Essays in Labor Economics and Industrial Organization

by

Iris Vrioni

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Economics)
in the University of Michigan
2024

Doctoral Committee:

Professor Ying Fan, Co-Chair
Professor Basit Zafar, Co-Chair
Professor Zach Brown
Professor Kevin Stange

Iris Vrioni
ivrioni@umich.edu
ORCID iD: 0000-0003-2895-9363

ACKNOWLEDGEMENTS

I am eternally grateful to my advisors Basit Zafar and Ying Fan for their wisdom, kindness, and dedication. They have given me the kind of life-changing advising that I never thought I would have the fortune to receive. I admire them immensely and am thankful to have had them in my corner as they pushed me to be a better, more productive, and more rigorous researcher.

In addition to my principal advisors, I am deeply grateful to Zach Brown for the many half-hours spent discussing my work over the years, for the practical advice and careful comments, and for being a role model. Martha Bailey gave me the opportunity to contribute to her monumental undertaking to understand the fertility choices of low-income women in the challenging U.S. policy environment. Our joint paper and the larger project were my first taste of meaningful academic work and I feel thankful to have been her student and research assistant. I had several pivotal conversations with Kevin Stange, John Bound, and Charlie Brown, and many practice presentations in the Labor and Industrial Organization Lunches at the University of Michigan.

The first chapter of this dissertation would not exist without Arjan Xhelaj, who made it possible for me to use data from the Academic Network of Albania. I am grateful to Don Carleton for allowing access to the Putnam Records in the Archives of American Mathematics, and to Carol Mead and staff at the Briscoe Center at UT Austin for taking a chance with a graduate student wanting to digitize loads of archival data for the first time. Their help accessing and navigating the Putnam records made the third chapter of this dissertation possible.

At the University of Michigan, I was fortunate to have exceptional friends. Emir Murathanoglu, Brock Rowberry, and Amy Ciardiello shared in my every high and low from the first to the last day of the PhD. My happiest times were the times we spent together. Hanna Onyshchenko, Carolina Santos, Agostina Brinatti, Chris Hollrah, Emily Horton, Nikhil Rao, Lea Bart, Katherine Richard, Tyler Radler gave me, in order, a second sister, my favorite hobby (running), a role model in discipline and calm, excellent afternoon coffee every day, perspective when I needed it, reliable office company and a tremendous example of resilience in research, an example of great boundaries, a better sense of style, a higher caliber chess opponent than I am.

Before the PhD, I received help and encouragement to pursue a PhD from Sarah West, Raymond Robertson, and Chad Topaz at Macalester College. Lorela Paco, my best friend since our first year at Macalester, remains a constant source of counsel and support.

Finally, I want to thank the most important people to me: my parents Monika and Arben, who have devoted their lives to the betterment of their children's, my sister Ana and her partner Klajdi, and my grandmother Hëna. They have given me immeasurable love and support through the PhD and my whole life. This dissertation is dedicated to them.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE

LIST OF TABLES

LIST OF APPENDICES

ABSTRACT

Chapter 1 studies how the interaction of student information with constraints dictated by market design determines higher education choices and outcomes. I study strategic application incentives in imperfect implementations of centralized assignment mechanisms in higher education. I ask whether, in markets with both a central match for public colleges and a broader private market, choices on the match are affected by the availability good private outside options. It is unclear whether the common market configuration with outside private options and application size restrictions generates strategic incentives in applications on the public match that is advantageous to students with higher socioeconomic backgrounds. I assemble data from the college match in Albania and utilize a policy change that incorporated all private colleges in the centralized platform to generate insight. The policy does two things: first, it removes private programs as an outside alternative to the match, differentially shifting outside options of students of different backgrounds. Second, it expands the set of choices on the match among which students must choose, tightening the constraint on application sizes for students who can afford private colleges and must substitute away from applying to public college to accommodate good private options in their application.

