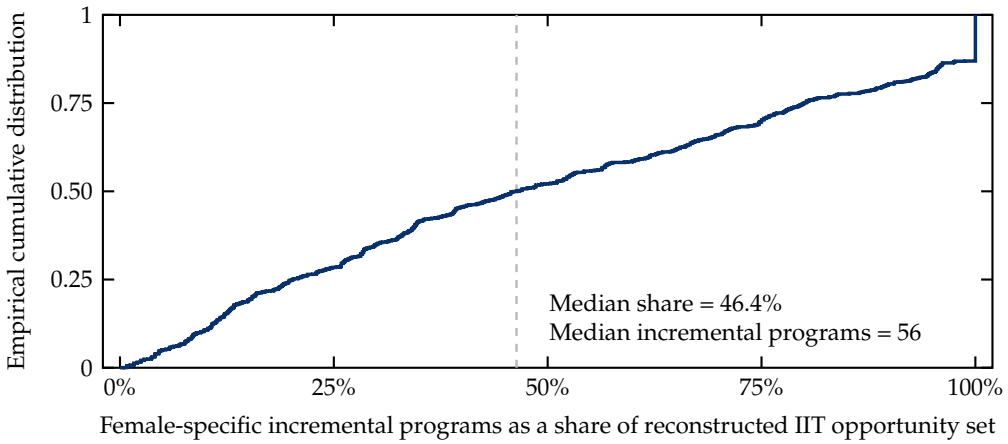
The distinction between $Z_p$ and $D_p$ is empirically important. Among the 244 continuing academic programs, 170 lie below the baseline 14-percent threshold and 162 receive positive additional capacity. The literal 14-percent rule correctly predicts the extensive-margin treatment

**Figure 4.** Female-Specific Access Frontiers and Candidate Opportunity Sets

**A. Female-Only versus Gender-Neutral closing-rank frontiers**




**B. Candidate-specific expansion of the physical-program opportunity set**




*Notes:* Panel A compares the most permissive finite Female-Only (FO) and Gender-Neutral (GN) closing ranks observed across the 2018 counseling rounds within each regular reservation category. The sample contains 825 paired program-category cells: 250 OPEN, 235 OBC-NCL, 219 SC, and 121 ST. Overall, 759 of 825 cells (92.0%) have an FO frontier above the corresponding GN frontier. Reservation categories use distinct rank lists, so closing ranks are comparable within, rather than across, categories. Panel B reports the empirical distribution across 1,712 point-identified regular FO recipients of the share of each candidate's reconstructed IIT opportunity set consisting of physical programs reachable through at least one FO route but through none of her eligible GN routes. The median candidate has 56 such female-specific incremental programs, accounting for 46.4% of her reconstructed opportunity set. Both panels are descriptive of the observed 2018 allocation architecture.

status of 230 programs, or 94.3 percent. Eleven below-threshold programs receive no positive additional capacity, while three programs at or above the threshold do receive it. We therefore treat the design as fuzzy rather than sharp. Table A22 provides the complete rule reconstruction.

Our preferred specification uses a symmetric ±4-percentage-point window around the threshold. The bandwidth was frozen before examining the final policy outcomes. This window contains 80 programs from 16 IITs at 35 distinct running-variable mass points, with 40 programs below and 40 at or above the threshold. The narrower ±3 window contains only 52 programs from 11 IITs and has a 21/31 split. Expanding to ±5 adds only nine programs and no additional IITs. We therefore use ±4 as the preferred compromise between locality and within-IIT support. The ±3 and ±5 windows provide pre-specified sensitivity checks.

For outcome $Y_p$, the preferred local specification can be written as

$$Y_p = \alpha_{\mathrm{IIT}(p)} + \tau D_p + \beta_1 r_p + \beta_2 Z_p r_p + \varepsilon_p, \qquad |r_p| \leq 4, \tag{6.7}$$

where $D_p$ is instrumented by $Z_p$. The first-stage equation replaces $Y_p$ with $D_p$. All specifications include IIT fixed effects, allow separate local-linear slopes on the two sides of the threshold, restrict the sample to IITs represented on both sides, and cluster standard errors at the IIT level.

The identifying interpretation is local. Under continuity of potential outcomes around the policy threshold, an exclusion restriction for the below-threshold assignment indicator, and monotonicity of the treatment response, the IV coefficient captures the effect of receiving positive additional capacity for programs whose treatment status is shifted by the policy rule (Hahn et al., 2001; Lee and Lemieux, 2010). We do not interpret the estimate as the average effect of the entire female-supernumerary reform across IIT programs.

The running variable also requires care. It is the ratio of integer female allotments to integer program capacity and therefore has discrete support. The preferred sample contains 35 mass points rather than a continuously varying forcing variable. We consequently report exact support explicitly, avoid relying on conventional continuous-density diagnostics, and treat weak-instrument-robust inference as central to the design (Kolesár and Rothe, 2018). Figure A2 displays the support, while Figures A3–A8 plot the corresponding IIT-fixed-effects-adjusted local relationships.

Predetermined characteristics provide no evidence of a systematic break at the threshold. In the preferred window, the estimated discontinuity in the 2016 female allotment share is −2.903 percentage points (SE = 2.529, $p = 0.270$), while the discontinuity in 2017 program capacity is 5.976 seats (SE = 12.122, $p = 0.629$). Table A25 reports the key checks, and Table A33 extends the exercise to historical capacity, category composition, female representation, and GN admission frontiers in 2015–2017.

### 6.3 From additional capacity to access and realized allocations

Panel A of Table 5 reports the first stage. In the preferred ±4 window, assignment below the 14-percent threshold increases the probability of receiving positive additional capacity by

$$0.482 \qquad (0.242).$$

The corresponding first-stage $t^2$ diagnostic is 3.97. The first stage is therefore economically substantial but not sufficiently strong to rely on conventional IV inference alone. For each structural outcome we report Anderson–Rubin tests of the null of a zero treatment effect.

Our primary access outcome asks whether the policy changes the extensive margin of observed female-specific access. For each regular program-category route, define an extension indicator equal to one when both frontiers are observed and

$$CR_{pc}^{FO} > CR_{pc}^{GN}.$$

We aggregate these indicators to the academic-program level using predetermined 2017 regular-category capacity weights. Denote the resulting weighted share by $X_p$. Missing finite FO frontiers are not interpreted as evidence of zero institutional FO capacity; for the observed-extension measure they contribute no observed positive extension. Table A28 separately documents frontier observability and weighted coverage.

Panel B of Table 5 contains the section's strongest local result. The reduced-form discontinuity in $X_p$ is 10.8 percentage points in the preferred specification, with an IIT-clustered standard error of 4.0 percentage points. The corresponding IV estimate is

$$\widehat{\tau}_X = 22.5 \text{ percentage points} \qquad (15.1).$$

The Anderson–Rubin $p$-value for $H_0 : \tau_X = 0$ is 0.017. The result is also stable in reduced form across the two neighboring frozen windows: 10.6 percentage points at ±3, 10.8 at ±4, and 10.4 at ±5. The local evidence therefore indicates that policy-induced positive additional capacity increases the extensive margin on which observed Female-Only access reaches beyond the corresponding Gender-Neutral frontier.

We obtain much less precise evidence about how far the frontier moves. We also construct $G_p$, a predetermined-capacity-weighted index of

$$\max\left\{0, \log\left(\frac{CR_{pc}^{FO}}{CR_{pc}^{GN}}\right)\right\}$$

over observed finite frontiers. At ±4, the IV point estimate is 0.128 (SE = 0.131). Mechanically exponentiating the coefficient gives 13.6 percent, but the Anderson–Rubin $p$-value is 0.418. We therefore do not interpret 13.6 percent as a precisely identified causal magnitude. Table A26 reports the full decomposition into finite-frontier activation, positive extension, and extension magnitude; Table A27 reports weak-instrument-robust confidence sets.

The final link in the mechanism is realized allocation. Here we classify a Female-Only recipient as beyond GN only when her allocated physical program lies beyond *every* eligible observed GN route available to her. This cross-route check is important. A reserved-category woman who fails to

clear her category-specific GN frontier can still have GN access through OPEN, and such a program should not be classified as female-specific incremental access.

The administrative reconstruction yields a striking descriptive result. Of 1,712 point-identified regular FO recipients, 32 receive programs that remain reachable through at least one eligible observed GN route. The remaining 1,680 receive programs beyond every eligible observed GN route. All 1,680 allocations are supported by an observed eligible FO frontier, and the exact administrative allocation route is recovered for all 1,712 recipients. Table A29 reports this certification.

Ten additional women are known to have occupied GN seats in 2018, but their individual identities are not observed in the analytical candidate file. Accounting for these women yields an overall beyond-GN share bounded between 97.56 and 98.13 percent among regular female recipients. Table A30 reports these partial-identification bounds overall, by social category, and within the frozen local samples.

These realized-use quantities should remain descriptive. They compare observed 2018 allocations with observed 2018 GN frontiers; they do not establish that 1,680 women would have lost their programs in a re-solved market without the policy.

The local IV evidence on realized recipient counts is substantially less precise. In the preferred window, Panel C of Table 5 reports an IV estimate of 2.220 (3.545) for the program-level number of point-identified beyond-GN recipients. The Anderson–Rubin $p$-value is 0.540. Normalizing by predetermined 2017 program capacity gives an IV estimate of 4.778 beyond-GN recipients per 100 pre-policy seats (SE = 3.408), with Anderson–Rubin $p = 0.353$; this specification is reported in Table A31. Thus, realized use of the observed female-specific frontier is widespread, but the local causal effect of marginal additional capacity on recipient counts is not precisely identified.

The preferred result is also sensitive to restricting the sample to programs with longer historical female-allotment records. Requiring observed 2016 history reduces the ±4 sample from 80 to 72 programs and attenuates the positive-extension IV estimate from 22.5 to 7.8 percentage points even before the historical control is added. Conditional on those same 72 programs, including the 2016 female-share control changes the estimate only from 7.8 to 6.7 percentage points. The analogous 2015–16 exercise produces the same pattern. Table A27 therefore indicates that much of the attenuation comes from sample composition rather than from historical covariate adjustment itself. For this reason, the full 80-program ±4 specification remains preferred, while the historical specifications are treated as sensitivity analyses.

Reservation-category diagnostics point to substantial heterogeneity in the observed architecture, but the local category-specific IV estimates are too imprecise to support separate causal conclusions. We therefore treat the category results in Table A32 as mechanism diagnostics rather than additional headline effects.

Taken together, the female-supernumerary reform provides a policy-induced application of the paper's central mechanism. The administrative architecture substantially expands observed female-specific opportunity sets, and local variation around the 14-percent rule provides evidence that genuinely additional capacity increases the extensive margin of that access. The data are less

informative about the magnitude of the resulting frontier movement and about causal changes in realized recipient counts.

This distinction is precisely what the broader opportunity-set framework is designed to reveal. Capacity rules do not affect allocation only by changing the number of seats in the market. They change which differentiated programs are attainable, through which priority routes, and at what frontier. Candidates then choose within those institutionally generated opportunities. The 2018 reform therefore provides direct policy evidence that the design of capacity can alter the opportunity object through which preferences are translated into centralized assignments.

**Table 5.** Female Supernumerary Capacity, Access Frontiers, and Realized Allocations

| Bandwidth (percentage points) | (1) ±3 | (2) ±4 ***Preferred*** | (3) ±5 |
|---|---|---|---|
| *Panel A. First stage* | | | |
| Below 14% threshold | 0.446<br>(0.295) | 0.482*<br>(0.242) | 0.529**<br>(0.216) |
| First-stage $t^2$ diagnostic | 2.29 | 3.97 | 5.99 |
| *Panel B. Female-access frontier* | | | |
| FO frontier exceeds GN, weighted share (pp) | 23.8<br>(17.2) | 22.5<br>(15.1) | 19.7<br>(11.8) |
| Reduced form | 10.6**<br>(3.7) | 10.8**<br>(4.0) | 10.4**<br>(4.5) |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.017 | 0.017 | 0.036 |
| Observed frontier-extension index, $G_p$ | 0.096<br>(0.164) | 0.128<br>(0.131) | 0.134<br>(0.104) |
| Implied change (%) | 10.1 | 13.6 | 14.4 |
| Reduced form | 0.043<br>(0.084) | 0.062<br>(0.074) | 0.071<br>(0.070) |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.622 | 0.418 | 0.323 |
| *Panel C. Realized use of the expanded frontier* | | | |
| Beyond-GN female recipients | 5.357<br>(4.983) | 2.220<br>(3.545) | 3.658<br>(4.383) |
| Reduced form | 2.391<br>(2.451) | 1.069<br>(1.706) | 1.937<br>(2.434) |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.352 | 0.540 | 0.439 |
| Programs | 52 | 80 | 89 |

*Notes:* Columns vary the symmetric bandwidth around the 14 percent 2017 female-share threshold; column (2), using a ±4-percentage-point bandwidth, is the preferred specification. The instrument indicates programs below the threshold, and the endogenous treatment indicates receipt of positive genuinely additional female-supernumerary capacity in 2018. All specifications are local linear with separate slopes, IIT fixed effects, and IIT-clustered standard errors, restricting the sample to IITs represented on both sides of the threshold. Panel B uses predetermined 2017 regular-category capacity weights. The first outcome measures the share of this weight for which the observed Female-Only (FO) frontier exceeds the corresponding Gender-Neutral (GN) frontier. The second outcome, $G_p$, measures the weighted magnitude of the observed female-specific frontier extension; the implied-change row reports $100[\exp(\hat{\beta}) - 1]$. Panel C reports point-identified regular FO recipients whose allocated physical program lies beyond every eligible observed GN route; this count is invariant to the unidentified identities of the ten women allocated through GN seats in 2018. Parentheses report IIT-clustered standard errors. Significance stars are based on conventional IIT-clustered Wald tests; Anderson–Rubin $p$-values provide weak-instrument-robust inference for the structural IV effects. $^{*}p < 0.10$, $^{**}p < 0.05$, $^{***}p < 0.01$.

# 7 Discussion and Conclusion

The empirical results point to a broader distinction between *choice* and the *opportunity environment in which choice is exercised*. In the IIT market, candidates do not choose from a common set of programs and differ only in which alternative they prefer. Institutional priority, eligibility rules, reservation

routes, seat pools, and program-specific admission frontiers first determine which alternatives are attainable. Preferences are then expressed within those candidate-specific menus. Empirical work often emphasizes one side of this process: either the rules that determine access or the preferences that determine choice. This paper studies the two jointly.

This distinction matters for interpreting allocation inequality. The observed top–bottom priority gap in chosen-program CTC is INR 17.50 lakh. Differences in average feasible CTC account for INR 4.27 lakh, or 24.4 percent, of that gap. Yet average opportunity quality is only one property of a feasible menu. When the same estimated choice rule is applied to the different menus faced by higher- and lower-priority candidates, their menu configurations generate an additional INR 9.07 lakh of separation. Under the preferred benchmark, this menu-structure component accounts for 68.5 percent of the allocation gap that remains after subtracting differences in average feasible CTC. Differences in institutional opportunity therefore account for roughly three-quarters of the observed CTC gap.

The implication is not that preferences are unimportant. Candidates sort strongly within the opportunities available to them, and revealed trade-offs vary across the priority distribution. Rather, the results show that much of the observed allocation gap does not require different behavioral rules across priority groups. The same rule can generate very different outcomes when it is applied to differently structured opportunity sets.

### 7.1 Opportunity inequality is multidimensional

A common way to summarize unequal access is through a scalar measure: the number of options available, the quality of the best attainable option, or the average quality of the feasible set. These measures are informative, but they need not characterize the economic decision problem created by the assignment mechanism.

Two candidates can face menus with the same average program quality and still face very different choices. High-CTC programs may be geographically close in one menu and far away in another. A strong placement program may come bundled with a less-preferred discipline. Prestigious institutes may be available only through fields that the candidate ranks below alternatives at other IITs. Because program characteristics are bundled, the value of adding or removing an alternative depends on how that alternative changes the set of trade-offs available within the menu.

This is why the distinction between *opportunity level* and *opportunity structure* is economically useful. The opportunity-level component asks whether one group has access, on average, to better alternatives. The menu-structure component asks what happens when a common behavioral rule is allowed to operate on the actual combinations of attributes that each group can attain. The latter is not summarized by average feasible CTC, menu size, or any single program characteristic. In the IIT market, most of the within-menu gap can be reproduced under common choice because higher- and lower-priority menus contain different combinations of compensation, distance, placement, institutional quality, discipline, and other program attributes.

This observation also changes how opportunity inequality should be measured. If access is

summarized only by the average or highest-quality feasible alternative, an assignment system can appear substantially more equal than the choice problems it actually creates. Opportunity is therefore not only a question of *how much* is available, but of *which trade-offs* the institution makes available together.

### 7.2 Observed sorting need not reveal unconstrained preferences

The results also caution against interpreting differences in final allocations as direct evidence of preference differences. In a rationed assignment market, observed sorting reflects at least two conceptually distinct objects: the opportunities produced by the mechanism and the decisions candidates make within those opportunities.

This issue is especially relevant when preferences appear to vary with priority. We estimate meaningful differences in revealed CTC–distance trade-offs across the priority distribution, although the top–bottom difference in the preferred specification is only marginally statistically distinguishable from zero. These differences may reflect genuine taste heterogeneity, but priority is not randomly assigned. Exam performance is correlated with preparation, family resources, geography, information, coaching, academic interests, and other characteristics that can also affect college choice. We therefore interpret these estimates as *priority-associated revealed-choice heterogeneity*, not as a causal effect of examination rank on preferences.

When a common estimated choice rule is imposed on candidates' own feasible menus, the predicted allocation gradient remains large. An analyst who observes only final assignments could therefore attribute to preferences a substantial amount of sorting that can instead be generated by differences in institutional opportunity.

The same qualification applies to the residual component of the decomposition. The residual rank-associated choice term should not be read as an estimate of pure taste heterogeneity. It contains any systematic difference between observed allocations and the common-choice benchmark, including differences in preferences, information, consideration, omitted program attributes, decision rules, and model approximation. The decomposition is therefore most useful as a disciplined separation between what can be reproduced by opportunity sets under common behavior and what remains unexplained by that benchmark.

This interpretation extends beyond college admissions. In school choice, public housing, medical matching, selective labor markets, and other rationed environments, agents with different priorities frequently choose from different attainable sets. Comparing their realized choices without reconstructing those constraints risks conflating the structure of access with the structure of preferences.

### 7.3 Implications for market and policy design

The results have a direct implication for the design of centralized allocation systems. Policies that modify priority or capacity affect outcomes not only by changing the number or average quality

of alternatives available to a group. They can change the configuration of the feasible menu and therefore the choices induced by a given behavioral rule.

The female-supernumerary reform provides a concrete illustration. The policy did not merely create a labeled pool of seats for women. It changed the female-specific admission frontier. Across observed regular-category frontiers, Female-Only access extends beyond Gender-Neutral access in 92.0 percent of paired cells. For point-identified regular FO recipients, the median female-specific increment is 56 physical programs, accounting for 46.4 percent of the reconstructed IIT opportunity set.

The local threshold analysis strengthens the connection to institutional design. Around the 14-percent historical-female-share rule, programs induced into positive genuinely additional capacity exhibit a 10.8-percentage-point reduced-form increase in the weighted share of routes on which the observed FO frontier extends beyond GN. The corresponding IV estimate is 22.5 percentage points, with an Anderson–Rubin $p$-value of 0.017. The evidence is less precise for the magnitude of the frontier shift and for realized recipient counts, so these estimates do not identify the full effect of the female-supernumerary policy. We therefore interpret the reform not as a complete evaluation of affirmative action, but as policy-induced evidence that an administrative capacity rule can move the opportunity object measured throughout the paper.

This distinction is relevant for affirmative action and capacity policy more generally. The consequences of adding a seat depend on where the seat is created and on the priority route through which candidates can reach it. A new seat in a program already attainable to the relevant candidates may have little effect on their feasible menu. The same seat placed at a program that extends a binding frontier can alter many candidates' attainable alternatives and the trade-offs among those alternatives. Counting seats alone can therefore be an incomplete description of policy incidence.

The same reasoning applies to priority reforms. A change in priority need not simply move a candidate from a lower-quality to a higher-quality menu. It can open alternatives that differ jointly in field, location, earnings, and institutional attributes. Evaluating such reforms only through changes in mean access or final enrollment can miss an important mechanism through which the allocation system affects behavior.

Our results do not imply that centralized assignment mechanisms should equalize menus across candidates. Priority rules exist precisely because seats are scarce and institutions have explicit objectives over their allocation. Instead, the analysis identifies a margin that policy evaluation should make visible: when priorities or capacities change, one should ask not only how many additional alternatives become available or how their average quality changes, but how the reform changes the structure of the choices candidates are able to make.

### 7.4 Interpretation and scope

Several features of the analysis determine the scope of these conclusions. First, the baseline feasible menus are reconstructed from observed admission frontiers. Closing ranks are equilibrium outcomes, so a contemporaneous frontier is not a primitive institutional constraint independent

of aggregate demand. It can also differ from the information candidates possessed when they submitted their rankings. The results using immediately preceding-year program attributes and prior-year closing-rank menus are therefore important: the central decomposition changes little when opportunity sets are constructed from information and frontiers available before the current admission cycle. This stability supports the interpretation of the baseline reconstruction as an economically meaningful measure of attainable opportunity rather than an artifact of contemporaneous cutoffs.

Second, feasibility in the data is not the same as a candidate's subjective consideration set. A program that is institutionally attainable may be unknown to the candidate, omitted from her submitted ranking, or effectively excluded because of family, financial, informational, or other constraints that we do not observe. Our object is consequently an *institutional opportunity set*: the programs attainable under observed priority, eligibility, and admission-frontier rules. The paper does not identify all psychological or informational constraints governing consideration.

Third, the common-choice decomposition is benchmark-dependent. The accounting identity separating average opportunity level from the remaining within-menu gap does not require a behavioral model. The subsequent division of the within-menu gap into menu structure and residual rank-associated choice does. Different common-choice reference rules alter the exact attribution between those two terms. The robustness exercises show that menu structure remains the majority component of the within-menu CTC gap under the principal alternatives, but the preferred 68.5-percent share should not be interpreted as an institution-free structural constant.

Fourth, the estimated program-attribute coefficients are revealed-choice associations. CTC, placement, institutional quality, and geography are not randomly assigned program characteristics. The implied willingness-to-travel measures summarize how observed choices covary with these attributes conditional on the included controls and reconstructed menus; they are not causal willingness-to-pay parameters. The attenuation of several attribute coefficients when academic-program fixed effects are added reinforces the importance of this interpretation.

Finally, our analysis is not a welfare evaluation of JoSAA or of the female-supernumerary reform. We do not observe applicants' outside options, the full set of relevant competitors outside the IIT allocation sample, or all determinants of utility. Nor do we re-solve the entire market under alternative priority and capacity systems. The common-choice exercises are measurement benchmarks, not welfare counterfactuals, and the beyond-GN classification in the female-policy section compares realized allocations with observed frontiers rather than with a re-equilibrated no-policy market.

# 8 Conclusion

These limitations leave the central empirical conclusion intact. Across a high-stakes centralized admissions system, priority is associated with large differences not only in the breadth and average quality of feasible opportunities, but also in their internal configuration. Applying common

revealed-choice behavior to those different menus reproduces most of the observed allocation gap. And when policy directly changes one component of that opportunity architecture, the corresponding access frontier moves as well.

The broader implication is that opportunity sets should be treated as an economic object rather than as background to the study of choice. Allocation outcomes in priority-rationed markets are jointly produced by the institution that determines what is attainable and the behavior that selects among those alternatives. Understanding inequality in such markets consequently requires studying not only what agents choose, but the institutionally generated opportunities over which those choices can be expressed.

## References


Abdulkadiroğlu, A. (2005). College admissions with affirmative action. *International Journal of Game Theory*, 33(4):535–549.

Abdulkadiroğlu, A., Agarwal, N., and Pathak, P. A. (2017). The welfare effects of coordinated assignment: Evidence from the new york city high school match. *American Economic Review*, 107(12):3635–3689.

Abdulkadiroğlu, A. and Sönmez, T. (2003). School choice: A mechanism design approach. *American Economic Review*, 93(3):729–747.

Agarwal, N. (2015). An empirical model of the medical match. *American Economic Review*, 105(7):1939–1978.

Agarwal, N. and Somaini, P. (2018). Demand analysis using strategic reports: An application to a school choice mechanism. *Econometrica*, 86(2):391–444.

Arcidiacono, P., Hotz, V. J., and Kang, S. (2012). Modeling college major choices using elicited measures of expectations and counterfactuals. *Journal of Econometrics*, 166(1):3–16.

Arslan, H. A. (2021). Preference estimation in centralized college admissions from reported lists: Ha arslan. *Empirical Economics*, 61(5):2865–2911.

Aygün, O. and Turhan, B. (2017). Large-scale affirmative action in school choice: Admissions to iits in india. *American Economic Review*, 107(5):210–213.

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

Baswana, S., Chakrabarti, P. P., Chandran, S., Kanoria, Y., and Patange, U. (2019a). Centralized admissions for engineering colleges in india. *INFORMS Journal on Applied Analytics*, 49(5):338–354.

Baswana, S., Chakrabarti, P. P., Kanoria, Y., Patange, U., and Chandran, S. (2019b). Joint seat allocation 2018: An algorithmic perspective.

Baswana, S., Chakrabarti, P. P., Kanoria, Y., Patange, U., Chandran, S., and Kumar, A. (2015). Joint seat allocation: An algorithmic perspective.

Bertrand, M., Hanna, R., and Mullainathan, S. (2010). Affirmative action in education: Evidence from engineering college admissions in india. *Journal of public economics*, 94(1-2):16–29.

Bonhomme, S., Jolivet, G., and Leuven, E. (2016). School characteristics and teacher turnover: Assessing the role of preferences and opportunities. *The Economic Journal*, 126(594):1342–1371.

Cameron, A. C., Gelbach, J. B., and Miller, D. L. (2011). Robust inference with multiway clustering. *Journal of Business & Economic Statistics*, 29(2):238–249.

Chetty, R., Deming, D. J., and Friedman, J. N. (2026). Diversifying society's leaders? the determinants and causal effects of admission to highly selective private colleges. *The Quarterly Journal of Economics*, 141(1):51–145.

Chetty, R., Friedman, J. N., Saez, E., Turner, N., and Yagan, D. (2020). Income segregation and intergenerational mobility across colleges in the united states. *The Quarterly Journal of Economics*, 135(3):1567–1633.

Choudhury, P., Ganguli, I., and Gaulé, P. (2023). Top talent, elite colleges, and migration: Evidence from the indian institutes of technology. *Journal of Development Economics*, 164:103120.

Conlon, C. T. and Mortimer, J. H. (2013). Demand estimation under incomplete product availability. *American Economic Journal: Microeconomics*, 5(4):1–30.

Dur, U., Kominers, S. D., Pathak, P. A., and Sönmez, T. (2018). Reserve design: Unintended consequences and the demise of boston's walk zones. *Journal of Political Economy*, 126(6):2457–2479.

Fack, G., Grenet, J., and He, Y. (2019). Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109(4):1486–1529.

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

Gibbons, S. and Vignoles, A. (2012). Geography, choice and participation in higher education in england. *Regional science and urban economics*, 42(1-2):98–113.

Grenet, J., He, Y., and Kübler, D. (2022). Preference discovery in university admissions: The case for dynamic multioffer mechanisms. *Journal of Political Economy*, 130(6):1427–1476.

GuimarÃ£es, P., Figueirdo, O., and Woodward, D. (2003). A tractable approach to the firm location decision problem. *The Review of Economics and Statistics*, 85(1):201–204.

Gupta, R. (2026). Fixing the leaky pipeline: Affirmative action in local elite colleges and subject choice. *Southern Economic Journal*, 93(1):457–473.

Hafalir, I. E., Yenmez, M. B., and Yildirim, M. A. (2013). Effective affirmative action in school choice. *Theoretical Economics*, 8(2):325–363.

Hahn, J., Todd, P., and der Klaauw, W. V. (2001). Identification and estimation of treatment effects with a regression-discontinuity design. *Econometrica*, 69(1):201–209.

Horowitz, J. L. (1991). Modeling the choice of choice set in discrete-choice random-utility models. *Environment and Planning a*, 23(9):1237–1246.

Joint Seat Allocation Authority (2019). Business rules for joint seat allocation for the academic programs offered by the iits, nits, iiest, iiits and other-gftis for the academic year 2019–20. June 16, 2019.

Joint Seat Allocation Authority (2023). Business rules for joint seat allocation for the academic programs offered by the iits, nits, iiest, iiits, and other-gftis for the academic year 2023–24. June 7, 2023.

JoSAA (2018). Business rules for joint seat allocation for academic programs offered by the iits, nits, iiest, iiits and other-gftis. https://josaa.nic.in/webinfocms/Handler/FileHandler.ashx?i=File&ii=109&iii=Y. Accessed August 1, 2025.

JoSAA (2020). Business rules for joint seat allocation for academic programs offered by the iits, nits, iiest, iiits and other-gftis. https://josaa.nic.in/. Accessed August 1, 2025.

JoSAA (2023). Business rules for joint seat allocation for academic programs offered by the iits, nits, iiest, iiits and other-gftis. https://josaa.nic.in/. Accessed August 2, 2025.

Kamada, Y. and Kojima, F. (2015). Efficient matching under distributional constraints: Theory and applications. *American Economic Review*, 105(1):67–99.

Kapor, A., Karnani, M., and Neilson, C. (2024). Aftermarket frictions and the cost of off-platform options in centralized assignment mechanisms. *Journal of Political Economy*, 132(7):2346–2395.

Kapor, A. J., Neilson, C. A., and Zimmerman, S. D. (2020). Heterogeneous beliefs and school choice mechanisms. *American Economic Review*, 110(5):1274–1315.

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

Kolesár, M. and Rothe, C. (2018). Inference in regression discontinuity designs with a discrete running variable. *American Economic Review*, 108(8):2277–2304.

Lee, D. S. and Lemieux, T. (2010). Regression discontinuity designs in economics. *Journal of Economic Literature*, 48(2):281–355.

Long, B. T. (2004). How have college decisions changed over time? an application of the conditional logistic choice model. *Journal of econometrics*, 121(1-2):271–296.

McFadden, D. (1972). Conditional logit analysis of qualitative choice behavior. Working paper, Institute of Urban and Regional Development, University of California, Berkeley.

Mello, U. (2022). Centralized admissions, affirmative action, and access of low-income students to higher education. *American Economic Journal: Economic Policy*, 14(3):166–197.

Schmidheiny, K. and Brülhart, M. (2011). On the equivalence of location choice models: Conditional logit, nested logit and poisson. *Journal of Urban Economics*, 69(2):214–222.

Sengupta, N., Munshi, R., and Kaur, R. (2025). A foot in the door: evaluating the supernumerary seats scheme for girls in the iits. *Higher Education*, pages 1–26.

Shorrer, R. I. and Sóvágó, S. (2024). Dominated choices under deferred acceptance mechanism: The effect of admission selectivity. *Games and Economic Behavior*, 144:167–182.

Sönmez, T. and Yenmez, M. B. (2022). Affirmative action in india via vertical, horizontal, and overlapping reservations. *Econometrica*, 90(3):1143–1176.

Suhonen, T. (2014). Field-of-study choice in higher education: does distance matter? *Spatial Economic Analysis*, 9(4):355–375.

Wiswall, M. and Zafar, B. (2015). Determinants of college major choice: Identification using an information experiment. *The Review of Economic Studies*, 82(2):791–824.

Ye, X. (2024). Improving college choice in centralized admissions: Experimental evidence on the importance of precise predictions. *Education Finance and Policy*, 19(2):308–340.

Zeng, Z. and Xie, Y. (2008). A preference-opportunity-choice framework with applications to intergroup friendship. *American Journal of Sociology*, 114(3):615–648.

Zimmerman, S. D. (2019). Elite colleges and upward mobility to top jobs and top incomes. *American Economic Review*, 109(1):1–47.

# A Data, Institutional Details, and Opportunity-Set Construction

This appendix documents the construction of the estimation sample, audits the reconstructed candidate opportunity sets, defines the priority measures used throughout the paper, and records the institutional features of JoSAA that the empirical framework simplifies. Section A.1 contains the institutional mapping, and Sections A.2 through A.4 contain the three analytical results stated in the main text.

Table A1 lists each data component with its source, years of coverage, unit of construction, and timing relative to choice submission. Timing is recorded because program attributes and admission frontiers are realized at different points in the admission cycle; Figure A1 shows the placement and compensation release windows against the JoSAA preference-submission dates for each cycle. Table A2 reports sample composition by admission cycle and social category.

Table A3 audits the reconstructed menus. The final sample contains 119,285 candidate choice sets and 18,632,124 candidate–program alternatives. Every choice set has exactly one chosen program, no duplicate alternatives, an internally consistent menu size, and at least two alternatives. For 118,567 candidates, or 99.40 percent, the allocated program is recovered directly from the published opening- and closing-rank frontier using the candidate's own eligible routes. The remaining 718 candidates, 0.60 percent of the sample, have the allocated program appended under the revealed-feasible rule described in Section 3.4; none of these has a zero direct frontier. Panel B shows that augmentation is small in every cycle and concentrated in 2019, where it affects 4.02 percent of candidates. Mean menu size is 156.19 under the direct frontier and 156.20 after augmentation, so the correction changes menu breadth negligibly.

Table A4 defines the four scalar priority measures. None of them uses a candidate's chosen program, allocated-seat category, or realized seat pool, which is what makes them predetermined with respect to the allocation being studied. Table A5 reports coverage by candidate category. The preferred measure P1 is observed for 119,284 of 119,285 candidates. All-India Rank alone is observed for only 41.0 percent of SC candidates and 26.0 percent of ST candidates, which is why the AIR-only measure P4 is used as a selected-subsample robustness check rather than as a baseline.

Section A.1 documents the institutional features of the mechanism that the empirical framework treats in reduced form, including multiple counseling rounds, freeze, float and slide decisions, de-reservation, and the multiple deferred-acceptance runs within a cycle. Table A6 summarizes each feature and the corresponding analytical treatment. Section A.2 proves Proposition 1 for JoSAA's virtual-route representation, Section A.3 proves Lemma 1, and Section A.4 derives the decomposition identity and the common-choice reproduction shares used in Section 5.4.

**Table A1.** Data Sources, Variable Construction, and Information Timing

| Variable / data component | Source | Years | Construction / unit | Timing relative to choice submission |
|---|---|---|---|---|
| Candidate rank and category | JEE/JoSAA candidate records | 2015–2023 | Candidate priority measures | Known before choice filling |
| Program and seat information | JoSAA seat/allocation records | 2015–2023 | IIT × program × seat category/pool | Admission-cycle information |
| Closing ranks | JoSAA OR–CR records | 2015–2023 | Used to reconstruct feasible menus | Realized during allocation; prior-year values used in ex-ante robustness |
| Distance | JEE exam-centre geography + IIT location | 2015–2023 | Haversine distance, km | Predetermined |
| Median CTC | IIT placement records | Relevant years | Annual median CTC; CPI-deflated to 2023 prices | Contemporaneous release can overlap normal-year choice filling; exact $t-1$ values used in robustness |
| Placement rate | IIT placement records | Relevant years | Percentage placed | Same timing issue as CTC |
| NIRF score | NIRF | 2016–2023 | Institute score | Publicly released before JoSAA choice filling in sample years |
| SHRUG covariates | SHRUG | Matched years | District poverty/remoteness measures | Predetermined |

*Notes:* The table summarizes the principal administrative and institutional data sources used in the analysis, their construction, and the timing of the information relative to candidates' submitted preference orderings. CTC denotes median cost to company; CPI denotes the Consumer Price Index; NIRF denotes the National Institutional Ranking Framework; OR–CR denotes opening and closing ranks. Exact $t-1$ program outcomes are used in the information-timing robustness analysis.

**Table A2.** Sample Composition by Admission Cycle and Social Category

| Admission cycle | Candidates | Mean $K_i$ | Median $K_i$ | OPEN | EWS | OBC-NCL | SC | ST |
|---|---|---|---|---|---|---|---|---|
| 2015 | 9,684 | 132.8 | 143.0 | 50.0% | – | 27.0% | 15.2% | 7.7% |
| 2016 | 10,315 | 143.1 | 153.0 | 50.0% | – | 27.5% | 15.3% | 7.3% |
| 2017 | 10,629 | 148.6 | 159.0 | 49.8% | – | 27.2% | 15.3% | 7.7% |
| 2018 | 11,657 | 151.6 | 161.0 | 50.6% | – | 28.0% | 15.1% | 6.3% |
| 2019 | 13,213 | 151.8 | 161.0 | 43.1% | 7.6% | 27.2% | 15.2% | 6.9% |
| 2020 | 14,955 | 153.1 | 161.0 | 41.8% | 10.8% | 28.7% | 15.7% | 3.0% |
| 2021 | 15,515 | 155.5 | 162.0 | 41.3% | 10.6% | 27.6% | 14.7% | 5.8% |
| 2022 | 16,331 | 166.5 | 176.0 | 38.9% | 11.3% | 27.2% | 15.2% | 7.4% |
| 2023 | 16,986 | 182.2 | 193.0 | 38.3% | 11.8% | 27.4% | 15.1% | 7.4% |
| **All years** | **119,285** | **156.2** | **165.0** | **43.9%** | **6.8%** | **27.5%** | **15.2%** | **6.5%** |

*Notes:* The table reports candidate-level composition of the reconstructed analysis sample. Each candidate appears once and receives equal weight. $K_i$ is the number of academic programs in candidate $i$'s reconstructed feasible menu. Social category refers to the candidate's own category, not the category of the allocated seat. OPEN denotes the general category, EWS the Economically Weaker Sections category, OBC-NCL Other Backward Classes–Non-Creamy Layer, SC Scheduled Castes, and ST Scheduled Tribes. Category columns report shares of candidates within each admission cycle. EWS was introduced into the admission system in 2019, so the EWS cells for 2015–2018 are shown as dashes. The sample is restricted throughout to reconstructed candidate choice sets with at least two feasible academic programs.

**Table A3.** Validation of Reconstructed Candidate Opportunity Sets

**Panel A. Final frontier validation**

| Validation measure | Count | Share |
|---|---|---|
| Final candidate choice sets | 119,285 | 100.00% |
| Final candidate–program alternatives | 18,632,124 | – |
| Choice sets with exactly one chosen program | 119,285 | 100.00% |
| Duplicate candidate–program alternatives | 0 | – |
| Choice sets with internally consistent $K_i$ | 119,285 | 100.00% |
| Choice sets satisfying $K_i \geq 2$ | 119,285 | 100.00% |
| Chosen program recovered directly from published OR–CR frontier | 118,567 | 99.40% |
| Chosen program appended as revealed-feasible | 718 | 0.60% |
| Of which: zero direct frontier | 0 | 0.00% |

**Panel B. Direct versus final frontier by admission cycle**

| Admission cycle | Candidates | Mean $K_i^D$ | Mean $K_i$ | Augmented | Augmentation share |
|---|---|---|---|---|---|
| 2015 | 9,684 | 132.78 | 132.79 | 3 | 0.03% |
| 2016 | 10,315 | 143.14 | 143.14 | 2 | 0.02% |
| 2017 | 10,629 | 148.62 | 148.62 | 14 | 0.13% |
| 2018 | 11,657 | 151.56 | 151.56 | 1 | 0.01% |
| 2019 | 13,213 | 151.75 | 151.79 | 531 | 4.02% |
| 2020 | 14,955 | 153.14 | 153.15 | 25 | 0.17% |
| 2021 | 15,515 | 155.53 | 155.53 | 58 | 0.37% |
| 2022 | 16,331 | 166.51 | 166.51 | 4 | 0.02% |
| 2023 | 16,986 | 182.22 | 182.22 | 80 | 0.47% |
| **All years** | **119,285** | **156.19** | **156.20** | **718** | **0.60%** |

*Notes:* The table audits the reconstructed candidate-specific feasible menus used in the analysis. The direct frontier is constructed from published JoSAA opening- and closing-rank (OR–CR) records using each candidate's eligible priority routes, including applicable category and seat-pool rules. Multiple eligible routes leading to the same IIT–program are collapsed to one academic program. If the documented allocated program is absent from the direct OR–CR frontier, that program is appended once as revealed-feasible. $K_i^D$ denotes the size of the direct frontier and $K_i$ the final revealed-augmented frontier. No other programs are added through this augmentation. The analysis retains candidate choice sets with $K_i \geq 2$.

**Table A4.** Priority Definitions Used in the Empirical Analysis

| Priority definition | Rank used | Percentile construction | Uses realized allocation? |
|---|---|---|---|
| Legacy: year × allocated-seat category | Existing `rank_used` / `rank_dec` | Admission year × allocated-seat category | Yes |
| P1: predetermined own-list priority *(preferred)* | OPEN: AIR/CRL; EWS, OBC-NCL, SC, ST: own-category merit-list rank | Admission year × candidate category | No |
| P2: P1 stratified by AIR availability | Same rank as P1 | Admission year × candidate category × AIR availability | No |
| P3: predetermined analytical-rank convention | OPEN, EWS, OBC-NCL: AIR; SC, ST: own-category merit-list rank | Admission year × candidate category | No |
| P4: AIR-only common-list subsample | AIR | Admission year | No |

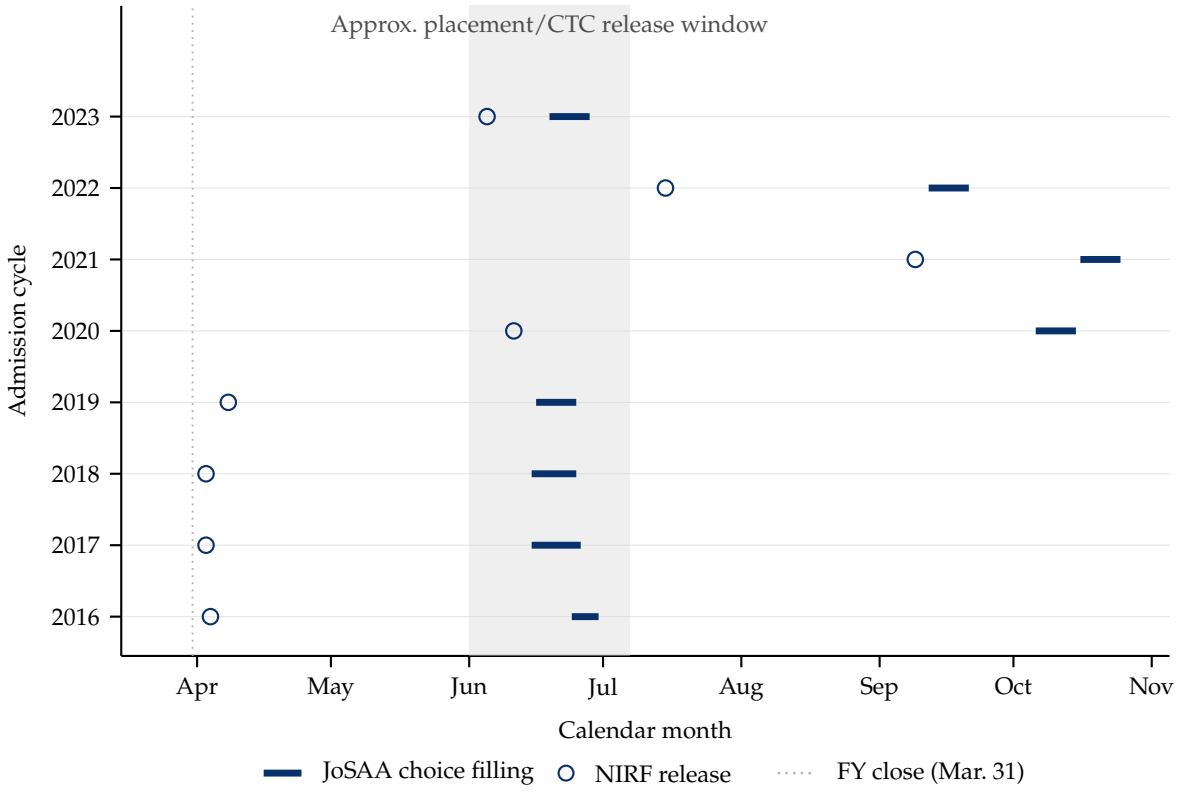
*Notes:* P1 is the preferred predetermined scalar priority summary. OPEN candidates use AIR/CRL rank, while EWS, OBC-NCL, SC, and ST candidates use their own-category merit-list rank; deciles are formed within admission year and candidate category. P2 additionally stratifies the percentile by AIR availability. P3 follows the stored analytical-rank convention without conditioning on realized allocated-seat category. P4 uses AIR only and therefore applies to a selected common-list subsample. JoSAA priority is route-specific and multidimensional; none of these scalar summaries is asserted to fully characterize priority for every virtual-program route. P1–P4 use no chosen program, allocated-seat category, or realized seat pool in their construction.