I find that ambitiousness in public college applications declines after policy implementation, and more for private high school students, supporting the hypothesis of non-truthful application behavior that is responsive to market structure. I then build a model of applications and enrollment behavior that accommodates strategic behavior and can disentangle the effects of heterogeneous beliefs, preferences, and outside options on choice to evaluate the distributional consequences of the interaction between non-truthful applications and market partitioning. I find substantial uncertainty about admission thresholds, in part responsible for strategic applications. The application behavior induced by a single match for all schools with list size restrictions worsens outcomes on average, especially for private school students, both for those who would change their applications, and indirectly through crowd-out for those who would not. The main channel is list size restrictions becoming more binding with more options on the match, but some of the effect is due to outside options becoming less good. The relative worsening of outcomes for higher-SES students may be redistributive, but better outcomes can be achieved for both groups under a single match by slightly extending

application lists.

In Chapter 2, joint with Martha J. Bailey, Vanessa Wanner Lang, Alexa Prettyman, Lea J. Bart, Daniel Eisenberg, Paula Fomby, Jennifer Barber, and Vanessa Dalton, we study the consequences of costliness in fertility regulation in the current US policy environment, which leaves 1.4 million uninsured Title-X clients with substantial cost sharing for contraception and reproductive health. We experimentally vary contraceptive subsidies to women in Planned Parenthood clinics of Michigan and find a substantial response to contraceptive cost, in particular for high-fixed cost methods. Mothers are the most financially constrained, but all groups increase their take-up. Our take-up estimates imply that a U.S. policy eliminating out-of-pocket costs for all Title X patients would reduce pregnancies by 5.3%, birth rates by 3.9%, and abortions by 8.3%.

Finally, Chapter 3 studies the role of information and beliefs that supervisors have in determining individuals' education and career trajectories. In it, I investigate whether academic supervisors have differential information quality about their male and female students and ask how supervisors learn about skills of individuals and groups. In particular for students of high talent, early-career signals may have important consequences on future opportunities, career trajectory, and intellectual output. I assemble new archival data on all participants of the Putnam Mathematical Competition over four decades to make progress on the above questions. The setting allows me to observe an objective measure of talent in the scores achieved by each student and the supervisors' prediction about the talent ordering of students they supervise. I find that ex-ante, for women and men with the same ex-post competition scores, supervisors expect women to do worse than their male peers from that college. Women are less likely than men to have been predicted to be in the top three performers of their school even when they obtain a score that places them in the top three. Female supervisors are no better at predicting female achievement than male supervisors. I find evidence that supervisors learn about individual women, but little evidence of supervisor learning about the group. With a model of learning, I investigate whether supervisors are biased in their perceptions of gender talent or hold beliefs consistent with the empirical distribution of performance by gender.

CHAPTER 1

Inequity in Centralized College Admissions with Public and Private Universities: Evidence from Albania

1.1 Introduction

The importance of fairness and efficiency in assigning students to scarce public school or university seats has led to an increase in the practice of centralized admissions at all levels of education across the world. In higher education alone, every year, over 20 million students are matched to colleges and majors through centralized mechanisms.[1]

Two prominent features of most implementations of centralized mechanisms are: (1) list size restrictions and (2) the exclusion of some institutions, generally private ones, from the match. List size restrictions limit the number of programs one can apply to through the mechanism and force students to weigh their preferences and ambitions against the chances of admission, inducing them to strategize on what programs to include in their application. At the same time, private universities manage their admissions outside of the mechanism and can serve as much as 75% of the market.[2] In this context, for applications within the

[1]Around 10 million students participate in the college match in China (Chen and Kesten, 2017), more than 2.5 million in Brazil (Otero et al., 2021), and millions more in Chile, Germany, India, Kenya, and Turkey, among others.