**Table A5.** Coverage of Predetermined Priority Measures by Candidate Category

| Candidate category | Candidates | P1 own-list observed | P1 coverage *(preferred)* | Analytical-convention coverage | AIR coverage |
|---|---|---|---|---|---|
| OPEN | 52,405 | 52,405 | 100.0% | 100.0% | 100.0% |
| EWS | 8,098 | 8,098 | 100.0% | 99.5% | 99.5% |
| OBC-NCL | 32,853 | 32,852 | 100.0% | 97.6% | 97.6% |
| SC | 18,141 | 18,141 | 100.0% | 100.0% | 41.0% |
| ST | 7,788 | 7,788 | 100.0% | 100.0% | 26.0% |

*Notes:* Counts aggregate the final reconstructed candidate universe across admission cycles. P1 is the preferred predetermined priority measure: OPEN candidates use AIR/CRL rank, while EWS, OBC-NCL, SC, and ST candidates use their own-category merit-list rank. The analytical convention uses AIR for OPEN, EWS, and OBC-NCL candidates and own-category rank for SC and ST candidates. P1 has essentially complete coverage: 119,284 of 119,285 candidates have an observed P1 rank; the single missing case is an OBC-NCL candidate. Coverage rates are rounded to one decimal place. AIR alone is substantially less complete for SC and ST candidates and is therefore used only as a selected-subsample robustness measure. Candidate category refers to the candidate's own category rather than the category of the allocated seat.

**Figure A1.** Timing of Program Information and JoSAA Preference Submission




*Notes:* The figure compares public NIRF release dates with JoSAA choice-filling windows for the 2016–2023 admission cycles. The shaded band is the approximate June-to-early-July period in which program placement and CTC outcomes are generally released, following the March 31 fiscal-year close and subsequent audit and compilation; it represents an institutional timing window rather than an exact publication date for every program-year. NIRF release precedes choice filling in every plotted cycle, but placement and CTC releases can overlap with preference submission in the normal June cycles, which motivates the robustness check that uses only immediately preceding-year information. The figure describes the pre-choice information environment and does not imply that NIRF, placement, or CTC information was displayed in the JoSAA choice interface.

## A.1 JoSAA mechanism details and analytical treatment

This appendix clarifies how the institutional features described in Section 2 map into the formal route-level environment and the empirical reconstruction used in Section 3. The distinction is useful because the historical JoSAA process contains both regular priority routes that can be represented directly through observed merit lists and program-section cutoffs and additional business rules that operate across runs, rounds, or special allocation channels.

The formal results in Section 2.9 are stated for the regular route-level allocation environment. In that environment, a physical academic program $j$ can be reached through one or more virtual routes $r \in \mathcal{R}_{jt}$, each governed by an applicable merit list, eligibility condition, and program–route priority frontier. The corresponding institutional opportunity set is

$$Q_{it}(\boldsymbol{\kappa}_t) = \left\{ j \in \mathcal{J}_t : \exists r \in \mathcal{R}_{jt} \text{ such that } e_{irt} = 1 \text{ and } q_{irt} \leq \kappa_{jrt} \right\}.$$

Our baseline empirical reconstruction retains the regular IIT routes for which this rank–eligibility–frontier mapping can be implemented consistently from the administrative records. These include the standard OPEN and category-specific routes and, after their introduction, the applicable Gender-Neutral and Female-Only program-sections. Different virtual routes to the same institute–degree program are collapsed to one physical academic program after route-level qualification has been evaluated.

**Multiple rounds and Min-Cutoff protections.** JoSAA is a multi-round mechanism rather than a single static allocation. Between rounds, candidates can accept, reject, freeze, float, slide, or withdraw subject to the applicable Business Rules, and their active preference lists can consequently change. The MRDA implementation also carries forward program-specific Min-Cutoff information so that a candidate who remains active is protected against mechanically losing a previously secured allocation solely because the market is rerun. The algorithmic reports therefore distinguish the priority frontier generated within a particular run from the protections governing continuation across rounds.

Our reconstruction does not attempt to reproduce the entire sequence of candidate states and Min-Cutoff calculations. Instead, for every regular candidate–program–route combination, we use the most permissive observed closing-rank frontier across the allocation rounds in admission cycle $t$. This is the "any-round" construction described in Section 3.2. It is intended to recover whether the program ever lay inside the candidate's realized institutional priority frontier during the historical allocation process, rather than to reproduce the allocation state in one arbitrarily selected round.

**De-reservation and repeated DA runs.** Within a round, JoSAA can execute multiple runs of deferred acceptance when unfilled seats are eligible for de-reservation. After a run terminates, eligible unfilled capacity is transferred to the prescribed parent program-section, capacities are updated, and the allocation algorithm is run again. The JoSAA algorithmic analysis establishes

monotonicity across these de-reservation runs: the options generated for candidates who remain active are weakly expanded rather than contracted as eligible capacity moves to parent sections.

The empirical reconstruction does not separately simulate the internal de-reservation sequence. Instead, it uses the route-specific admission frontiers observed after the historical allocation process has operated. Consequently, de-reservation enters our opportunity measure through the realized frontiers generated by the mechanism rather than through a counterfactual reconstruction of each intermediate run.

**Preparatory-course routes.** Preparatory-course allocation follows separate institutional rules and uses dedicated preparatory merit lists. Under the IIT Business Rules, specified unfilled category or PwD seats can first be made available to candidates on the corresponding preparatory lists before subsequent de-reservation occurs. These routes therefore do not share the same rank–frontier mapping as the regular IIT routes used in the baseline reconstruction.

We exclude preparatory-course routes from the baseline analytical market. This exclusion is a comparability choice rather than a claim that preparatory allocations are economically unimportant. It ensures that the baseline opportunity sets are constructed from routes governed by the regular merit-list frontiers represented consistently in our candidate and OR–CR data.

**PwD routes.** PwD program-sections are governed by separate PwD merit lists and interact with preparatory-course and de-reservation provisions. Although these channels are part of the JoSAA allocation system, their priority architecture is not implemented comparably to the regular non-PwD routes in our baseline candidate-level reconstruction. We therefore exclude PwD routes from the principal analytical market.

This restriction should not be interpreted as treating PwD seats as part of the corresponding non-PwD route. They are distinct institutional channels and are removed rather than mechanically reassigned in the baseline reconstruction.

**Defence Service allocations.** Defence Service allocations constitute another special route. Their historical implementation differs from the regular nested category architecture. The JoSAA algorithmic documentation records preferential DS allocation under the 2015 rules and a shift to supernumerary DS treatment from 2016 onward. Later Business Rules continue to treat DS as a separate preferential allocation channel.

Because this allocation cannot be represented as an ordinary program–category cutoff route comparable to OPEN, OBC-NCL, SC, ST, or GEN-EWS, Defence Service allocations are excluded from the baseline rank-frontier reconstruction. Proposition 1 therefore makes no claim about the DS allocation procedure.

**Other special or unresolved routes.** The baseline reconstruction also excludes malformed, unresolved, or other special allocation records for which candidate eligibility, governing merit

list, and the corresponding published frontier cannot be mapped consistently. The objective is to maintain a common interpretation of a retained route: candidate $i$ is eligible for route $r$, $q_{irt}$ is the rank governing that route, and the observed closing rank provides the relevant institutional frontier.

We do not convert excluded special channels into regular routes or infer counterfactual eligibility from the candidate's realized seat category.

**Mapping to the empirical opportunity sets.** These institutional choices generate the hierarchy used throughout the paper. The formal object

$$Q_{it}(\boldsymbol{\kappa}_t)$$

describes cutoff-qualified institutional opportunity in a fixed regular program–route environment. The direct empirical reconstruction

$$Q_{it}^{D}$$

implements the same route-level qualification logic using the historical JoSAA OR–CR records and the most permissive observed frontier across rounds. The final empirical menu

$$F_{it}$$

adds the administratively observed allocated program only when that program is absent from $Q_{it}^{D}$.

The revealed augmentation is deliberately minimal. Among the 119,285 candidate menus in the principal sample, the observed allocated program is recovered directly from $Q_{it}^{D}$ for 118,567 candidates, or 99.40 percent. Only 718 candidate menus, 0.60 percent, require the allocated program to be added. These cases demonstrate that the published OR–CR frontier is not a sufficient statistic for every feature of the historical MRDA implementation; they do not identify which particular administrative rule generated each discrepancy.

Accordingly, we do not use the augmentation to infer additional unobserved opportunities. For an augmented candidate, only the documented realized allocation is added. Every other academic program in $F_{it}$ must be generated by the regular route-specific frontier reconstruction described in Section 3.3.

**Table A6.** JoSAA institutional features and baseline analytical treatment

| Institutional feature | Mechanism role | Baseline analytical treatment |
|---|---|---|
| Regular OPEN and category routes | Separate merit lists, capacities, and program-sections generate route-specific priority frontiers. | Included. Qualification is evaluated route by route and collapsed to the physical academic-program level. |
| Female-Only and Gender-Neutral pools | From 2018, eligible female candidates can compete through both pools under the prescribed virtual-route ordering. | Included when the relevant pool is administratively identified. Pool-specific frontiers enter the route-level reconstruction. |
| GEN-EWS | Introduced in 2019 as an additional category-specific priority route. | Included from 2019 using the corresponding candidate rank and program-section frontier. |
| Multiple allocation rounds | Candidate actions and administrative updates alter the active allocation environment across rounds. | Included through the any-round reconstruction using the most permissive observed route-specific frontier. |
| Min-Cutoff | Protects eligible continuing candidates across rounds and can generate allocations not represented by a simple contemporaneous cutoff comparison. | Not simulated directly. Residual cases are handled only through the minimal revealed-feasible augmentation when the documented allocation is absent from $Q_{it}^{D}$. |
| De-reservation | Unfilled eligible capacity is transferred to designated parent program-sections across repeated DA runs. | Not simulated run by run. Its realized consequences are incorporated through the observed route-specific OR–CR frontiers. |
| Preparatory courses | Use separate preparatory merit lists and interact with subsequent de-reservation. | Excluded from the baseline reconstruction. |
| PwD routes | Use separate PwD merit lists and interact with preparatory and de-reservation rules. | Excluded from the baseline reconstruction. |
| Defence Service | Special preferential/supernumerary allocation channel outside the regular nested route structure. | Excluded from the baseline reconstruction. |
| Malformed, unresolved, or other noncomparable routes | No consistently recoverable regular merit-list/frontier mapping. | Excluded rather than reassigned to a regular route. |

*Notes:* The table summarizes the institutional features relevant to the baseline IIT opportunity-set reconstruction. $Q_{it}^{D}$ denotes the direct empirical route-level reconstruction and $F_{it}$ the final empirical menu after minimal revealed augmentation. Exclusion of a special route indicates that it does not enter the baseline regular-market reconstruction; it does not imply that the corresponding institutional provision is unimportant.

## A.2 Proof of Proposition 1

We prove the result for a fixed regular DA run. The argument applies after the institution has constructed the virtual programs, virtual merit lists, and virtual preference lists relevant to that run. The purpose of the proposition is not to establish the correctness of the full JoSAA algorithm, which is proved in the algorithmic literature, but to establish the particular cutoff-based implication used in our empirical reconstruction.

For candidate $i$, let

$$\widetilde{\mathcal{L}}_i$$

denote the virtual preference list generated from her active physical-program list $\mathcal{L}_i$. By assumption, the virtual-route construction preserves the physical-program ordering: all eligible virtual routes corresponding to a physical program $j$ are considered before the algorithm proceeds to a physical program $k$ whenever

$$j \succ_i k.$$

Let $r \in \mathcal{R}_{jt}$ denote a virtual route associated with physical program $j$. Candidate $i$ is eligible for this route when $e_{ir} = 1$. Let $q_{ir}$ denote her position on the merit list relevant to $r$, with lower values representing stronger priority.

At termination of the run, define the route closing rank by

$$\kappa_r = \max_{h:\, \mu_h^r = r} q_{hr}$$

when route $r$ contains at least one allocated candidate. If route $r$ has unfilled capacity, set

$$\kappa_r = +\infty.$$

When several candidates share the closing rank, JoSAA's equal-rank rule creates the required supernumerary capacity rather than retaining only an arbitrary subset of candidates tied at that rank. Consequently,

$$q_{ir} \leq \kappa_r$$

is the appropriate weak qualification condition.

We now prove the proposition by contradiction.

Suppose candidate $i$ is assigned physical program $\mu_i$, and suppose there exists a physical program

$$j \in Q_i(\kappa) \cap \mathcal{L}_i$$

such that

$$j \succ_i \mu_i.$$

Because $j \in Q_i(\kappa)$, there exists at least one route

$$r \in \mathcal{R}_{jt}$$

such that

$$e_{ir} = 1 \qquad \text{and} \qquad q_{ir} \leq \kappa_r.$$

Because $j \in \mathcal{L}_i$ and $j \succ_i \mu_i$, the preference-preserving virtual refinement places the eligible virtual routes associated with $j$ before the routes associated with $\mu_i$. Candidate-proposing DA must therefore process candidate $i$ through the eligible routes associated with $j$ before she can proceed to the route that ultimately yields $\mu_i$.

Candidate $i$ does not receive physical program $j$. Hence she cannot be permanently retained by route $r$. There are two possible cases.

**Case 1: Route $r$ has unfilled capacity at termination.** Then

$$\kappa_r = +\infty.$$

An eligible candidate who applies to a route with available capacity is retained under DA. Candidate $i$ therefore could not have been rejected from route $r$. This contradicts the fact that the algorithm subsequently proceeds to a less-preferred physical program $\mu_i$.

**Case 2: Route $r$ is filled at termination.** Let $\kappa_r$ be the priority of the weakest candidate retained by route $r$. Under candidate-proposing DA, a candidate can be rejected from a filled route only when the candidates retained by that route have strictly stronger priority, subject to the institutional equal-rank rule.

Candidate $i$ satisfies

$$q_{ir} \leq \kappa_r.$$

If

$$q_{ir} < \kappa_r,$$

candidate $i$ has strictly stronger priority than at least the weakest candidate retained at $r$, so she cannot be rejected while that weaker candidate remains assigned.

If instead

$$q_{ir} = \kappa_r,$$

candidate $i$ is tied at the closing rank. JoSAA's multiple-candidate same-rank rule prevents the mechanism from rejecting only some candidates at that closing priority merely because the nominal route capacity is exhausted; the necessary additional seats are created on a supernumerary basis. Thus candidate $i$ again cannot be rejected solely because of capacity while an equal-priority candidate remains assigned.

Both cases contradict the supposition that candidate $i$ passes through route $r$ and is eventually assigned the strictly less-preferred physical program $\mu_i$.

Therefore no such $j$ exists, and

$$\mu_i \succeq_i j \qquad \forall j \in Q_i(\kappa) \cap \mathcal{L}_i.$$

This proves Proposition 1. □

**Scope of the proposition.** Proposition 1 is intentionally stated for a fixed regular DA run. JoSAA's full implementation is richer. Across runs within a round, de-reservation can change the capacities of virtual program-sections. Across rounds, candidate actions and document verification can alter active preference lists and eligibility, while Min-Cutoff rules protect previously obtained allocations. Special channels such as preparatory courses and the DS common quota introduce additional institutional rules.

The proposition therefore provides a route-level justification for using published admission frontiers to measure institutional qualification; it is not a claim that the complete multi-run, multi-round historical mechanism is equivalent to one static DA market with fixed capacities.

Our baseline empirical sample excludes the special allocation channels for which this regular route interpretation cannot be implemented comparably. For the remaining regular routes, the reconstruction takes the union of the observed qualification conditions across eligible routes and observed allocation rounds. We separately verify in the data that the resulting direct frontier contains the administratively observed assignment for 99.4 percent of the final candidate sample.

### A.3 Proof of Lemma 1

*Proof.* Fix admission cycle $t$ and the program–route cutoff environment $\boldsymbol{\kappa}_t = \{\kappa_{jrt}\}_{j,r}$. Suppose candidate $i$ routewise dominates candidate $k$, so that

$$\mathcal{E}_{kt} \subseteq \mathcal{E}_{it}$$

and

$$q_{irt} \leq q_{krt} \qquad \forall r \in \mathcal{E}_{kt}.$$

Take an arbitrary physical academic program

$$j \in Q_{kt}(\boldsymbol{\kappa}_t).$$

By the definition of the cutoff-qualified institutional opportunity set $Q_{kt}(\boldsymbol{\kappa}_t)$, there exists at least one route $r \in \mathcal{R}_{jt}$ such that candidate $k$ is eligible for that route and clears the corresponding program–route cutoff:

$$e_{krt} = 1 \qquad \text{and} \qquad q_{krt} \leq \kappa_{jrt}. \tag{A.1}$$

Because $e_{krt} = 1$,

$$r \in \mathcal{E}_{kt}.$$

Eligibility dominance, $\mathcal{E}_{kt} \subseteq \mathcal{E}_{it}$, therefore implies

$$r \in \mathcal{E}_{it},$$

and hence

$$e_{irt} = 1.$$

Routewise priority dominance further gives

$$q_{irt} \leq q_{krt}.$$

Combining this inequality with (A.1) yields

$$q_{irt} \leq q_{krt} \leq \kappa_{jrt}.$$

Candidate $i$ therefore also clears the cutoff for route $r$ to physical academic program $j$. Since $r \in \mathcal{R}_{jt}$, the definition of $Q_{it}(\boldsymbol{\kappa}_t)$ implies

$$j \in Q_{it}(\boldsymbol{\kappa}_t).$$

Because $j$ was an arbitrary element of $Q_{kt}(\boldsymbol{\kappa}_t)$, it follows that

$$Q_{kt}(\boldsymbol{\kappa}_t) \subseteq Q_{it}(\boldsymbol{\kappa}_t).$$

□

**Remark on scalar priority measures.** Lemma 1 concerns the complete route-specific eligibility and priority information entering the institutional opportunity set. It does not imply that a scalar empirical summary of priority is sufficient to order candidates' opportunity sets.

In particular, the predetermined P1 measure introduced in Section 4.5 summarizes a candidate's position using AIR/CRL for OPEN candidates and the corresponding own-category merit-list rank for reserved-category candidates. P1 is useful for organizing empirical comparisons without conditioning on realized allocation information, but it does not encode every route through which a candidate may compete.

Accordingly, the empirical relationship between P1 and reconstructed opportunities is an object to be measured rather than a mechanical consequence of Lemma 1. The alternative-priority exercises reported in the appendix assess the sensitivity of the main results to this scalar summarization.

## A.4 Derivation of the decomposition identity and common-choice reproduction shares

This appendix establishes the accounting identities and benchmark-dependence properties stated in the decomposition identity in Remark 1 and Remark 2. Recall that $H$ and $L$ denote the highest- and lowest-predetermined-priority groups, respectively, and that the empirical decomposition pools candidate–cycle observations across admission cycles. For program outcome $y$,

$$Y_{it}^{y} = y_{a_{it}t}$$

is the outcome associated with candidate $i$'s observed allocation in admission cycle $t$, $B_y(F_{it})$ is the equal-weight opportunity-level benchmark defined over the reconstructed menu $F_{it}$, and

$$P^{c} = \left\{P_t^{c}\right\}_t$$

denotes the collection of admission-cycle-specific common-choice operators. The corresponding common-choice outcome for candidate $i$ in admission cycle $t$ is $C_y(F_{it}; P_t^{c})$.

*Derivation of the decomposition identity.* By definition, the observed allocation gap and the opportunity-level component are

$$\Delta_Y^{y} = \mathrm{E}\left[Y_{it}^{y} \mid (i,t) \in H\right] - \mathrm{E}\left[Y_{it}^{y} \mid (i,t) \in L\right], \tag{A.2}$$

$$\Delta_O^{y} = \mathrm{E}\left[B_y(F_{it}) \mid (i,t) \in H\right] - \mathrm{E}\left[B_y(F_{it}) \mid (i,t) \in L\right]. \tag{A.3}$$

The within-menu sorting gap is defined as

$$\Delta_W^{y} = \Delta_Y^{y} - \Delta_O^{y}. \tag{A.4}$$

Therefore,

$$\Delta_Y^{y} = \Delta_O^{y} + \Delta_W^{y}, \tag{A.5}$$

which establishes the first accounting decomposition.

For an admissible common-choice operator $P^c = \{P_t^c\}_t$, define

$$\Delta_C^y(P^c) = \mathrm{E}\big[C_y(F_{it}; P_t^c) \mid (i,t) \in H\big] - \mathrm{E}\big[C_y(F_{it}; P_t^c) \mid (i,t) \in L\big]. \tag{A.6}$$

The menu-structure component is

$$\Delta_S^y(P^c) = \Delta_C^y(P^c) - \Delta_O^y, \tag{A.7}$$

and the residual rank-associated choice component is

$$\Delta_R^y(P^c) = \Delta_Y^y - \Delta_C^y(P^c). \tag{A.8}$$

Adding the opportunity-level, menu-structure, and residual components gives

$$\begin{aligned}\Delta_O^y + \Delta_S^y(P^c) + \Delta_R^y(P^c) &= \Delta_O^y + \big[\Delta_C^y(P^c) - \Delta_O^y\big] + \big[\Delta_Y^y - \Delta_C^y(P^c)\big] \\ &= \Delta_Y^y. \end{aligned} \tag{A.9}$$

Hence,

$$\Delta_Y^y = \Delta_O^y + \Delta_S^y(P^c) + \Delta_R^y(P^c), \tag{A.10}$$

which proves equation (4.32).

Combining this identity with $\Delta_W^y = \Delta_Y^y - \Delta_O^y$ yields

$$\begin{aligned}\Delta_W^y &= \Delta_Y^y - \Delta_O^y \\ &= \Delta_S^y(P^c) + \Delta_R^y(P^c), \end{aligned} \tag{A.11}$$

which proves equation (4.33).

It remains to establish the invariance and benchmark-dependence statements. The quantities $\Delta_Y^y$ and $\Delta_O^y$ depend only on observed allocations, reconstructed menus, the focal program outcome, and the definitions of the comparison groups. Neither quantity contains the common-choice operator $P^c$. Because

$$\Delta_W^y = \Delta_Y^y - \Delta_O^y,$$

the within-menu sorting gap is likewise invariant to the choice of $P^c$.

Now consider two admissible collections of admission-cycle-specific common-choice operators, $P^c = \{P_t^c\}_t$ and $\widetilde{P}^c = \{\widetilde{P}_t^c\}_t$. Using the definition of the menu-structure component,

$$\begin{aligned}\Delta_S^y(P^c) - \Delta_S^y(\widetilde{P}^c) &= \left[\Delta_C^y(P^c) - \Delta_O^y\right] - \left[\Delta_C^y(\widetilde{P}^c) - \Delta_O^y\right] \\ &= \Delta_C^y(P^c) - \Delta_C^y(\widetilde{P}^c). \end{aligned} \tag{A.12}$$

This establishes equation (4.34).