[2]In Brazil, more than 75% of students enrolled in a college degree attend a private institution. In my empirical setting, 27% of students are enrolled in private universities.

match, it may matter strategically what outside options a student has. Those with better outside options can apply more ambitiously within the mechanism and may ultimately be assigned to more preferred programs than students with worse outside options. This strategic response to market structure may have both efficiency and equity consequences. If outside alternatives, which are often expensive, are more desirable for high-SES students, it gives them not only higher direct value from choosing these options, but also the ability to take more risk within the match.

This paper assesses the importance of market structure, in particular the extent of centralization, on strategic applications in centralized assignment systems, focusing on Deferred Acceptance (DA) mechanisms with list size restrictions. With data from a market that changed its structure from partial centralization with a public match and private decentralized admissions to a fully centralized system, I build a model of student applications and enrollment decisions to college-major pairs ("programs") on the platform. The model and estimation take advantage of the unique features of the setting, which allow separate identification of student beliefs about chances of admission and their preferences for programs. The level of uncertainty and bias about the expected selectivity of programs are central in determining how constraining the list size restrictions are. In addition, preferences for inside and outside options determine the extent to which it is important to students to make sure they are admitted to an acceptable program in the match rather than take a chance with their favorite programs. I quantify these elements and assess choice, resulting allocations and welfare in commonly observed market structures.

My empirical setting is the centralized admissions system in Albania, which underwent a unique change that incorporated all private colleges into the centralized admissions process. Prior to 2016, the Albanian college admissions were representative of common partially centralized systems with a market that was partitioned into a public match with decentralized private college admissions that happened roughly at the same time. Assignments to public universities were mediated by a clearinghouse, which took in stated student preferences for programs as well as program priorities.[3] Student preferences were reported through rank order lists restricted to 10 or fewer programs, and the clearinghouse produced matches through a standard DA algorithm. In 2016, a major higher education reform changed admissions in two significant ways. First, all private colleges got incorporated into the centralized match with no programs allowed to conduct admissions outside of the national match. This expanded options on the national platform by over 50% with list size restrictions remaining the same as before the reform, and any possibility of enrolling in college outside of the match was eliminated. Second, the reform changed the assignment procedure to a live multi-offer mechanism with 7 phases in its main round. Students submit their application lists to the clearinghouse and programs submit their priorities. Then, as a program-proposing Gale-Shapley algorithm would begin, in the first phase, programs make initial proposals to students ranked at the top. Under the new procedures, students observe all offers and decide within 48 hours whether to enroll in any of the proposing programs or forgo all first phase offers and wait for a better offer in the next phase. The multi-offer phases continue until the last phase or until each school has filled its seats.

This setting is attractive because it allows me to overcome two key challenges that have so far prevented a clear answer on the extent to which outside options matter for strategic applications in centralized mechanisms. The first is that of observing choice over in- and out-of-match programs in order to infer preferences for private programs. Data on applications and enrollments are generally only available for programs on the match, which makes it impossible to understand the value of outside options to students. I overcome this challenge with data from the post-reform period in which choices to apply and enroll in all programs are made on the platform.

The second and more complex challenge is one of identification. In a setting with constrained applications, both preferences for programs and perceived probabilities of admission play a crucial role in determining the extent to which market structure affects on-platform

[3]Program priorities come in the from of weights for GPA and end-of-high school exams that the clearinghouse uses to produce weighted average scores for each student applying to each program.

applications. It is necessary to separate preferences from student beliefs about probabilities of admission. If students have perfect information about which schools on the match they would be admitted to, there is no scope for strategic applications in general and no role for outside options to affect applications through the strategic channel in particular. The importance of private outside options is tied to the level of uncertainty and bias that students have about their chances of admission to centralized programs. In most centralized settings the only observed choice is the selection of application lists, which are insufficient to separately identify preferences from beliefs about admission chances (Agarwal and Somaini, 2020). The school choice literature often makes strict assumptions about belief formation where agents have rational expectations about program cutoffs (Agarwal and Somaini, 2018; Idoux, 2022). This assumption is inadequate to evaluate the question because it drastically limits the ability of the model to explain choices as arising from a strategic channel.[4] I instead allow beliefs to depart form rational expectations and capture the uncertainty and bias in the market in a reduced form. I overcome the identification issue by taking advantage of the post-reform mechanism, which allowed multiple offers at the admissions phase. Students observe which programs have admitted them, and many are admitted to 2 or more programs in their choice set. The choices to enroll and which program students enroll in pins down preferences, and the application portfolio choice and decisions to wait for future phases can be exploited to identify beliefs about probabilities of admission.