Similarly, from the definition of the residual component,

$$\begin{aligned}\Delta_R^y(P^c) - \Delta_R^y(\widetilde{P}^c) &= \left[\Delta_Y^y - \Delta_C^y(P^c)\right] - \left[\Delta_Y^y - \Delta_C^y(\widetilde{P}^c)\right] \\ &= -\left[\Delta_C^y(P^c) - \Delta_C^y(\widetilde{P}^c)\right] \\ &= -\left[\Delta_S^y(P^c) - \Delta_S^y(\widetilde{P}^c)\right]. \end{aligned} \tag{A.13}$$

Thus, changing the common-choice benchmark reallocates the fixed within-menu sorting gap one-for-one between the menu-structure and residual components, while leaving $\Delta_Y^y$, $\Delta_O^y$, and $\Delta_W^y$ unchanged.

□

*Derivation of the common-choice reproduction shares.* From the definition of the menu-structure component,

$$\Delta_S^y(P^c) = \Delta_C^y(P^c) - \Delta_O^y,$$

so that

$$\Delta_C^y(P^c) = \Delta_O^y + \Delta_S^y(P^c). \tag{A.14}$$

When $\Delta_Y^y \neq 0$, dividing equation (A.14) by the observed allocation gap gives

$$\frac{\Delta_C^y(P^c)}{\Delta_Y^y} = \frac{\Delta_O^y + \Delta_S^y(P^c)}{\Delta_Y^y}. \tag{A.15}$$

the decomposition identity in Remark 1 also implies

$$\Delta_Y^y = \Delta_C^y(P^c) + \Delta_R^y(P^c),$$

because

$$\Delta_C^y(P^c) = \Delta_O^y + \Delta_S^y(P^c).$$

Therefore,

$$\frac{\Delta_C^y(P^c)}{\Delta_Y^y} = 1 - \frac{\Delta_R^y(P^c)}{\Delta_Y^y}. \tag{A.16}$$

Combining equations (A.15) and (A.16) yields

$$\mathcal{A}_y(P^c) = \frac{\Delta_C^y(P^c)}{\Delta_Y^y} = \frac{\Delta_O^y + \Delta_S^y(P^c)}{\Delta_Y^y} = 1 - \frac{\Delta_R^y(P^c)}{\Delta_Y^y},$$

which proves equation (4.36).

For the second result, the decomposition identity in Remark 1 gives

$$\Delta_W^y = \Delta_S^y(P^c) + \Delta_R^y(P^c).$$

When $\Delta_W^y \neq 0$, dividing through by the within-menu sorting gap produces

$$1 = \frac{\Delta_S^y(P^c)}{\Delta_W^y} + \frac{\Delta_R^y(P^c)}{\Delta_W^y}.$$

Rearranging,

$$\mathcal{M}_y(P^c) = \frac{\Delta_S^y(P^c)}{\Delta_W^y} = 1 - \frac{\Delta_R^y(P^c)}{\Delta_W^y},$$

which proves equation (4.37).

Neither ratio is mechanically bounded between zero and one. For example, $\Delta_S^y(P^c)$ can exceed $\Delta_W^y$ if the residual component has the opposite sign, and the common-choice prediction can exceed the observed gap if $\Delta_R^y(P^c)$ is negative. The terminology "share" therefore describes the decomposition ratio rather than imposing a probability interpretation.

□

# B Choice Model, Support, and Behavioral Robustness

This appendix reports the specification, support, and inference checks underlying the choice estimates in Section 4.3 and the priority gradient in Section 5.2. The exercises address four questions in turn: whether the estimated attribute trade-offs depend on which attributes and fixed effects enter the model, whether the priority gradient depends on the compensation support available within candidates' menus, whether the reported inference depends on the clustering convention, and whether the results depend on the independence of irrelevant alternatives.

Table A7 reports a specification ladder under both the conditional-choice model estimated by PPML and a linear probability model. The distance coefficient is stable across all four columns and both estimators, moving only from −0.126 to −0.135 per 100 kilometers. The compensation trade-off is not stable in the same way. It falls from 111.7 kilometers per lakh with distance and CTC alone, to 97.7 in the full-attribute model, to 36.8 once academic-program fixed effects are included, and the placement and NIRF trade-offs become statistically indistinguishable from zero in that final column. The pattern is expected, because program fixed effects absorb the across-program variation in attributes that the earlier columns use, leaving only within-program variation across cycles. We report the full-attribute specification as the baseline and read all attribute coefficients as revealed-choice associations rather than as causal willingness-to-pay parameters, consistent with Section 4.4.

Table A8 restricts the candidate–program support. A concern about the priority gradient in compensation trade-offs is that lower-priority candidates may face menus with little compensation variation, so that their estimated trade-off is identified from thin support. Panel A addresses this by restricting to candidates whose menus have above-median within-decile compensation dispersion, which retains 50.1 percent of choice sets, and by restricting to alternatives with observed compensation. The top-minus-bottom difference is 24.97 kilometers per lakh in the full sample, 24.91 under the within-decile dispersion restriction, and 36.79 under the pooled dispersion restriction. The sign and approximate magnitude are therefore not produced by differential support, though none of the three differences is precisely estimated.

Table A9 compares candidate-clustered and two-way clustered standard errors for the average compensation trade-off. Two-way clustering by candidate choice set and academic-program-by-admission-year produces standard errors between 3.6 and 7.3 times larger than candidate-clustering alone. The larger dimension is the program-by-year dimension, which candidate-clustering ignores entirely. We report two-way clustered inference throughout the paper.

Table A10 relaxes the substitution restriction using an IIT-nested specification. The estimated inclusive-value parameter is 0.430 and is distinguishable from one, so within-IIT substitution is substantial and the conditional-logit restriction is rejected. Attribute trade-offs change in interpretable directions: the compensation trade-off falls from 97.7 to 81.0 kilometers per lakh and the placement trade-off from 19.6 to 14.0, while the NIRF trade-off rises from 16.2 to 23.2. The priority pattern survives, with a top-minus-bottom difference of 41.54 rather than 24.97 kilometers per lakh. Because the sequential nested estimator does not deliver the cross-stage covariance

required for inference on the combined difference, we treat this as a point-estimate exercise.

**Table A7.** Choice-Model Specification Ladder and Alternative Estimator

| | (1) | (2) | (3) | (4) |
|---|---|---|---|---|
| | Distance only | + CTC | Full attributes | + Program FE |
| *Panel A. Conditional-choice model (PPML)* | | | | |
| Distance (100 km) | −0.126*** | −0.134*** | −0.133*** | −0.135*** |
| | (0.003) | (0.003) | (0.003) | (0.002) |
| Median CTC (Rs. 10 lakh, 2023 prices) | | 1.499*** | 1.303*** | 0.498*** |
| | | (0.058) | (0.053) | (0.048) |
| Placement rate (10 pp) | | | 0.261*** | 0.010 |
| | | | (0.019) | (0.010) |
| NIRF score | | | 0.022*** | −0.002 |
| | | | (0.002) | (0.003) |
| *Implied distance trade-offs* | | | | |
| CTC (km per Rs. 1 lakh) | | 111.7 | 97.7 | 36.8 |
| | | [102.0, 121.4] | [88.7, 106.7] | [29.7, 43.8] |
| Placement (km per percentage point) | | | 19.6 | 0.8 |
| | | | [16.6, 22.6] | [−0.8, 2.3] |
| NIRF (km per score point) | | | 16.2 | −1.6 |
| | | | [12.5, 19.9] | [−6.0, 2.9] |
| *Panel B. Linear probability model robustness* | | | | |
| Distance (100 km) | −0.000769*** | −0.000791*** | −0.000806*** | −0.000869*** |
| | (0.000024) | (0.000025) | (0.000024) | (0.000023) |
| Median CTC (Rs. 10 lakh, 2023 prices) | | 0.012413*** | 0.010726*** | 0.001331*** |
| | | (0.000581) | (0.000590) | (0.000192) |
| Placement rate (10 pp) | | | 0.001332*** | −0.000025 |
| | | | (0.000101) | (0.000023) |
| NIRF score | | | 0.000083*** | −0.000011 |
| | | | (0.000014) | (0.000010) |
| *Implied distance trade-offs* | | | | |
| CTC (km per Rs. 1 lakh) | | 156.9 | 133.1 | 15.3 |
| | | [143.0, 170.8] | [118.7, 147.5] | [11.0, 19.6] |
| Placement (km per percentage point) | | | 16.5 | −0.3 |
| | | | [14.0, 19.1] | [−0.8, 0.2] |
| NIRF (km per score point) | | | 10.3 | −1.3 |
| | | | [7.0, 13.6] | [−3.5, 0.9] |
| *Panel C. Specification and sample* | | | | |
| No-graduated-batch indicator | No | Yes | Yes | Yes |
| NIRF-missing indicator | No | No | Yes | Yes |
| Candidate choice-set fixed effects | Yes | Yes | Yes | Yes |
| Academic-program fixed effects | No | No | No | Yes |
| Two-way clustered standard errors | Yes | Yes | Yes | Yes |
| Candidate–program alternatives | 18,629,219 | 18,629,219 | 18,629,219 | 18,629,219 |
| Candidate choice sets | 119,259 | 119,259 | 119,259 | 119,259 |
| Program–year clusters | 2,291 | 2,291 | 2,291 | 2,291 |

*Notes:* Panel A reports the conditional-choice specification ladder estimated by PPML with candidate choice-set fixed effects; Panel B estimates the same specifications by linear probability model. Distance is measured in 100 km, CTC in Rs. 10 lakh of annual median compensation in 2023 prices, placement in 10-percentage-point units, and NIRF in score points. Column (2) adds CTC and the no-graduated-batch indicator; column (3) adds placement, NIRF, and the NIRF-missing indicator; column (4) additionally includes academic-program fixed effects. Standard errors in parentheses are two-way clustered by candidate choice set and academic-program-by-admission-year. Implied distance trade-offs are coefficient ratios relative to distance; brackets report 95% delta-method confidence intervals. These ratios place the two estimators in common distance-equivalent units; the linear probability model is used as a functional-form robustness check rather than as a structural discrete-choice model. $^{***}p < 0.01$, $^{**}p < 0.05$, $^{*}p < 0.10$.

**Table A8.** Robustness of the Priority Gradient in CTC–Distance Trade-offs

| Specification | CTC–distance trade-off | | | Top-minus-bottom difference | | Choice sets (retained share) |
|---|---|---|---|---|---|---|
| | Top priority | Bottom priority | Difference | 95% CI | $p$-value | |
| *Panel A. Feasible-support and observed-CTC restrictions* | | | | | | |
| Full sample *(preferred)* | 103.64 | 78.67 | 24.97 | [-4.22, 54.16] | 0.094 | 119,258 (100.0%) |
| High within-P1-decile CTC variation | 96.60 | 71.69 | 24.91 | [-8.32, 58.14] | 0.142 | 59,701 (50.1%) |
| High CTC variation, pooled cutoff | 103.42 | 66.63 | 36.79 | [-8.76, 82.33] | 0.113 | 59,606 (50.0%) |
| Observed CTC only | 102.02 | 48.03 | 54.00 | [26.11, 81.89] | $< 0.001$ | 110,681 (92.8%) |
| Observed CTC + within-P1-decile variation | 96.45 | 42.82 | 53.63 | [22.14, 85.12] | $< 0.001$ | 56,155 (47.1%) |
| *Panel B. Exact prior-year information* | | | | | | |
| Contemporaneous attributes on exact $t-1$ sample | 131.30 | 77.54 | 53.77 | [23.09, 84.45] | $< 0.001$ | 109,601 |
| Exact $t-1$ attributes | 104.34 | 82.83 | 21.51 | [-10.07, 53.09] | 0.182 | 109,601 |
| *Panel C. Prior-year-cutoff reconstructed menus* | | | | | | |
| Strict prior-year-cutoff menu | 127.02 | 72.58 | 54.43 | [23.20, 85.67] | $< 0.001$ | 66,294 |
| Revealed-augmented prior-year-cutoff menu | 131.24 | 85.21 | 46.03 | [7.91, 84.15] | 0.018 | 107,685 |
| *Panel D. Substitution-model robustness* | | | | | | |
| Conditional logit / PPML benchmark | 103.64 | 78.67 | 24.97 | [-4.22, 54.16] | 0.094 | 119,258 |
| IIT-nested logit P1 | 95.52 | 53.98 | 41.54 | – | – | – |

*Notes:* Entries report CTC–distance trade-offs in kilometers per Rs. 1 lakh of annual program CTC. All priority groups use the predetermined own-list priority measure (P1). The top and bottom columns correspond to the highest and lowest P1 deciles, respectively.

Panel A applies pre-specified restrictions on feasible-set CTC support and observed CTC; retained shares are relative to the preferred full P1 choice-model sample. Panel B compares contemporaneous program attributes with exact prior-year ($t-1$) information on the common 2016–2023 choice-model sample. Panel C reconstructs feasible menus using prior-year admission cutoffs. Panel D compares the baseline conditional-logit/PPML specification with the two-stage IIT-nested-logit specification.

Confidence intervals and $p$-values refer to the top-minus-bottom difference. For the IIT-nested-logit specification, the combined trade-off is reported as a point-estimate robustness because the sequential estimator does not provide the cross-stage covariance required for valid inference on the combined difference; the corresponding confidence interval and $p$-value are therefore not reported. The estimated nested inclusive-value parameter is 0.430 (SE = 0.028). The reported priority gradients describe revealed-choice heterogeneity and are not interpreted as causal effects of priority on preferences.

**Table A9.** Sensitivity of Average CTC Trade-Off Inference to the Clustering Convention

*Panel A. Distance–CTC trade-offs*

| Specification | Estimate | Candidate-clustered SE [95% CI] | Two-way clustered SE [95% CI] | SE ratio |
|---|---|---|---|---|
| Preferred full-attribute model | 97.7 | 0.66 [96.4, 99.0] | 4.59 [88.7, 106.7] | 6.99 |
| Full attributes + program FE | 36.8 | 0.99 [34.8, 38.7] | 3.60 [29.7, 43.8] | 3.62 |
| Exact $t-1$ full-attribute model | 99.1 | 0.68 [97.8, 100.4] | 4.95 [89.4, 108.8] | 7.27 |

*Panel B. Cluster dimensions*

| Sample | Candidate clusters | Program–year clusters |
|---|---|---|
| Main choice sample, 2015–2023 | 119,259 | 2,291 |
| Exact $t-1$ information sample, 2016–2023 | 109,601 | 2,075 |

*Notes:* The table compares covariance estimators while holding choice-model point estimates fixed. Candidate-clustered inference allows arbitrary dependence within candidate choice sets; two-way inference additionally clusters by academic-program-by-admission-year. Panel A reports implied distance–CTC trade-offs in kilometers per Rs. 1 lakh of annual program median CTC in 2023 prices. The preferred full-attribute and program-FE specifications use the 2015–2023 main choice sample; the exact $t-1$ specification uses the 2016–2023 prior-information sample. Each inference cell reports the delta-method standard error with the 95% confidence interval in brackets; the SE ratio is the two-way standard error divided by the candidate-clustered standard error. The paper therefore uses two-way clustering by candidate choice set and program–year as its preferred inference convention.

**Table A10.** IIT-Nested Logit Robustness

*Panel A. Average distance-equivalent program trade-offs*

| Program attribute | Conditional logit | IIT-nested logit | Nested / conditional |
|---|---|---|---|
| CTC (km per Rs. 1 lakh) | 97.7 | 81.0 | 0.83 |
| Placement (km per percentage point) | 19.6 | 14.0 | 0.72 |
| NIRF (km per score point) | 16.2 | 23.2 | 1.43 |

*Panel B. Priority-associated CTC–distance trade-off*

| | | | |
|---|---|---|---|
| Top P1 decile | 103.64 | 95.52 | 0.92 |
| Bottom P1 decile | 78.67 | 53.98 | 0.69 |
| Top-minus-bottom | 24.97 | 41.54 | 1.66 |

*Panel C. Inclusive-value parameter*

| Nested specification | $\lambda$ | SE | 95% CI | $p$-value: $\lambda = 1$ |
|---|---|---|---|---|
| Average-trade-off specification | 0.506 | 0.028 | [0.452, 0.560] | $< 0.001$ |
| P1 priority-varying specification | 0.430 | 0.028 | [0.376, 0.484] | $< 0.001$ |

*Panel D. Nest structure and support*

| Nest feature | Value |
|---|---|
| Candidate–IIT nests | 2,479,238 |
| Mean feasible IITs per candidate | 20.8 |
| Share with at least two feasible IIT nests | 100.0% |
| Lower-stage informative choice sets | 113,401 |
| Upper-stage informative choice sets | 119,242 |
| Program–year clusters | 2,291 |
| IIT–year clusters | 203 |

*Notes:* The table evaluates sensitivity to the conditional logit's independence-of-irrelevant-alternatives assumption using a sequential IIT-nested logit, with academic programs nested within IITs. At the lower stage, candidates choose among programs within an IIT using CTC, placement, and the no-graduated-batch indicator; at the upper stage, they choose among feasible IITs using distance, NIRF, NIRF missingness, and the lower-stage inclusive value. Distance is constant within a candidate–IIT nest.

Panels A and B report distance-equivalent point estimates. In the nested model, lower-stage coefficients in original utility units are recovered as $\beta = \lambda\alpha$, where $\lambda$ is the inclusive-value parameter. Lower-stage standard errors are two-way clustered by candidate choice set and program–year; upper-stage standard errors are two-way clustered by candidate choice set and IIT–year. Combined confidence intervals for nested trade-offs are not reported because the sequential estimator does not provide the cross-stage covariance required for valid inference.

Panel C reports $\lambda$ and tests $H_0 : \lambda = 1$; both estimates satisfy $0 < \lambda \leq 1$. The P1 priority-varying specification uses 119,258 choice sets and 18,629,216 candidate–program alternatives. The nested-logit exercise is a substitution-pattern robustness check and is not interpreted causally.

# C Opportunity–Structure–Choice Decomposition

This appendix reports the full set of decomposition results summarized in Section 5.5. Two properties of the decomposition organize the exercises. The first step, which separates the observed allocation gap into a difference in average feasible compensation and a remaining within-menu gap, is an accounting identity and requires no behavioral model. The second step, which divides the within-menu gap into a menu-structure component and a residual rank-associated component, requires a common-choice reference rule and is therefore benchmark-dependent. As established in Remark 1 and derived in Section A.4, changing the reference rule leaves the total gap, the opportunity-level component, and the within-menu gap unchanged, and reallocates only the within-menu gap between its two parts. The exercises below accordingly report the accounting components and the benchmark-dependent attribution separately.

Table A11 applies the decomposition to placement rates and NIRF scores rather than compensation. Table A12 varies the two constructions that the attribution depends on. Panel A varies the empirical definition of priority while holding the preferred reference rule fixed, and panel B varies the reference rule while holding priority groups and the candidate sample fixed. Panel B also reports top- and bottom-priority reference rules, which are included as deliberately extreme diagnostics of benchmark dependence and bracket the menu-structure share of the within-menu gap at 87.2 and 26.4 percent.

Table A13 reports parametric uncertainty for the model-dependent components from 2,000 draws of the year-specific common-choice coefficients, following Section 4.8. The observed sample, reconstructed menus, total gap, opportunity-level component, and within-menu gap are held fixed, so these intervals describe estimation uncertainty in the reference-rule parameters and are not confidence intervals for the accounting components.

Tables A14 through A17 address the concern that contemporaneous closing ranks are equilibrium outcomes rather than predetermined constraints, and may also postdate the information candidates held when submitting preferences, as discussed in Section 3.7. Table A14 reports the decomposition under exact immediately preceding-year program attributes and under two prior-year-frontier menu reconstructions. Table A16 reports coverage of exact prior-year information, which for compensation is 78.0 percent of feasible alternatives and 89.1 percent of chosen programs, and is lower for NIRF because the ranking does not exist for the earliest cycles. Table A17 reports the construction and sample retention for the two ex-ante designs. The strict design requires a candidate's allocated program to be predicted feasible from prior-year cutoffs and therefore retains 60.5 percent of the 2016–2023 candidate universe, while the revealed-augmented design retains 98.3 percent. Table A15 reports alternative admission periods and leave-one-cycle-out samples.

Across these exercises the opportunity-level share of the total gap stays between roughly 23 and 25 percent. Setting aside the two extreme reference diagnostics, the menu-structure share of the within-menu gap ranges from 57.2 to 71.6 percent, with the lower end produced by the AIR-only priority definition on a smaller candidate sample and the upper end by a leave-one-cycle-out sample. We therefore treat the preferred 68.5-percent share as a transparent benchmark-dependent

attribution, and the conclusion that menu structure accounts for the majority of the within-menu gap as the claim the evidence supports.

**Table A11.** Opportunity–Structure–Choice Decomposition for Alternative Program Outcomes

| | Placement<br>Percentage points | NIRF<br>Score points |
|---|---|---|
| *Panel A. Accounting decomposition* | | |
| Total chosen-program gap | 21.78 | 16.16 |
| Opportunity-level component | 10.04 | 6.12 |
| Within-menu sorting | 11.74 | 10.04 |
| *Panel B. Decomposition of within-menu sorting under common choice* | | |
| Menu structure under common choice | 6.90 | 3.31 |
| Residual rank-associated choice | 4.84 | 6.74 |
| *Panel C. Share of total chosen-program gap* | | |
| Opportunity level | 46.1% | 37.9% |
| Menu structure under common choice | 31.7% | 20.5% |
| Residual rank-associated choice | 22.2% | 41.7% |
| *Panel D. Share of within-menu gap* | | |
| Menu structure under common choice | 58.8% | 32.9% |
| Residual rank-associated choice | 41.2% | 67.1% |

*Panel E. Parametric uncertainty for model-dependent components*

| | Placement | | NIRF | |
|---|---|---|---|---|
| Measure | Estimate | 95% parametric simulation interval | Estimate | 95% parametric simulation interval |
| Menu-structure component | 6.90 | [6.32, 7.45] | 3.31 | [3.02, 3.55] |
| Residual rank-associated choice | 4.84 | [4.29, 5.42] | 6.74 | [6.50, 7.02] |
| Menu-structure share of total gap | 31.7% | [29.0%, 34.2%] | 20.5% | [18.7%, 21.9%] |
| Residual share of total gap | 22.2% | [19.7%, 24.9%] | 41.7% | [40.2%, 43.4%] |
| Menu-structure share of within-menu gap | 58.8% | [53.8%, 63.4%] | 32.9% | [30.1%, 35.3%] |

*Panel F. Analysis sample*

| Priority group | Candidates | Candidates |
|---|---|---|
| Top predetermined-priority decile | 11,806 | 10,959 |
| Bottom predetermined-priority decile | 9,788 | 9,478 |

*Notes:* The table applies the opportunity–structure–choice decomposition to placement rates and NIRF scores using the predetermined own-list priority measure (P1). For OPEN candidates, P1 uses AIR/CRL rank; for EWS, OBC-NCL, SC, and ST candidates, it uses the corresponding own-category merit-list rank. Priority deciles are constructed within admission year and candidate category.

Panel A decomposes the total top-minus-bottom outcome gap into an opportunity-level component and within-menu sorting. Panel B further decomposes within-menu sorting into menu structure under the same year-specific pooled common-choice rule used for CTC and residual rank-associated choice. The latter should not be interpreted as identified pure taste heterogeneity. Placement is measured in percentage points and NIRF in score points.

Panel E reports 95% parametric simulation intervals from 2,000 draws of the year-specific common-choice coefficients using the frozen two-way-clustered covariance matrices. The observed sample, reconstructed menus, P1 groups, total outcome gap, opportunity-level component, and within-menu gap are held fixed. The intervals therefore reflect common-choice parameter uncertainty conditional on these objects rather than comprehensive uncertainty for the accounting decomposition. Panel F reports outcome-specific candidate counts. Components may not sum exactly at the displayed precision because of rounding. The decomposition is accounting- and model-based and is not interpreted causally.

**Table A12.** Sensitivity of the Opportunity–Structure–Choice Decomposition

| Specification | Total gap | Opportunity level | Within-menu gap | Menu structure | Residual choice | Menu-structure share of within-menu gap |
|---|---|---|---|---|---|---|
| *Panel A. Priority-definition sensitivity; year-specific pooled common-choice rule fixed* | | | | | | |
| Legacy: year × allocated-seat category | 17.305 | 4.164 | 13.141 | 8.780 | 4.362 | 66.8% |
| P1: predetermined own-list priority *(preferred)* | 17.502 | 4.268 | 13.235 | 9.065 | 4.169 | 68.5% |
| P2: P1 stratified by AIR availability | 15.787 | 3.846 | 11.941 | 8.019 | 3.921 | 67.2% |
| P3: predetermined analytical-rank convention | 17.476 | 4.246 | 13.229 | 9.033 | 4.196 | 68.3% |
| P4: AIR-only common-list subsample | 13.441 | 2.671 | 10.770 | 6.166 | 4.604 | 57.2% |
| *Panel B. Common-choice benchmark sensitivity; P1 groups and candidate sample fixed* | | | | | | |
| Year-specific pooled *(preferred)* | 17.502 | 4.268 | 13.235 | 9.065 | 4.169 | 68.5% |
| Full-period pooled | 17.502 | 4.268 | 13.235 | 9.190 | 4.045 | 69.4% |
| Symmetric top/bottom probabilities | 17.502 | 4.268 | 13.235 | 7.702 | 5.532 | 58.2% |
| Top-priority reference | 17.502 | 4.268 | 13.235 | 11.536 | 1.699 | 87.2% |
| Bottom-priority reference | 17.502 | 4.268 | 13.235 | 3.498 | 9.737 | 26.4% |

*Notes:* All gaps and components are expressed in annual program median CTC (2023 LPA). Panel A varies the empirical priority definition while holding the preferred year-specific pooled common-choice rule fixed. P1 is the preferred predetermined own-list priority measure. Because the priority grouping and, for P4, the candidate sample can change across rows, the total, opportunity-level, and within-menu gaps need not be identical in Panel A.
Panel B holds P1 priority groups and the candidate sample fixed and varies only the common-choice benchmark. Consequently, the observed total gap, opportunity-level component, and within-menu gap are identical across Panel B rows by construction; only the decomposition of within-menu sorting into menu structure and residual rank-associated choice varies. The final column reports the menu-structure component as a share of the within-menu gap. The top- and bottom-priority reference rules are deliberately extreme reference diagnostics rather than co-equal preferred specifications. The decomposition is accounting- and model-based and is not interpreted causally.

**Table A13.** Parametric Uncertainty for the Preferred P1 Decomposition

| Measure | Estimate | Simulation SD | 95% parametric simulation interval |
|---|---|---|---|
| *Panel A. Fixed accounting objects* | | | |
| Observed chosen-program CTC gap | 17.502 | | Fixed |
| Opportunity-level component | 4.268 | | Fixed |
| Within-menu sorting | 13.235 | | Fixed |
| Opportunity share of total gap | 24.4% | | Fixed |
| *Panel B. Parametric uncertainty for model-dependent components* | | | |
| Common-choice predicted CTC gap | 13.333 | 0.405 | [12.571, 14.127] |
| Menu-structure component | 9.065 | 0.405 | [8.303, 9.859] |
| Residual rank-associated choice | 4.169 | 0.405 | [3.375, 4.931] |
| Menu-structure share of total gap | 51.8% | 2.3 pp | [47.4%, 56.3%] |
| Residual share of total gap | 23.8% | 2.3 pp | [19.3%, 28.2%] |
| Menu-structure share of within-menu gap | 68.5% | 3.1 pp | [62.7%, 74.5%] |
| Residual share of within-menu gap | 31.5% | 3.1 pp | [25.5%, 37.3%] |

*Notes:* The table reports parameter uncertainty for the preferred P1 decomposition of the top-minus-bottom chosen-program CTC gap. Panel A lists accounting objects held fixed in the simulation. Panel B reports results from 2,000 parametric draws of the year-specific common-choice coefficient vectors for admission cycles 2015–2023. Covariance matrices are two-way clustered by candidate choice set and academic-program-by-admission-year. The observed sample, reconstructed menus, total chosen-program CTC gap, opportunity-level component, and within-menu gap are held fixed. The intervals therefore reflect estimation uncertainty in the common-choice parameters conditional on these objects, rather than comprehensive uncertainty for the accounting decomposition. Simulation standard deviations for shares are reported in percentage points. CTC is annual program median compensation in lakh rupees per annum (LPA), expressed in 2023 prices. The decomposition is accounting- and model-based and is not interpreted causally.

**Table A14.** Decomposition Under Alternative Information and Ex-Ante Menu Reconstructions

| Specification | Candidates | | Share of total chosen-program CTC gap | | | Share of within-menu gap |
|---|---|---|---|---|---|---|
| | Top decile | Bottom decile | Opportunity level | Menu structure | Residual choice | Menu structure |
| *Panel A. Information timing and CTC support* | | | | | | |
| Preferred P1, contemporaneous reconstructed menus | 11,806 | 9,788 | 24.4% | 51.8% | 23.8% | 68.5% |
| Contemporaneous CTC, exact $t-1$ period/sample | 10,837 | 8,927 | 24.4% | 48.6% | 27.0% | 64.3% |
| Exact $t-1$ CTC | 10,732 | 8,259 | 25.0% | 52.8% | 22.2% | 70.4% |
| *Panel B. Prior-year-cutoff menu reconstruction* | | | | | | |
| Strict direct prior-year-cutoff menu | 9,233 | 2,895 | 23.3% | 50.8% | 25.8% | 66.3% |
| Revealed-augmented prior-year-cutoff menu | 10,837 | 7,619 | 23.9% | 50.9% | 25.2% | 66.9% |

*Notes:* All rows use the predetermined own-list priority measure (P1) and the preferred common-choice benchmark appropriate to the corresponding information or menu construction. The preferred row uses contemporaneous reconstructed menus. Panel A assesses sensitivity to the timing of program-information measurement. The exact-$t-1$ rows use admission cycles 2016–2023 and the available CTC support for the corresponding measure; because lagged CTC is not observed for all candidate–program alternatives, the contemporaneous and exact-$t-1$ rows should not be interpreted as a perfectly identical-candidate comparison.
Panel B reconstructs opportunity sets using prior-year admission cutoffs. The strict-direct specification retains candidates whose realized chosen program is predicted feasible under those cutoffs, while the revealed-augmented specification adds the realized chosen program when the strict prior-year menu omits it. These exercises assess sensitivity to contemporaneous realized cutoffs and are not equilibrium policy counterfactuals. Decomposition shares are relative to each row's observed top-minus-bottom chosen-program CTC gap. CTC is annual program median compensation in lakh rupees per annum (LPA), expressed in 2023 prices. The decomposition is accounting- and model-based and is not interpreted causally.

**Table A15.** Stability of the P1 Decomposition Across Admission Cycles

| Specification | Candidates | | Share of total chosen-program CTC gap | | | Share of within-menu gap |
|---|---|---|---|---|---|---|
| | Top decile | Bottom decile | Opportunity level | Menu structure | Residual choice | Menu structure |
| *Panel A. Preferred sample* | | | | | | |
| 2015–2023 *(preferred)* | 11,806 | 9,788 | 24.4% | 51.8% | 23.8% | 68.5% |
| *Panel B. Alternative admission periods* | | | | | | |
| 2016–2023 | 10,837 | 8,927 | 24.4% | 48.6% | 27.0% | 64.3% |
| 2018–2023 | 8,743 | 7,302 | 24.0% | 47.7% | 28.3% | 62.8% |
| *Panel C. Leave-one-admission-cycle-out* | | | | | | |
| Omit 2015 | 10,837 | 8,927 | 24.4% | 48.6% | 27.0% | 64.3% |
| Omit 2016 | 10,775 | 8,937 | 24.4% | 52.0% | 23.6% | 68.8% |
| Omit 2017 | 10,743 | 9,014 | 24.1% | 51.3% | 24.6% | 67.6% |
| Omit 2018 | 10,640 | 8,887 | 24.4% | 52.0% | 23.5% | 68.8% |
| Omit 2019 | 10,502 | 8,729 | 24.5% | 52.2% | 23.3% | 69.2% |
| Omit 2020 | 10,330 | 8,572 | 24.4% | 52.3% | 23.3% | 69.2% |
| Omit 2021 | 10,295 | 8,471 | 24.7% | 53.9% | 21.4% | 71.6% |
| Omit 2022 | 10,197 | 8,413 | 24.1% | 52.5% | 23.3% | 69.2% |
| Omit 2023 | 10,129 | 8,354 | 24.5% | 51.4% | 24.1% | 68.1% |

*Notes:* All rows use the predetermined own-list priority measure (P1) and the preferred year-specific pooled common-choice rule. Priority deciles are constructed within admission year and candidate category. Panel A reports the preferred 2015–2023 sample. Panel B restricts the decomposition to alternative admission periods. Panel C omits one admission cycle at a time. Decomposition shares are relative to each row's observed top-minus-bottom chosen-program CTC gap. The exercise assesses whether the decomposition is sensitive to the inclusion of particular admission cycles or periods; it is not used to select the preferred sample. CTC is annual program median compensation in lakh rupees per annum (LPA), expressed in 2023 prices. The decomposition is accounting- and model-based and is not interpreted causally.

**Table A16.** Coverage of Exact Prior-Year Program Information

*Panel A. Overall exact $t-1$ information coverage*

| Attribute | Feasible alternatives | | Chosen programs | |
|---|---|---|---|---|
| | Observed | Coverage | Observed | Coverage |
| Median CTC | 13,526,369 | 78.0% | 97,709 | 89.1% |
| Placement rate | 13,526,369 | 78.0% | 97,709 | 89.1% |
| NIRF score | 12,737,134 | 73.4% | 88,612 | 80.8% |

*Panel B. Exact $t-1$ coverage by admission cycle*

| | Feasible alternatives | | | Chosen programs | | |
|---|---|---|---|---|---|---|
| Admission cycle | CTC | Placement | NIRF | CTC | Placement | NIRF |
| 2016 | 78.1% | 78.1% | 0.0% | 89.4% | 89.4% | 0.0% |
| 2017 | 71.6% | 71.6% | 70.6% | 87.2% | 87.2% | 82.9% |
| 2018 | 74.7% | 74.7% | 85.9% | 87.5% | 87.5% | 92.1% |
| 2019 | 77.9% | 77.9% | 81.7% | 88.3% | 88.3% | 89.5% |
| 2020 | 83.8% | 83.8% | 82.7% | 92.3% | 92.3% | 90.0% |
| 2021 | 80.1% | 80.1% | 78.7% | 90.0% | 90.0% | 88.1% |
| 2022 | 80.7% | 80.7% | 81.9% | 90.1% | 90.1% | 90.6% |
| 2023 | 74.7% | 74.7% | 79.0% | 87.5% | 87.5% | 90.1% |

*Notes:* The table documents coverage of exact prior-year ($t-1$) program information in the 2016–2023 prior-information sample. Coverage is the share of feasible candidate–program alternatives, or realized chosen programs, for which the corresponding exact $t-1$ attribute is observed. Panel A reports overall coverage and Panel B reports coverage by admission cycle. The sample contains 17,346,233 candidate–program alternatives, 109,601 candidate choice sets, and 2,075 program–year cells. The 2015 cycle is excluded because the required 2014 program-information vintage is not used in this exercise. Exact prior-year NIRF is unavailable for the 2016 admission cycle because no 2015 NIRF vintage exists. CTC and placement coverage coincide because their prior-year program records have the same missingness pattern.

**Table A17.** Opportunity Sets from Prior-Year Closing Ranks

*Panel A. Frontier construction and sample retention*

| Construction step | Count | Share of universe |
|---|---|---|
| Candidate universe, 2016–2023 | 109,601 | 100.0% |
| Strict prior-year-cutoff menu nonempty | 107,695 | 98.3% |
| Chosen program predicted feasible from prior-year cutoffs | 66,304 | 60.5% |
| Strict-direct analysis sample, $K_i \geq 2$ | 66,294 | 60.5% |
| Revealed-augmented analysis sample, $K_i \geq 2$ | 107,685 | 98.3% |
| Stable-rules strict-direct sample | 66,043 | 60.3% |
| Stable-rules revealed-augmented sample | 105,304 | 96.1% |

*Panel B. Analysis-sample menu size by ex-ante design*

| Measure | Strict direct | Revealed augmented |
|---|---|---|
| Candidate choice sets | 66,294 | 107,685 |
| Prior-year-cutoff candidate–program alternatives | 10,558,468 | 15,175,994 |
| Current-menu alternatives, same candidates | 11,693,215 | 17,286,433 |
| Mean prior-year-cutoff menu size | 159.3 | 140.9 |
| Median prior-year-cutoff menu size | 171 | 149 |
| Candidate universe retained | 60.5% | 98.3% |

*Panel C. Opportunity-set size by admission cycle*

| | | Mean menu size | | | |
|---|---|---|---|---|---|
| Admission cycle | Choice sets | Current | Strict prior-year | Augmented prior-year | Chosen program predicted feasible (%) |
| 2016 | 10,107 | 145.4 | 121.6 | 121.9 | 67.3% |
| 2017 | 10,592 | 149.1 | 136.5 | 136.7 | 79.8% |
| 2018 | 11,358 | 154.3 | 138.8 | 139.2 | 59.9% |
| 2019 | 12,679 | 157.3 | 135.2 | 135.7 | 52.7% |
| 2020 | 14,747 | 154.9 | 133.9 | 134.5 | 42.6% |
| 2021 | 15,362 | 156.9 | 141.2 | 141.5 | 68.8% |
| 2022 | 15,972 | 169.1 | 149.8 | 150.2 | 67.0% |
| 2023 | 16,868 | 183.4 | 156.1 | 156.5 | 59.3% |

*Notes:* The table documents ex-ante feasible sets constructed from exact prior-year JoSAA closing ranks. For admission cycle $t$, a program enters the strict frontier if at least one eligible admission route has a relevant candidate rank no worse than the corresponding program–seat-category–pool closing rank in $t-1$. OPEN routes use AIR; reserved-category routes use the corresponding category rank. The strict-direct design retains candidates whose chosen program is predicted feasible and whose menu contains at least two programs. The revealed-augmented design restores the observed chosen program when the strict rule misses it and retains menus with $K_i \geq 2$. Female-Only pools begin in 2018 and EWS in 2019; stable-rules samples exclude candidates relying on newly introduced routes without prior-year cutoffs. Panel C uses the revealed-augmented sample. Program attributes and geography remain measured in the target admission year; only feasible-set construction changes. The 2015 cycle is excluded because 2014 cutoffs predate the unified JoSAA mechanism.

# D Distributional Heterogeneity

This appendix examines whether the revealed attribute trade-offs estimated in Section 4.3 differ systematically across social category, place of origin, and seat pool. The exercises are included primarily as uniformity checks on the common-choice benchmark defined in Section 4.6, which applies a single reference rule to every candidate within an admission cycle. If measured trade-offs differed sharply across groups, that benchmark would be harder to defend. Two data features constrain what can be estimated and require the exercise to be split across two samples.

The Economically Weaker Section quota took effect for admissions beginning in 2019 and redefined the OPEN base category, so a single social-category specification cannot span the full sample period. Table A18 therefore reports the five-category specification on 2019–2023, and Table A19 reports a four-category specification using the categories observed throughout 2015–2023. In the 2019–2023 sample, compensation trade-offs range from 77.1 kilometers per lakh for SC candidates to 101.6 for EWS candidates, and the OPEN-minus-ST contrast is −0.2 with a confidence interval spanning zero. In the full-period legacy sample the corresponding OPEN-minus-ST contrast is 21.0 and excludes zero. The two windows therefore disagree about the OPEN–ST comparison in compensation trade-offs, while agreeing that placement trade-offs are larger for ST candidates in both. We report both rather than selecting between them, and note that the samples differ in period, in category construction, and in whether EWS candidates are separately identified. The placement and NIRF contrasts are more stable across the two windows than the compensation contrast.

Panel B of Table A18 reports trade-offs by home-district poverty and remoteness. The compensation trade-off is 80.1 kilometers per lakh one standard deviation below mean district poverty and 93.1 one standard deviation above, a difference of 13.0 that is precisely estimated.

Table A20 compares trade-offs across the Gender-Neutral and Female-Only seat pools for 2018–2023. The compensation trade-off is essentially identical across pools, at 86.1 and 86.0 kilometers per lakh, while the placement trade-off is 5.5 kilometers per percentage point lower and the NIRF trade-off 4.0 kilometers per score point higher in the Female-Only pool, with both contrasts stable in sign across every cycle in panel B. This comparison should be read as a contrast across seat pools and not as a gender preference contrast. Candidate gender is not recorded in the 2016 and 2017 data, and the Female-Only pool exists only from 2018, so within the estimable window the female indicator and the policy-created pool are largely the same object.

**Table A18.** Heterogeneity by Social Category and Place of Origin

*Panel A. Social-category heterogeneity, 2019–2023*

| Social category | CTC trade-off (km per Rs. 1 lakh) | Placement trade-off (km per percentage point) | NIRF trade-off (km per score point) |
|---|---|---|---|
| OPEN | 85.5 [74.3, 96.7] | 18.0 [14.5, 21.5] | 24.5 [20.0, 29.0] |
| EWS | 101.6 [86.9, 116.3] | 22.6 [18.2, 27.0] | 23.0 [17.5, 28.5] |
| OBC-NCL | 85.3 [73.9, 96.7] | 20.7 [16.9, 24.5] | 21.5 [17.0, 26.1] |
| SC | 77.1 [65.4, 88.9] | 22.5 [18.5, 26.6] | 19.4 [14.6, 24.2] |
| ST | 85.7 [69.7, 101.7] | 27.2 [21.5, 32.9] | 27.1 [20.5, 33.7] |
| OPEN minus ST | −0.2 [−12.4, 12.0] | −9.3 [−13.0, −5.5] | −2.6 [−6.6, 1.5] |
| $p$-value: equal across categories | < 0.001 | < 0.001 | < 0.001 |

*Panel B. Home-district poverty and remoteness*

| | CTC distance trade-off (km per Rs. 1 lakh) | | | |
|---|---|---|---|---|
| Home-district characteristic | −1 SD | +1 SD | Difference (+1 SD minus −1 SD) | $p$-value |
| Poverty | 80.1 [71.4, 88.9] | 93.1 [81.8, 104.4] | 13.0 [6.0, 20.0] | < 0.001 |
| Remoteness | 83.6 [74.2, 93.0] | 88.5 [78.8, 98.1] | 4.9 [1.1, 8.6] | 0.010 |

*Panel C. Specification and sample*

| Measure | Social category | Place of origin |
|---|---|---|
| Candidate–program alternatives | 12,523,368 | 17,484,247 |
| Candidate choice sets | 77,000 | 112,101 |
| Program–year clusters | 1,345 | 2,291 |
| Rank-varying attribute slopes | Yes | Yes |
| Two-way clustered standard errors | Yes | Yes |

*Notes:* The table reports rank-adjusted heterogeneity in distance-equivalent program trade-offs from conditional-choice models estimated by Poisson pseudo-maximum likelihood with candidate choice-set fixed effects. Standard errors are two-way clustered by candidate choice set and academic-program-by-admission-year; brackets report 95% delta-method confidence intervals. Panel A uses the common 2019–2023 social-category sample and allows distance, CTC, placement, and NIRF coefficients to vary by social category and with the within-admission-year, within-category rank percentile. Rank percentile is centered at 0.50, so category-specific trade-offs are evaluated at the median rank percentile. The OPEN-minus-ST row reports a nonlinear contrast, and the final row tests equality across all five categories. Panel B estimates standardized home-district poverty and remoteness jointly; the other characteristic is held at its mean when one varies from −1 to +1 standard deviation. Poverty is a district-level measure and should not be interpreted as household poverty. CTC is annual program median compensation in 2023 lakh rupees per annum; placement is measured in percentage points and NIRF in score points. The estimates describe heterogeneity in revealed choice behavior conditional on reconstructed feasible menus and rank and are not interpreted causally.

**Table A19.** Social-Category Heterogeneity Using Legacy Categories, 2015–2023

*Panel A. Rank-adjusted trade-offs by social category*

| Social category | CTC trade-off (km per Rs. 1 lakh) | Placement trade-off (km per percentage point) | NIRF trade-off (km per score point) |
|---|---|---|---|
| OPEN | 87.3 | 17.9 | 18.3 |
| | [78.0, 96.5] | [15.1, 20.8] | [14.9, 21.8] |
| OBC-NCL | 87.2 | 20.0 | 15.4 |
| | [77.6, 96.9] | [16.8, 23.1] | [11.8, 19.0] |
| SC | 73.8 | 21.0 | 13.4 |
| | [64.5, 83.1] | [17.9, 24.2] | [9.7, 17.0] |
| ST | 66.3 | 24.3 | 17.3 |
| | [54.4, 78.2] | [20.2, 28.4] | [12.8, 21.8] |
| OPEN minus ST | 21.0 | −6.4 | 1.0 |
| | [10.8, 31.2] | [−9.0, −3.7] | [−1.4, 3.4] |
| *p-value: equal across categories* | < 0.001 | < 0.001 | < 0.001 |

*Panel B. Estimation sample*

| Sample feature | Count |
|---|---|
| Candidate–program alternatives | 17,357,344 |
| Candidate choice sets | 111,161 |
| Program–year clusters | 2,291 |
| Admission cycles | 9 |
| Retained social categories | 4 |

*Notes:* The table extends the rank-adjusted social-category heterogeneity exercise in Table A17 to the full 2015–2023 admission period using the four social categories observed throughout the sample: OPEN, OBC-NCL, SC, and ST. EWS entered the admission system in 2019; EWS candidates in 2019–2023 are excluded from this robustness sample and are not recoded as OPEN. The specification otherwise matches Table A17: distance, CTC, placement, and NIRF slopes vary by social category and with the within-admission-year, within-category rank percentile. Rank percentile is centered at 0.50, so reported trade-offs are evaluated at the median rank percentile. The OPEN-minus-ST row reports a nonlinear contrast, and the final row tests equality across the four retained categories. Models are estimated by Poisson pseudo-maximum likelihood with candidate choice-set fixed effects; standard errors are two-way clustered by candidate choice set and academic-program-by-admission-year. Brackets report 95% delta-method confidence intervals. CTC is annual program median compensation in 2023 lakh rupees per annum; placement is measured in percentage points and NIRF in score points. The estimates describe rank-adjusted heterogeneity in revealed choice behavior and are not interpreted causally.

**Table A20.** Female-Identified Seat-Pool Heterogeneity

*Panel A. Rank- and admission-cycle-adjusted average trade-offs, 2018–2023*

| Program attribute | Gender-Neutral seat | Female-Only seat | Female-Only minus Gender-Neutral | $p$-value |
|---|---|---|---|---|
| CTC trade-off (km per Rs. 1 lakh) | 86.1<br>[74.5, 97.7] | 86.0<br>[76.9, 95.0] | −0.2<br>[−8.5, 8.1] | 0.968 |
| Placement trade-off (km per percentage point) | 21.6<br>[17.9, 25.2] | 16.0<br>[12.6, 19.5] | −5.5<br>[−7.5, −3.5] | < 0.001 |
| NIRF trade-off (km per score point) | 18.1<br>[14.0, 22.3] | 22.2<br>[18.3, 26.1] | 4.0<br>[1.6, 6.4] | < 0.001 |

*Panel B. Female-Only minus Gender-Neutral contrasts by admission cycle*

| Admission cycle | CTC trade-off (km per Rs. 1 lakh) | Placement trade-off (km per percentage point) | NIRF trade-off (km per score point) |
|---|---|---|---|
| 2018 | −1.0<br>[−10.8, 8.8] | −4.8<br>[−7.1, −2.6] | 6.8<br>[4.4, 9.1] |
| 2019 | 0.4<br>[−8.6, 9.5] | −5.5<br>[−8.0, −3.1] | 6.3<br>[4.0, 8.6] |
| 2020 | 3.1<br>[−5.4, 11.6] | −5.2<br>[−7.4, −3.1] | 5.0<br>[2.8, 7.3] |
| 2021 | 3.3<br>[−5.5, 12.1] | −5.0<br>[−7.4, −2.6] | 4.0<br>[1.5, 6.5] |
| 2022 | −3.6<br>[−14.2, 7.0] | −5.3<br>[−7.9, −2.7] | 2.5<br>[−0.4, 5.4] |
| 2023 | −2.9<br>[−13.7, 8.0] | −7.0<br>[−9.9, −4.2] | 1.0<br>[−2.7, 4.7] |
| $p$-value: equal across admission cycles | 0.390 | 0.634 | < 0.001 |
| $p$-value: all contrasts jointly zero | 0.458 | < 0.001 | < 0.001 |

*Panel C. Estimation sample*

| Sample feature | Count |
|---|---|
| Candidate–program alternatives | 14,290,099 |
| Candidate choice sets | 88,657 |
| Female-Only recipient choice sets | 16,454 |
| Gender-Neutral recipient choice sets | 72,203 |
| Program–year clusters | 1,595 |

*Notes:* The table reports heterogeneity by the gender pool of the candidate's allocated IIT seat during 2018–2023. A Female-Only recipient is point-identified as female by the allocation rule; a Gender-Neutral recipient is not necessarily male because female candidates are eligible for both seat pools. The estimates therefore describe seat-pool heterogeneity rather than a female–male effect. Models are estimated by Poisson pseudo-maximum likelihood with candidate choice-set fixed effects. Distance, CTC, placement, and NIRF slopes vary with Female-Only-recipient status, rank percentile, and admission cycle. Rank percentile is centered at 0.50, so reported trade-offs are evaluated at the median rank percentile. Panel A reports admission-cycle-weighted average trade-offs. Panel B reports Female-Only-minus-Gender-Neutral contrasts by admission cycle together with nonlinear Wald tests of equality across cycles and of whether all annual contrasts are jointly zero. Brackets report 95% delta-method confidence intervals; standard errors are two-way clustered by candidate choice set and academic-program-by-admission-year. CTC is annual program median compensation in 2023 lakh rupees per annum; placement is measured in percentage points and NIRF in score points. The estimates are conditional on reconstructed feasible menus and are not interpreted causally.

# E Female Supernumerary

This appendix reports the exhibits underlying Section 6. The material is organized in five groups: the institutional architecture of the 2018 reform and the reconstruction of the 14-percent assignment rule; the observed access-frontier architecture the reform produced; the frozen local design and its validity checks; the local effects of additional capacity on observed frontiers; and the identification limits governing realized allocations.

**Policy architecture and rule reconstruction.** Table A21 reports seat-pool accounting before and after the reform. Historical female allotment shares were 8.9, 7.9, and 9.2 percent in 2015 through 2017. In 2018 total capacity across 250 academic programs was 12,079 seats, split into a Gender-Neutral pool of 10,227 and a Female-Only pool of 1,852, of which 800 seats were genuinely additional supernumerary capacity. Table A22 reconstructs the assignment rule described in Section 6.1. Of 244 continuing academic programs, 170 had a 2017 female allotment share below 14 percent and 162 received positive additional capacity, totaling 794 additional seats. The literal rule reproduces the binary treatment for 230 of 244 programs and the exact added-seat count for 191, with 11 below-threshold programs receiving no additional capacity and 3 at-or-above programs receiving some. Compliance is therefore high but imperfect, which is why the design in Section 6.2 is fuzzy rather than sharp.

**Observed access-frontier architecture.** Table A23 compares the most permissive observed Female-Only and Gender-Neutral closing-rank frontiers across 825 matched program-by-reservation-category cells. The Female-Only frontier extends farther down the relevant rank list in 92.0 percent of cells, with a median Female-Only-to-Gender-Neutral ratio of 1.657. Coverage varies substantially by category, from 99.6 percent of OPEN cells to 64.5 percent of ST cells. Table A24 translates this into candidate-level opportunity sets for the 1,712 point-identified regular Female-Only recipients in 2018. The median female-specific increment is 56 physical programs, which is 46.4 percent of the median reconstructed opportunity set. These are descriptive features of the realized 2018 architecture and are not effects of marginal capacity.

**Frozen local design and validity.** Table A25 reports the frozen local support and the principal predetermined continuity checks. At the preferred bandwidth of four percentage points the sample contains 80 programs across 16 IITs at 35 distinct running-variable mass points. Predetermined 2015 and 2016 female allotment shares and 2017 program capacity show no discontinuity that the data can distinguish from zero at any bandwidth. Table A33 reports the full predetermined covariate battery and Figure A7 displays it. Figure A2 shows the distribution of programs across mass points. Because the running variable is discrete with a small number of mass points and clustering is at the IIT level with 16 clusters, inference follows the discrete-running-variable and weak-identification conventions described in Section 6.2 rather than conventional bandwidth-asymptotic standard errors.

**Effects of additional capacity on observed frontiers.** Table A26 reports the local effects at three bandwidths, with figures showing the corresponding discontinuities. The strongest result is in panel B: at the preferred bandwidth, additional capacity raises the predetermined-weighted share of routes on which the observed Female-Only frontier exceeds the Gender-Neutral frontier by 10.8 percentage points in reduced form and 22.5 percentage points in the instrumented specification, with an Anderson–Rubin $p$-value of 0.017 that is stable across bandwidths. Figure A4 displays this discontinuity. The extensive-margin activation result in panel A is weaker, at 4.2 percentage points in reduced form with an Anderson–Rubin $p$-value of 0.360, and is shown in Figures A3 and A8. Panel C reports the magnitude of the frontier extension, displayed in Figure A5, and is not precisely estimated. One measurement point governs the reading of panel A: the absence of a finite observed Female-Only closing rank is an observability outcome, since a frontier is recorded only when a positive finite closing rank appears in some 2018 round, and it is not evidence that a program had no institutional Female-Only capacity.

**Identification limits and realized allocations.** Table A27 reports the first-stage diagnostic, which is 3.97 at the preferred bandwidth, together with sample composition and weak-instrument-robust inference. Restricting to programs with observed 2016 history reduces the sample from 80 to 72 programs and attenuates the positive-extension estimate from 22.5 to 7.8 percentage points before any historical control is added; adding the control changes it only to 6.7. The attenuation is therefore attributable to sample composition rather than to covariate adjustment, which is why the full 80-program specification remains preferred and the historical specifications are treated as sensitivity analyses. Table A28 reports observed frontier support and predetermined-weight coverage, which is 96.0 percent overall but 65.5 percent for ST cells. Table A29 certifies realized allocations against exact administrative routes: all 1,712 allocation routes are recovered, 1,680 recipients received programs beyond every eligible observed Gender-Neutral route, and all 1,680 are supported by an observed eligible Female-Only frontier. Table A30 reports bounds on realized beyond-Gender-Neutral use, since the identities of women allocated through Gender-Neutral seats are only partially identified. Table A31 and Figure A6 report the local effects of additional capacity on realized beyond-Gender-Neutral allocation counts, which are not precisely identified. Table A32 reports category-specific estimates that are too imprecise to support separate causal conclusions and are presented as mechanism diagnostics.

Taken together, these exhibits support a narrow claim. The reform substantially expanded the observed female-specific access frontier, and locally induced additional capacity increased the share of routes on which that frontier extends beyond the Gender-Neutral frontier. The magnitude of the frontier shift and the downstream effect on realized allocation counts are not precisely identified, and no exhibit in this appendix re-solves the assignment market under a no-policy counterfactual. The appendix is therefore evidence that an administrative capacity rule moves the opportunity object measured throughout the paper, and not a welfare evaluation of the reform, consistent with Section 7.4.

**Table A21.** Institutional Seat-Pool Accounting Before and After the 2018 Female-Supernumerary Reform

*Panel A. Aggregate capacity and female representation*

| | Pre-reform | | | Reform year |
|---|---|---|---|---|
| Measure | 2015 | 2016 | 2017 | 2018 |
| Academic programs | 216 | 235 | 245 | 250 |
| Total seat capacity | 10,054 | 10,574 | 10,988 | 12,079 |
| Historical female allotments | 898 | 835 | 1,013 | – |
| Historical female allotment share (%) | 8.9 | 7.9 | 9.2 | – |
| Gender-Neutral pool capacity | – | – | – | 10,227 |
| Female-Only pool capacity | – | – | – | 1,852 |
| Additional female-supernumerary capacity $x$ | – | – | – | 800 |

*Panel B. 2018 reservation-category seat pools*

| Reservation category | GN pool | FO pool | Total | FO share (%) |
|---|---|---|---|---|
| OPEN | 5,185 | 942 | 6,127 | 15.4 |
| OBC-NCL | 2,761 | 496 | 3,257 | 15.2 |
| SC | 1,518 | 284 | 1,802 | 15.8 |
| ST | 763 | 130 | 893 | 14.6 |
| Total | 10,227 | 1,852 | 12,079 | 15.3 |

*Notes:* Before 2018, IIT seats were allocated without separate gender pools; the female entries in Panel A therefore report historical female allotments contained within total program capacity. In 2018, each academic program was divided into a Gender-Neutral (GN) pool and a Female-Only (FO) pool. The FO pool includes genuinely additional female-supernumerary capacity, denoted by $x$, but FO capacity is a seat-pool quantity and should not be interpreted as realized female admissions.

Under the baseline 14% construction, for a continuing program with 2017 female representation below 14%, additional capacity was computed from

$$x_p^* = \frac{0.14C_{p,2017} - f_{p,2017}}{0.86},$$

or equivalently,

$$f_{p,2017} + x_p^* = 0.14\left(C_{p,2017} + x_p^*\right).$$

Fractional values of $x_p^*$ were rounded up when constructing the seat matrix. For programs satisfying $0.14C_{p,2017} < f_{p,2017} \leq 0.20C_{p,2017}$, the FO pool contained $f_{p,2017}$ seats and the GN pool contained $C_{p,2017} - f_{p,2017}$ seats. For programs with $f_{p,2017} > 0.20C_{p,2017}$, the FO and GN pools were set to 20% and 80% of capacity, respectively. For academic programs introduced for the first time in 2018, the FO pool was set to 14% of capacity and the remainder to the GN pool. The implementation rules also permitted an institute-specific desired female percentage between 14% and 20%, in which case that target replaced 14% in the corresponding calculation.

Panel B reports the 2018 GN and FO capacity split across the four regular vertical reservation categories, combining regular and PwD seats within each category. The FO share is FO capacity divided by total category capacity.

**Table A22.** Reconstruction of the 14 % Assignment Rule, Treatment Compliance, and Frozen Local Support

| Statistic | Full policy universe | Symmetric bandwidth around 14% threshold ±3 | ±4 | ±5 |
|---|---|---|---|---|
| *Panel A. Policy reconstruction and treatment compliance* | | | | |
| Continuing academic programs | 244 | – | – | – |
| Programs below 14% in 2017 | 170 | – | – | – |
| Programs with positive additional capacity, $x_p > 0$ | 162 | – | – | – |
| Total genuinely additional seats | 794 | – | – | – |
| Programs predicted positive by literal 14% rule | 170 | – | – | – |
| Binary treatment-rule matches | 230 | – | – | – |
| Below 14% with $x_p = 0$ | 11 | – | – | – |
| At or above 14% with $x_p > 0$ | 3 | – | – | – |
| Exact added-seat-count matches | 191 | – | – | – |
| Unique exact running-variable mass points | 99 | – | – | – |
| *Panel B. Frozen local support* | | | | |
| Programs | – | 52 | 80 | 89 |
| Below / at-or-above threshold | – | 21 / 31 | 40 / 40 | 47 / 42 |
| IITs | – | 11 | 16 | 16 |
| Running-variable mass points | – | 26 | 35 | 41 |
| First stage | – | 0.446 (0.295) | 0.482 (0.242) | 0.529 (0.216) |
| First-stage $t^2$ diagnostic | – | 2.29 | 3.97 | 5.99 |

*Notes:* Panel A reports the reconstruction of the baseline 14 percent assignment rule for the 244 academic programs observed in both 2017 and 2018. The running variable is the program's 2017 female allotment share relative to the 14 percent threshold. The empirical treatment is an indicator for receiving positive genuinely additional female-supernumerary capacity in 2018, $D_p = \mathbf{1}\{x_p > 0\}$. The literal 14 percent rule predicts positive additional capacity for 170 programs and correctly classifies the observed extensive-margin treatment for 230 of 244 programs (94.3%). It also reproduces the exact number of additional seats for 191 programs (78.3%). The 794 genuinely additional seats reported in Panel A refer only to continuing programs; including the six seats associated with programs introduced in 2018 yields the aggregate 2018 total of 800 reported in Table A21. The 11 programs below the threshold with $x_p = 0$ and the three programs at or above the threshold with $x_p > 0$ generate fuzzy, rather than sharp, treatment assignment.
Panel B reports the outcome-blind local samples retained after restricting to IITs represented on both sides of the threshold within each symmetric bandwidth. The ±4-percentage-point window is the frozen preferred specification. First-stage estimates come from local-linear regressions with separate slopes, IIT fixed effects, and IIT-clustered standard errors; standard errors are in parentheses. Because the running variable is discrete, the table also reports the number of distinct running-variable mass points in each estimation window.

**Table A23.** Female-Only and Gender-Neutral Access Frontiers by Reservation Category

| Reservation category | Paired cells | Closing-rank frontiers | | | Relative frontier expansion | | |
|---|---|---|---|---|---|---|---|
| | *N* | Median GN CR | Median FO CR | Median FO–GN gap | Median FO/GN ratio | Median log gap | FO CR > GN CR (% of cells) |
| OPEN | 250 | 4,919 | 8,122 | 2,711.5 | 1.624 | 0.485 | 99.6 |
| OBC-NCL | 235 | 2,163 | 4,004 | 1,639.0 | 1.813 | 0.595 | 98.3 |
| SC | 219 | 1,182 | 1,912 | 633.0 | 1.599 | 0.469 | 91.8 |
| ST | 121 | 464 | 524 | 92.0 | 1.369 | 0.314 | 64.5 |
| **All regular categories** | **825** | – | – | – | **1.657** | **0.505** | **92.0** |

*Notes:* The table summarizes the most permissive finite Female-Only (FO) and Gender-Neutral (GN) closing-rank frontiers observed across the 2018 counseling rounds for matched academic-program–reservation-category cells. Regular categories are OPEN, OBC-NCL, SC, and ST. Within each category, FO and GN closing ranks are measured on the same category-specific rank list. A positive FO–GN gap, an FO/GN ratio above one, or an FO closing rank above the corresponding GN closing rank indicates that the observed FO frontier extends farther down the relevant rank list. Because OPEN, OBC-NCL, SC, and ST use distinct rank scales, pooled raw closing ranks and absolute rank gaps are not reported in the all-category row. The pooled relative statistics use 825 matched program–category cells spanning 250 academic programs. These comparisons describe the observed 2018 access-frontier architecture and are not causal effects of marginal female-supernumerary capacity.

**Table A24.** Candidate-Specific Female-Specific Expansion of Reconstructed IIT Opportunity Sets

| Reservation category | *N* | Opportunity sets | | | Female-specific expansion | | | |
|---|---|---|---|---|---|---|---|---|
| | | Median GN reach | Median FO reach | Median menu $F_i$ | Median increment | Increment share (%) | Zero GN reach (%) | Expansion vs. GN (%) |
| OPEN | 912 | 82.0 | 151.0 | 151.0 | 56.0 | 45.7 | 10.4 | 64.6 |
| OBC-NCL | 466 | 56.0 | 141.5 | 147.5 | 68.0 | 61.7 | 25.5 | 73.3 |
| SC | 254 | 72.0 | 125.5 | 145.5 | 51.0 | 50.2 | 3.9 | 83.1 |
| ST | 80 | 128.5 | 61.5 | 161.5 | 20.5 | 18.8 | 0.0 | 23.2 |
| **All regular FO recipients** | **1,712** | **76.0** | **139.0** | **151.0** | **56.0** | **46.4** | **13.1** | **63.0** |

*Notes:* The table reports candidate-specific physical-program opportunity sets for the 1,712 point-identified regular Female-Only (FO) recipients in 2018. Column heads are abbreviated: GN reach and FO reach are the medians of the candidate's Gender-Neutral–reachable and FO-reachable physical-program counts; menu $F_i$ is the median reconstructed opportunity set; increment is the median female-specific incremental program count; increment share is that increment as a percentage of the reconstructed opportunity set; and expansion vs. GN is the increment divided by the GN-reachable count. Program reach is evaluated separately on every merit-list route for which the candidate is eligible and is then unioned at the physical-program level. GN-reachable programs are programs attainable through at least one eligible observed GN route, while FO-reachable programs are attainable through at least one observed FO route. The reconstructed opportunity set is the candidate's literal empirical menu $F_i$. A female-specific incremental program is a physical IIT program reachable through at least one FO route but through none of the candidate's eligible observed GN routes. Zero GN reach reports the share of candidates with no physical IIT program reachable through an observed eligible GN route; expansion vs. GN is reported only for candidates with positive GN reach. Route cutoffs use the most permissive finite closing rank observed across the 2018 counseling rounds. These quantities describe the observed 2018 access architecture and are not counterfactual opportunity sets obtained by re-solving the allocation system without Female-Only pools.

**Table A25.** Frozen Local Design and Key Predetermined Continuity Checks

| | Symmetric bandwidth around 14% threshold | | |
|---|---|---|---|
| Predetermined measure | ±3 | ±4 | ±5 |
| *Panel A. Key predetermined continuity checks* | | | |
| Female allotment share, 2015 (pp) | -6.934 | -3.471 | -1.391 |
| | (5.228) | (5.904) | (3.887) |
| *p-value* | 0.221 | 0.568 | 0.727 |
| Female allotment share, 2016 (pp) | -2.050 | -2.903 | -1.897 |
| | (1.875) | (2.529) | (2.003) |
| *p-value* | 0.303 | 0.270 | 0.360 |
| 2017 program capacity | 18.699 | 5.976 | 10.158 |
| | (16.920) | (12.122) | (16.183) |
| *p-value* | 0.295 | 0.629 | 0.540 |
| Log 2017 program capacity | 0.445 | 0.205 | 0.271 |
| | (0.364) | (0.285) | (0.341) |
| *p-value* | 0.249 | 0.483 | 0.440 |
| *Panel B. Frozen local support* | | | |
| Programs | 52 | 80 | 89 |
| IITs | 11 | 16 | 16 |
| Running-variable mass points | 26 | 35 | 41 |

*Notes:* Panel A reports predetermined continuity checks at the 14 percent historical female-share threshold. Each coefficient is the estimated discontinuity for programs immediately below the threshold from the frozen both-side-IIT local-linear specification with separate slopes on the two sides, IIT fixed effects, and IIT-clustered standard errors. Parentheses report IIT-clustered standard errors. Female allotment shares are expressed in percentage points. Panel B reports the local-support samples obtained after restricting to IITs represented on both sides of the threshold within each symmetric bandwidth. The ±4-percentage-point window was frozen as the preferred specification before examining the final policy outcomes. Because the running variable is constructed from integer female allotments and program capacities, it has discrete support; the final row therefore reports the number of distinct running-variable mass points in each window. For predetermined measures with missing historical observations, Panel A uses the available observations for that measure and re-imposes the both-side-IIT restriction after dropping missing values; the Panel B program counts therefore need not equal the estimation sample for every continuity regression. The full predetermined-covariate continuity battery is reported in Table A33.

**Table A26.** Effects of Additional Capacity on Observed Female-Access Frontiers

| | Symmetric bandwidth around 14% threshold | | |
|---|---|---|---|
| Bandwidth | (1)<br>±3 | (2)<br>±4<br>*Preferred* | (3)<br>±5 |
| *Panel A. Observed Female-Only-frontier activation* | | | |
| Finite FO-frontier share, IV (pp) | 17.4<br>(19.1) | 8.8<br>(11.7) | 8.0<br>(9.8) |
| Reduced form (pp) | 7.8<br>(5.0) | 4.2<br>(4.5) | 4.2<br>(4.5) |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.156 | 0.360 | 0.356 |
| *Panel B. Positive observed female-access extension* | | | |
| FO frontier exceeds GN, weighted share, IV (pp) | 23.8<br>(17.2) | 22.5<br>(15.1) | 19.7<br>(11.8) |
| Reduced form (pp) | 10.6<br>(3.7) | 10.8<br>(4.0) | 10.4<br>(4.5) |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.017 | 0.017 | 0.036 |
| *Panel C. Magnitude of the observed frontier extension* | | | |
| Observed frontier-extension index, $G_p$, IV | 0.096<br>(0.164) | 0.128<br>(0.131) | 0.134<br>(0.104) |
| Implied change (%) | 10.1 | 13.6 | 14.4 |
| Reduced form | 0.043<br>(0.084) | 0.062<br>(0.074) | 0.071<br>(0.070) |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.622 | 0.418 | 0.323 |
| Programs | 52 | 80 | 89 |

*Notes:* The table decomposes the effect of positive genuinely additional female-supernumerary capacity on observed female-access frontiers. The instrument is an indicator for a program's 2017 female allotment share being below the 14 percent threshold, and the endogenous treatment is $D_p = \mathbf{1}\{x_p > 0\}$. All specifications are local linear with separate slopes, IIT fixed effects, and IIT-clustered standard errors, and restrict the sample to IITs represented on both sides of the threshold within the indicated symmetric bandwidth. The ±4-percentage-point window in column (2) is the frozen preferred specification.

Panel A reports the predetermined-2017-capacity-weighted share of regular reservation-category routes for which a finite Female-Only (FO) frontier is observed. This is an *observed finite-frontier activation* measure; absence of a finite FO closing rank is not interpreted as evidence of zero institutional FO capacity.

Panel B reports the predetermined-capacity-weighted share of regular-category routes for which the observed FO frontier extends beyond the corresponding Gender-Neutral (GN) frontier. Panel C reports the predetermined-capacity-weighted magnitude of that observed extension, $G_p$, constructed from $\max\{0, \log(CR^{FO}/CR^{GN})\}$ over observed finite frontiers. The implied-change row reports $100[\exp(\widehat{\beta}) - 1]$.

Parentheses report IIT-clustered standard errors. Anderson–Rubin $p$-values test $H_0 : \beta = 0$ and provide weak-instrument-robust inference. The table therefore reports structural IV estimates without conventional significance stars. Historical-female-share specifications and their associated sample-composition changes are reported separately in Table A27.

**Table A27.** Historical-Data Sample Composition, Covariate Adjustment, and Weak-Instrument-Robust Inference

| | Full sample | 2016 history observed | | 2015–16 history observed | |
|---|---|---|---|---|---|
| Specification | (1) Unadjusted | (2) Unadjusted | (3) + 2016 history | (4) Unadjusted | (5) + 2015–16 history |
| *Panel A. Positive observed female-access extension* | | | | | |
| Programs | 80 | 72 | 72 | 67 | 67 |
| IITs | 16 | 15 | 15 | 13 | 13 |
| First-stage $t^2$ diagnostic | 3.97 | 7.68 | 9.58 | 7.79 | 8.93 |
| Reduced form (pp) | 10.8 | 4.5 | 4.2 | 3.9 | 2.5 |
| | (4.0) | (5.1) | (5.5) | (5.4) | (5.2) |
| IV estimate (pp) | 22.5 | 7.8 | 6.7 | 6.7 | 4.1 |
| | (15.1) | (9.6) | (9.5) | (10.0) | (9.3) |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.017 | 0.396 | 0.458 | 0.490 | 0.644 |
| *Panel B. Observed frontier-extension index, $G_p$* | | | | | |
| Reduced form | 0.062 | 0.061 | 0.084 | 0.076 | 0.051 |
| | (0.074) | (0.095) | (0.101) | (0.100) | (0.082) |
| IV estimate | 0.128 | 0.107 | 0.133 | 0.130 | 0.084 |
| | (0.131) | (0.152) | (0.148) | (0.160) | (0.136) |
| Implied change (%) | 13.6 | 11.3 | 14.2 | 13.8 | 8.8 |
| Anderson–Rubin $p$-value, $\beta = 0$ | 0.418 | 0.527 | 0.418 | 0.464 | 0.549 |

*Panel C. Anderson–Rubin inference across frozen bandwidths*

| Outcome | Bandwidth | IV point estimate | AR $p$-value | AR 95% confidence set |
|---|---|---|---|---|
| Positive-extension share, $X_p$ | ±3 | 0.238 | 0.017 | $(-\infty, -0.4161] \cup [0.0453, \infty)$ |
| | ±4 | 0.225 | 0.017 | $(-\infty, -3.6367] \cup [0.0362, \infty)$ |
| | ±5 | 0.197 | 0.036 | [0.0144, 1.6289] |
| Observed frontier-extension index, $G_p$ | ±3 | 0.096 | 0.622 | $\mathbb{R}$ |
| | ±4 | 0.128 | 0.418 | $\mathbb{R}$ |
| | ±5 | 0.134 | 0.323 | [−0.6198, 0.4033] |

*Notes:* Panels A and B use the frozen ±4-percentage-point design. Column (1) uses the full 80-program sample. Columns (2)–(3) use the same 72 programs with observed 2016 female history, while columns (4)–(5) use the same 67 programs with observed 2015–16 female history. Thus, comparisons of columns (1) with (2) or (4) capture changes in sample composition, whereas comparisons of columns (2) with (3) and columns (4) with (5) isolate within-sample historical-covariate adjustment. All specifications are local linear with separate slopes, IIT fixed effects, and IIT-clustered standard errors; parentheses report standard errors. Panel C reports Anderson–Rubin weak-instrument-robust inference for the unadjusted ±3, ±4, and ±5 specifications; point estimates there are reported in proportions rather than percentage points. Confidence sets contain the structural-effect values not rejected at the 5 percent level. The ±4 bandwidth is the frozen preferred design throughout.

† The confidence set reaches the lower boundary of the search grid and is therefore truncated from below.

‡ The confidence set reaches the upper boundary of the search grid and is therefore truncated from above.

**Table A28.** Observed Frontier Support and Predetermined-Weight Coverage

| Reservation category | Predetermined 2017 support | | | Observed finite frontiers | | | Observed female-access extension | |
|---|---|---|---|---|---|---|---|---|
| | Positive-weight cells | 2017 weight | GN finite, FO absent | Finite GN | Finite FO | Missing-FO weight | Finite-FO weight coverage (%) | Positive-extension weight share (%) |
| OPEN | 244 | 5,385 | 0 | 244 | 244 | 0 | 100.0 | 99.6 |
| OBC-NCL | 244 | 2,868 | 15 | 244 | 229 | 65 | 97.7 | 96.5 |
| SC | 243 | 1,592 | 29 | 243 | 214 | 82 | 94.8 | 88.4 |
| ST | 242 | 799 | 119 | 239 | 120 | 273 | 65.5 | 46.1 |
| **All regular categories** | **973** | **10,644** | **163** | **970** | **807** | **420** | **96.0** | **93.1** |

*Notes:* The table summarizes observed closing-rank support for regular program–reservation-category cells in the 244 continuing-program policy universe. Predetermined 2017 weight is the corresponding regular-category capacity used to aggregate route-level outcomes. A finite GN or FO frontier means that a positive finite Gender-Neutral (GN) or Female-Only (FO) closing rank is observed in at least one 2018 counseling round; the most permissive observed closing rank across rounds defines the route frontier. "GN finite, FO absent" counts positive-weight cells with an observed finite GN frontier but no finite observed FO frontier, while missing-FO weight is the associated predetermined 2017 capacity. Finite-FO weight coverage is the share of predetermined weight with an observed finite FO frontier. Positive-extension weight share is the share of predetermined weight for which the observed FO frontier exceeds the corresponding GN frontier. Absence of a finite FO closing rank is treated as zero *observed* frontier extension and should not be interpreted as evidence of zero institutional FO capacity.

**Table A29.** Exact Administrative-Route Certification of Realized Beyond-GN Allocations

| | Count | |
|---|---|---|
| *Panel A. Candidate-level realized-use certification* | | |
| Point-identified regular FO recipients | | 1,712 |
| FO recipients whose allocated program is GN-reachable | | 32 |
| FO recipients beyond every eligible observed GN route | | 1,680 |
| Beyond-GN recipients with an observed eligible FO frontier | | 1,680 |
| Beyond-GN recipients supported only by revealed feasibility | | 0 |
| Exact allocated-route failures | | 0 |
| *Panel B. Exact administrative-route recovery* | | |
| OPEN | 919 / 919 | (100.0%) |
| OBC-NCL | 460 / 460 | (100.0%) |
| SC | 254 / 254 | (100.0%) |
| ST | 79 / 79 | (100.0%) |
| All regular categories | 1,712 / 1,712 | (100.0%) |

*Notes:* The table reports an independent exact-route certification for the 1,712 point-identified regular Female-Only (FO) recipients in the 2018 analytical sample. These observations are not the total number of regular female recipients; the identities of women allocated through Gender-Neutral (GN) seats are partially identified and are treated separately in Table A30. For each recipient, the allocated category-by-pool route is matched to the corresponding most-permissive observed OR–CR frontier across 2018 rounds. OPEN routes use AIR/CRL rank; OBC-NCL, SC, and ST routes use the corresponding category rank. A recipient is classified as beyond GN when her allocated physical program lies beyond every eligible observed GN route available to her. Thirty-two FO recipients received programs that were nevertheless reachable through at least one eligible GN route; the remaining 1,680 were beyond every such GN route. All 1,680 beyond-GN allocations are supported by an observed eligible FO frontier, and all 1,712 exact administrative allocation routes are recovered. These classifications describe realized allocations relative to observed 2018 frontiers and do not constitute a re-solved no-policy allocation counterfactual.

**Table A30.** Gender Partial Identification and Bounds on Realized Beyond-GN Use

| | Identification bounds | |
|---|---|---|
| Statistic | Lower | Upper |
| *Panel A. Aggregate regular-female bounds* | | |
| Unidentified GN women in regular analytical scope | 0 | 10 |
| Total regular female recipients | 1,712 | 1,722 |
| Beyond-GN regular female recipients | 1,680 | 1,680 |
| GN-reachable regular female recipients | 32 | 42 |
| Beyond-GN share of regular female recipients (%) | 97.56 | 98.13 |
| *Panel B. Beyond-GN share bounds by reservation category* | | |
| OPEN (%) | 98.81 | 99.89 |
| OBC-NCL (%) | 95.80 | 97.85 |
| SC (%) | 91.67 | 95.28 |
| ST (%) | 78.89 | 88.75 |
| *Panel C. Beyond-GN share bounds in frozen local samples* | | |
| Bandwidth ±3 (%) | 94.17 | 97.32 |
| Bandwidth ±4 (%) | 95.98 | 97.86 |
| Bandwidth ±5 (%) | 96.35 | 97.98 |

*Notes:* Exactly ten women occupied Gender-Neutral (GN) seats in the 2018 IIT allocation, but their individual identities are not observed. The number of these women falling within the regular analytical scope is therefore bounded between zero and ten, implying between 1,712 and 1,722 regular female recipients. The number of realized allocations beyond every eligible observed GN frontier is point identified at 1,680: an unidentified additional woman necessarily occupies a GN seat and therefore contributes zero to the beyond-GN numerator. The corresponding overall beyond-GN share is consequently bounded between 97.56 and 98.13 percent. Panel B reports marginal sharp bounds by reservation category under the same global ten-woman constraint. Panel C reports the corresponding bounds for the frozen ±3, ±4, and ±5 local program samples.

**Table A31.** Local Effects of Additional Capacity on Realized Beyond-GN Allocations

| Specification | Frozen bandwidth | | | Historical-control sensitivity | |
|---|---|---|---|---|---|
| | ±3 | ±4 *(preferred)* | ±5 | ±4 2016 history | ±4 2015–16 history |
| *Panel A. Beyond-GN recipient count* | | | | | |
| Reduced form | 2.391 | 1.069 | 1.937 | −0.940 | −1.584 |
| | (2.451) | (1.706) | (2.434) | (2.652) | (2.224) |
| IV estimate | 5.357 | 2.220 | 3.658 | −1.490 | −2.635 |
| | (4.983) | (3.545) | (4.383) | (4.370) | (3.888) |
| Anderson–Rubin $p$-value | 0.352 | 0.540 | 0.439 | 0.728 | 0.490 |
| *Panel B. Beyond-GN recipients per 100 pre-policy seats* | | | | | |
| Reduced form | 1.896 | 2.301 | 1.999 | 2.697 | 2.692 |
| | (3.581) | (2.401) | (2.110) | (2.467) | (2.435) |
| IV estimate | 4.247 | 4.778 | 3.776 | 4.274 | 4.477 |
| | (5.968) | (3.408) | (3.021) | (3.084) | (3.093) |
| Anderson–Rubin $p$-value | 0.608 | 0.353 | 0.358 | 0.293 | 0.291 |
| *Panel C. First stage and estimation sample* | | | | | |
| First-stage $t^2$ statistic | 2.29 | 3.97 | 5.99 | 9.58 | 8.93 |
| Programs | 52 | 80 | 89 | 72 | 67 |

*Notes:* Panel A reports local effects on the program-level number of point-identified regular Female-Only recipients whose realized physical program lies beyond every eligible observed Gender-Neutral (GN) route. Panel B reports the same outcome per 100 predetermined 2017 program seats. Each reduced-form and IV cell reports the coefficient with its IIT-clustered standard error in parentheses. All specifications use local-linear regressions with separate slopes on either side of the 14 percent threshold and IIT fixed effects. The instrument indicates that a program's 2017 female share lies below the 14 percent threshold; the endogenous treatment indicates positive additional female-supernumerary capacity in 2018. The ±4 specification without historical controls is preferred. Anderson–Rubin $p$-values test $H_0 : \beta = 0$ and provide weak-instrument-robust inference. Panel C reports the corresponding first-stage $t^2$ statistic and number of academic programs. Conventional significance stars are omitted for the structural IV estimates.

**Table A32.** Reservation-Category Mechanism Diagnostics

| Statistic | Reduced form | IV estimate |
|---|---|---|
| *Panel A. Category-specific observed-frontier diagnostics at* ±4 | | |
| OPEN: positive-extension share | 3.7<br>(3.3) | 7.8<br>(6.1) |
| OPEN: observed extension index | −0.005<br>(0.090) | −0.011<br>(0.189) |
| OBC-NCL: positive-extension share | 10.0<br>(14.7) | 20.7<br>(31.3) |
| OBC-NCL: observed extension index | 0.138<br>(0.138) | 0.286<br>(0.214) |
| SC: positive-extension share | 45.1<br>(19.5) | 93.6<br>(71.8) |
| SC: observed extension index | 0.198<br>(0.144) | 0.412<br>(0.335) |
| ST: positive-extension share | −9.6<br>(21.0) | −20.0<br>(43.2) |
| ST: observed extension index | −0.051<br>(0.167) | −0.106<br>(0.337) |

*Panel B. Descriptive realized beyond-GN use by candidate social category*

| Reservation category | Beyond-GN / point-identified FO | All-female beyond-GN share bounds |
|---|---|---|
| OPEN | 911 / 912 | 98.8–99.9% |
| OBC-NCL | 456 / 466 | 95.8–97.9% |
| SC | 242 / 254 | 91.7–95.3% |
| ST | 71 / 80 | 78.9–88.8% |

*Notes:* Panel A reports reservation-category-specific diagnostics from the frozen unadjusted ±4-percentage-point design. For each category, the positive-extension share measures the predetermined-weighted share of observed category routes for which the Female-Only frontier extends beyond the corresponding Gender-Neutral frontier; these coefficients are reported in percentage points. The observed extension index measures the corresponding magnitude of observed frontier extension in log-index units. Each reduced-form and IV cell reports the coefficient with its IIT-clustered standard error in parentheses. These category-specific estimates are diagnostic and should not be interpreted as four independent headline causal effects. Panel B reports point-identified beyond-GN recipient counts among regular Female-Only recipients together with the corresponding all-female partial-identification bounds from Table A30. Conventional significance stars are omitted.

**Table A33.** Full Predetermined Covariate Continuity Battery at the 14 Percent Threshold

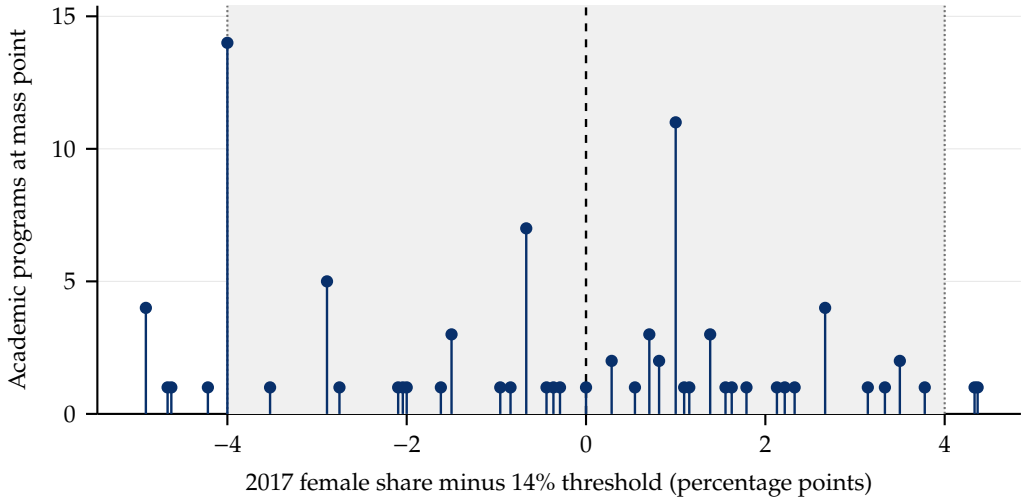
| Predetermined characteristic | ±3 | ±4 | ±5 |
|---|---|---|---|
| *Panel A. 2017 program capacity and category composition* | | | |
| 2017 program capacity | 18.699 (16.920); 0.295 | 5.976 (12.122); 0.629 | 10.158 (16.183); 0.540 |
| Log 2017 program capacity | 0.445 (0.364); 0.249 | 0.205 (0.285); 0.483 | 0.271 (0.341); 0.440 |
| OPEN female allotment share, 2017 | -0.406 (1.056); 0.709 | 0.050 (1.225); 0.968 | -0.027 (1.304); 0.984 |
| OPEN share of regular non-PwD capacity, 2017 | -0.583 (1.190); 0.635 | -0.343 (1.212); 0.781 | 0.011 (0.968); 0.991 |
| OBC-NCL share of regular non-PwD capacity, 2017 | 1.383 (1.692); 0.433 | 1.796 (1.451); 0.235 | 1.336 (1.136); 0.258 |
| SC share of regular non-PwD capacity, 2017 | 0.598 (0.539); 0.293 | -0.336 (0.816); 0.686 | -0.136 (0.732); 0.855 |
| ST share of regular non-PwD capacity, 2017 | -1.398 (1.061); 0.217 | -1.117 (0.807); 0.186 | -1.211 (0.758); 0.131 |
| Programs | 52 | 80 | 89 |
| *Panel B. 2016 historical program characteristics* | | | |
| 2016 program capacity | 23.379 (24.167); 0.362 | -4.905 (16.820); 0.775 | 0.255 (20.271); 0.990 |
| Log 2016 program capacity | 0.493 (0.463); 0.318 | -0.067 (0.334); 0.843 | 0.019 (0.396); 0.962 |
| Female allotment share, 2016 | -2.916 (2.272); 0.235 | -2.903 (2.529); 0.270 | -1.897 (2.003); 0.360 |
| OPEN female allotment share, 2016 | -1.497 (2.914); 0.621 | -0.843 (2.497); 0.741 | -0.056 (2.211); 0.980 |
| OPEN share of regular non-PwD capacity, 2016 | 1.641 (1.056); 0.159 | 0.804 (0.628); 0.222 | 0.554 (0.532); 0.316 |
| OBC-NCL share of regular non-PwD capacity, 2016 | -0.180 (0.745); 0.815 | 0.020 (0.623); 0.975 | -0.024 (0.456); 0.959 |
| SC share of regular non-PwD capacity, 2016 | -0.839 (0.688); 0.257 | 0.130 (0.651); 0.845 | 0.390 (0.514); 0.461 |
| ST share of regular non-PwD capacity, 2016 | -0.622 (0.849); 0.485 | -0.954 (0.542); 0.100 | -0.919 (0.509); 0.093 |
| Programs | 40 | 72 | 81 |
| *Panel C. 2015 historical program characteristics* | | | |
| 2015 program capacity | 18.534 (24.945); 0.482 | -10.437 (19.595); 0.604 | -4.136 (22.558); 0.858 |
| Log 2015 program capacity | 0.387 (0.490); 0.455 | -0.157 (0.377); 0.685 | -0.054 (0.435); 0.902 |
| Female allotment share, 2015 | -8.088 (5.448); 0.181 | -3.471 (5.904); 0.568 | -1.391 (3.887); 0.727 |
| OPEN female allotment share, 2015 | -10.218 (8.081); 0.247 | -3.411 (7.699); 0.666 | -1.113 (5.655); 0.847 |
| Programs | 38 | 67 | 76 |
| *Panel D. Predetermined Gender-Neutral admission frontiers, 2017* | | | |
| Log GN closing rank, OPEN, 2017 | -0.081 (0.327); 0.809 | 0.162 (0.260); 0.543 | 0.208 (0.234); 0.388 |
| Log GN closing rank, OBC-NCL, 2017 | -0.075 (0.288); 0.798 | 0.162 (0.227); 0.486 | 0.194 (0.209); 0.367 |
| Log GN closing rank, SC, 2017 | -0.019 (0.257); 0.944 | 0.169 (0.221); 0.457 | 0.259 (0.190); 0.192 |
| Log GN closing rank, ST, 2017 | -0.150 (0.224); 0.518 | -0.029 (0.236); 0.905 | 0.138 (0.178); 0.450 |
| Programs | 52 | 80 | 89 |
| *Panel E. Historical Gender-Neutral admission frontiers, 2016* | | | |
| Log GN closing rank, OPEN, 2016 | -0.307 (0.327); 0.374 | 0.177 (0.309); 0.575 | 0.217 (0.264); 0.426 |
| Log GN closing rank, OBC-NCL, 2016 | -0.293 (0.282); 0.328 | 0.168 (0.295); 0.577 | 0.229 (0.249); 0.373 |
| Log GN closing rank, SC, 2016 | -0.313 (0.268); 0.275 | 0.170 (0.291); 0.567 | 0.271 (0.239); 0.276 |
| Log GN closing rank, ST, 2016 | -0.189 (0.223); 0.421 | 0.070 (0.307); 0.822 | 0.275 (0.252); 0.294 |
| Programs | 40 | 72 | 81 |
| *Panel F. Historical Gender-Neutral admission frontiers, 2015* | | | |
| Log GN closing rank, OPEN, 2015 | -0.294 (0.289); 0.343 | 0.109 (0.307); 0.729 | 0.200 (0.253); 0.445 |
| Log GN closing rank, OBC-NCL, 2015 | -0.305 (0.249); 0.260 | 0.142 (0.316); 0.661 | 0.209 (0.253); 0.424 |
| Log GN closing rank, SC, 2015 | -0.288 (0.239); 0.266 | 0.063 (0.306); 0.840 | 0.220 (0.226); 0.349 |
| Log GN closing rank, ST, 2015 | 0.044 (0.286); 0.883 | 0.196 (0.292); 0.515 | 0.357 (0.198); 0.096 |
| Programs | 38 | 67 | 76 |

*Table A33 continued*

| Predetermined characteristic | ±3 | ±4 | ±5 |
|---|---|---|---|

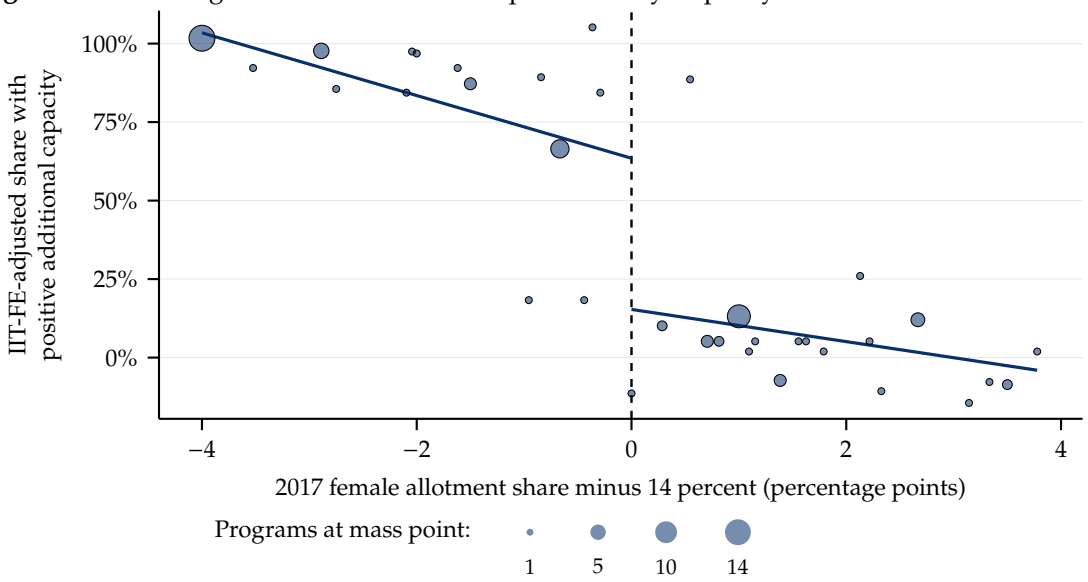
*Notes:* Each entry reports the estimated discontinuity, with the IIT-clustered standard error in parentheses and the corresponding $p$-value after the semicolon. Estimates are taken directly from the certified FP05B within-IIT predetermined-balance battery. Within each bandwidth, the sample retains only IITs represented by academic programs on both sides of the 14 percent threshold; this restriction is re-imposed after observations with missing values of the relevant covariate are removed. All specifications are local linear with separate slopes on either side of the threshold and include IIT fixed effects. Positive estimates indicate a higher value immediately below the threshold. The 2017 female share used to construct the running variable, IIT-level characteristics, 2018 treatment variables, and post-policy outcomes are excluded by construction. The ±4-percentage-point window is the frozen preferred bandwidth.

**Figure A2.** Running-Variable Support around the 14 Percent Threshold



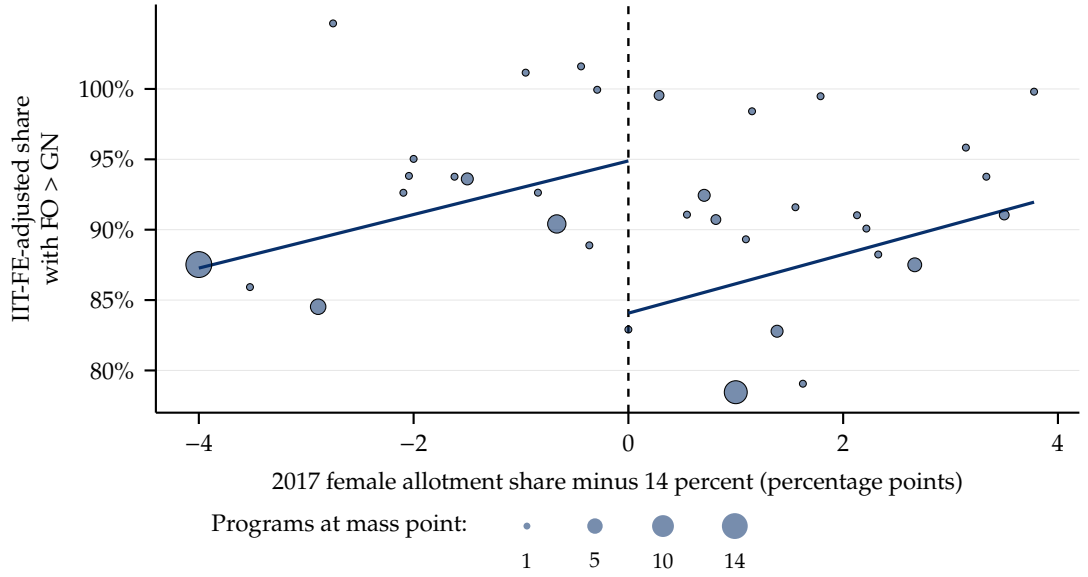

*Notes:* Each stem is a distinct realized value of the running variable, a program's 2017 female allotment share minus 14 % points, its height is the number of academic programs at that value. Because running variable is a ratio of integer female allotments to integer program capacities, its support is discrete: 89 programs occupy 41 distinct values. The shaded band bounded by dotted lines is the frozen ±4-percentage-point estimation window, which contains 80 programs at 35 mass points, split evenly into 40 below the threshold and 40 at or above it. The single program lying exactly at 14 percent is coded as untreated.

**Figure A3.** First Stage: Additional Female-Supernumerary Capacity at the 14 Percent Threshold



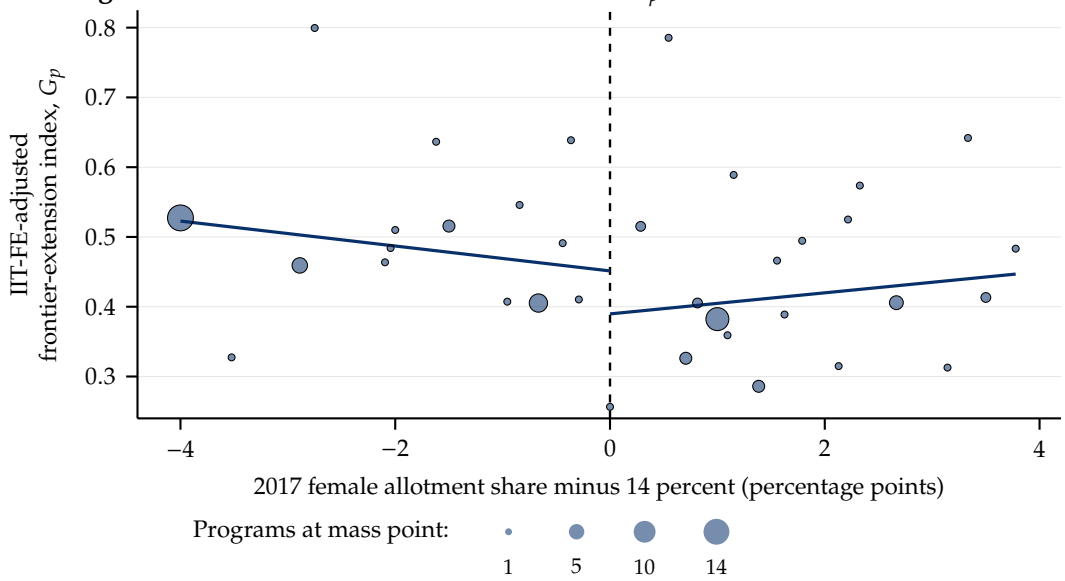

*Notes:* The figure plots the first stage of the frozen ±4-percentage-point design around the 14 percent 2017 female-allotment-share threshold: 80 programs at 35 mass points across 16 IITs, 40 below the threshold and 40 at or above it. The outcome is $D_p = \mathbf{1}\{x_p > 0\}$, an indicator for receiving positive genuinely additional female-supernumerary capacity in 2018. Throughout Figures A3–A8, points are mass-point means after removing IIT fixed effects and recentering to preserve the sample mean, and fitted lines use the same local-linear specification with separate slopes; each displayed discontinuity therefore equals the corresponding formal estimate. Here that estimate is 0.482 (IIT-clustered SE 0.242), with first-stage $t^2 = 3.97$. The instrument equals one only for programs strictly below 14 percent, so the program exactly at the threshold is coded untreated.

**Figure A4.** Positive Observed Female-Access Extension at the 14 Percent Threshold



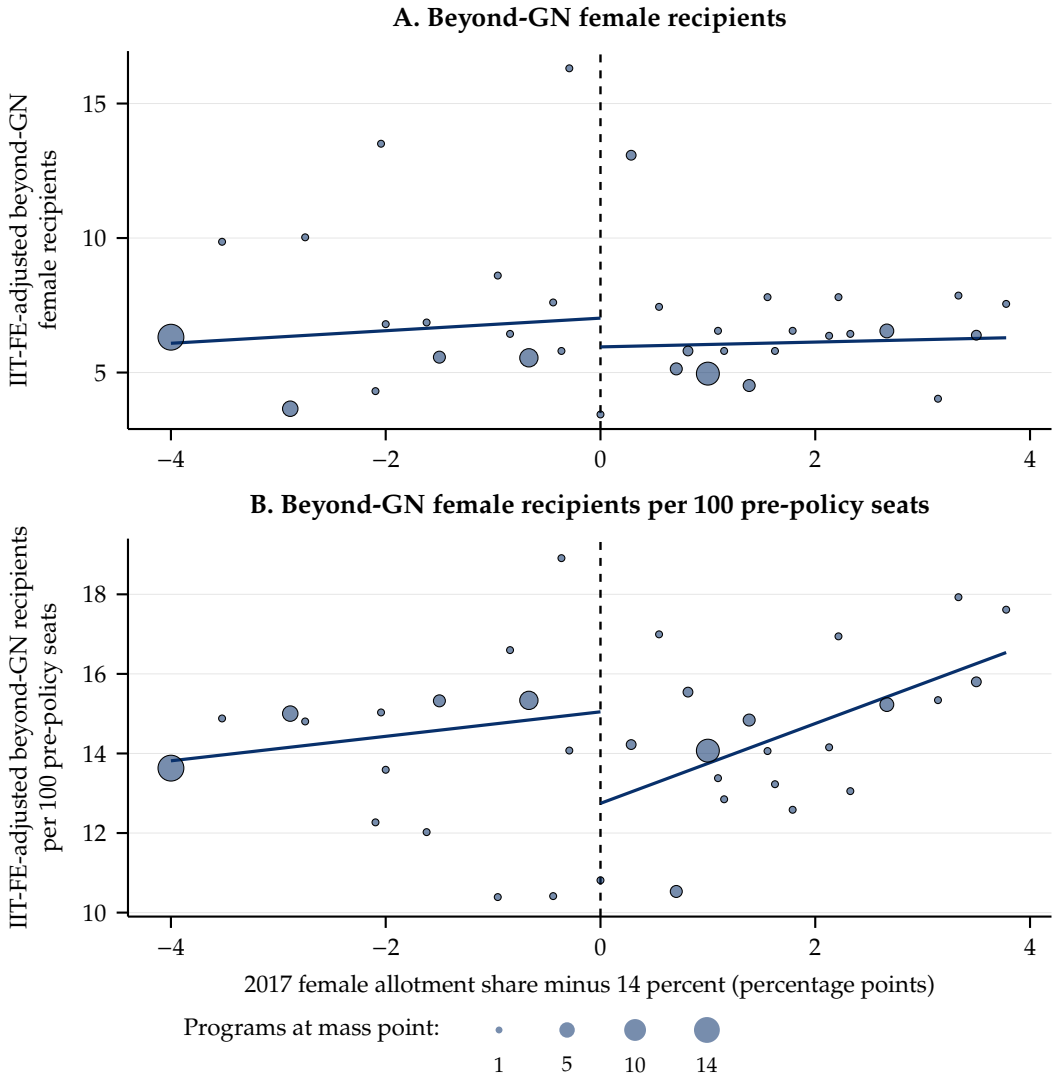

*Notes:* The outcome is the predetermined-2017-capacity-weighted share of regular reservation-category routes on which the observed Female-Only (FO) closing-rank frontier extends beyond the corresponding Gender-Neutral (GN) frontier. The sample and the IIT-fixed-effects adjustment are as in Figure A3 (80 programs, 35 mass points, 16 IITs). The displayed discontinuity equals the preferred reduced form, 10.8 percentage points (IIT-clustered SE 4.0); the structural IV estimate is 22.5 percentage points, with Anderson–Rubin $p = 0.017$ for $H_0 : \beta = 0$. Formal estimates are reported in Tables 5 and A26.

**Figure A5.** Observed Frontier-Extension Index $G_p$ at the 14 Percent Threshold



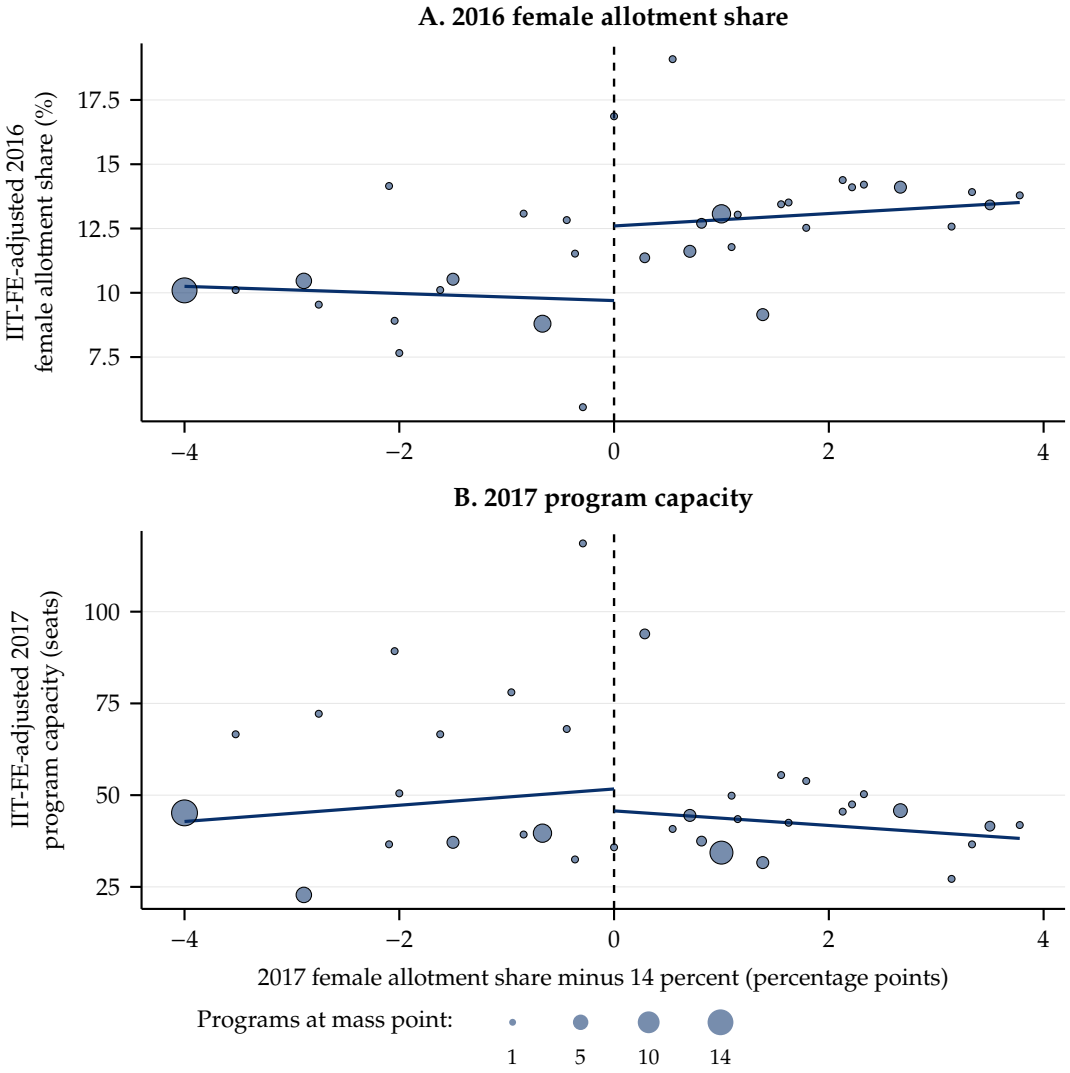

*Notes:* The outcome is $G_p$, the predetermined-2017-capacity-weighted magnitude of positive observed female-specific frontier extension, constructed from $\max\{0, \log(CR^{FO}/CR^{GN})\}$ over observed finite regular-category frontiers. The sample and the IIT-fixed-effects adjustment are as in Figure A3 (80 programs, 35 mass points, 16 IITs). The displayed discontinuity equals the preferred reduced form, 0.062 (IIT-clustered SE 0.074); the IV estimate is 0.128 (SE 0.131), implying a 13.6 percent change, but the Anderson–Rubin $p$-value is 0.418. Formal estimates are reported in Tables 5 and A26.

**Figure A6.** Realized Beyond-GN Allocations at the 14 Percent Threshold



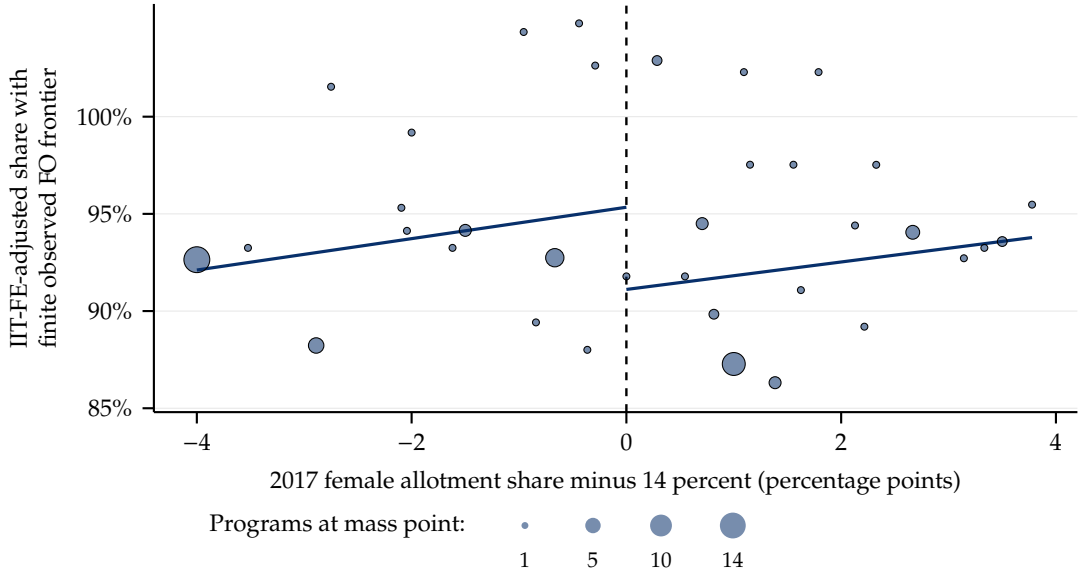

*Notes:* Panel A reports the program-level count of point-identified regular Female-Only recipients whose allocated physical program lies beyond every eligible observed Gender-Neutral route; Panel B divides that count by predetermined 2017 program capacity and reports recipients per 100 pre-policy seats. The sample and the IIT-fixed-effects adjustment are as in Figure A3 (80 programs, 35 mass points, 16 IITs). The displayed discontinuities equal the preferred reduced forms, 1.069 recipients (IIT-clustered SE 1.706) and 2.301 recipients per 100 seats (SE 2.401); the IV estimates are 2.220 and 4.778, with Anderson–Rubin $p$-values of 0.540 and 0.353. These outcomes compare realized 2018 allocations with observed GN frontiers and are not re-solved no-policy counterfactuals. Formal estimates are reported in Tables 5 and A31.

**Figure A7.** Predetermined-Covariate Continuity at the 14 Percent Threshold




*Notes:* Both panels plot predetermined program characteristics that could not have been affected by the 2018 reform. Panel A reports the 2016 female allotment share for the 72 programs with certified observed 2016 history, spanning 15 IITs and 34 mass points; Panel B reports 2017 program capacity for all 80 programs in the preferred sample, spanning 16 IITs and 35 mass points. Points and fitted lines are constructed as in Figure A3. The displayed discontinuities equal the formal continuity estimates: −2.903 percentage points (IIT-clustered SE 2.529; $p = 0.270$) for the 2016 female allotment share and 5.976 seats (SE 12.122; $p = 0.629$) for 2017 capacity. The full predetermined-covariate continuity battery is reported in Table A33.

**Figure A8.** Observed Finite Female-Only Frontier Activation at the 14 Percent Threshold




*Notes:* The outcome is the predetermined-2017-capacity-weighted share of regular reservation-category routes for which a finite Female-Only (FO) closing-rank frontier is observed. The sample and the IIT-fixed-effects adjustment are as in Figure A3 (80 programs, 35 mass points, 16 IITs). The displayed discontinuity equals the preferred reduced form, 4.2 percentage points (IIT-clustered SE 4.5); the IV estimate is 8.8 percentage points (SE 11.7), with Anderson–Rubin $p = 0.360$. Absence of a finite observed FO closing rank is an observability outcome, not evidence that the program had zero institutional FO capacity. Formal estimates are reported in Table A26; category-specific frontier coverage is reported in Table A28.