Using rich data on applications and enrollments in years 2013-2019, I first provide descriptive evidence that when private options are available outside the centralized match, high-SES students apply to and enroll in more selective public programs than their lower-SES peers with the same high school and exam performance. High-SES students are also more likely to end up without an assignment in the public match. These are striking facts because these differences cannot be explained by geographic access to more selective programs for high-SES

[4] Non-degenerate beliefs about distribution of cutoffs from which perceived probabilities of admission arise exist only because of sampling variation and hence probabilities of admission depend only on student preferences and the size of the market, holding the mechanism constant.

students or lack of affordability of selective public programs given that public institutions are tuition-free.

I then analyze the policy change that enforced participation by all colleges in the platform. I use an event study design to measure the effects of centralizing all available alternatives on application behavior. Comparing applications of high-SES students to lower-SES students before and after the reform, I find that high-SES students change their applications more: they reduce the number of public programs they apply to by 1.2 more than lower-SES students. In addition, the selectivity gap in applications between the two groups shrinks after the policy change, both for the overall portfolio and for the reach programs. The shrinkage is driven by high-SES students decreasing the selectivity of their applications by more than lower-SES students do. Finally, the variance in the selectivity of public programs in the applications declines, indicating that students are giving up more selective programs rather than shifting the entire application toward less selective programs.

Motivated by the reduced-form results, I quantify the welfare and distributional impacts of a partitioned market structure. Based on the features of the post-reform period, I build a structural model of student decision to apply to college, application portfolio selection, and enrollment and waiting decisions on the waitlist. In the model, students take national exams and after observing their score, decide whether and where to apply to college. Each graduating high-schooler applies through the match if there is at least one program on the platform that they prefer to their outside option. Crucially, applications are allowed to be strategic: students are allowed to prefer more than ten programs to their outside options, but are only allowed to apply to ten, which induces them to exclude certain programs in such a way that the resulting portfolio maximizes their expected utility from the lottery over outcomes induced by the application portfolio. Portfolios are constructed as in Chade and Smith (2006)—students understand that the marginal value of each option included in the portfolio depends not only on the probability of admission to this program and the value of attending it, but also on the admission probabilities and value from all the other choices on the portfolio. Therefore portfolios are chosen as an optimization problem over all possible lotteries induced by portfolios of size ten.

Students form beliefs over probabilities of admission that depend only on the final cutoff at the last phase of admission and I assume they disregard the distribution of possible cutoffs in intermediate stages of the mechanism. This assumption is consistent with the information available to students at the time of application and alleviates the intractability problem that arises from the fact that each portfolio choice induces a distribution of waitlist states in each of the 7 rounds of these dynamic admissions.[5] I model beliefs over cutoffs as based on the previous year's cutoff for that program. Students expect a mean shift from the previous year's cutoff and have uncertainty over the realization of cutoffs, which I model as a distribution in possible cutoffs that is centered around the shifted previous year's threshold. The scale of distribution is allowed to be heterogeneous by SES group and by program selectivity. By parameterizing beliefs as a normal distribution over program cutoffs with a shift and a scale parameter, I capture a very complicated multi-dimensional object with data and a small set of parameters.

I assume that choices on the first phase of admissions are made with the same information and preferences as those in the application stage. While it is possible to model learning in this context, I find that differences in the first-phase cutoffs from the previous year do not predict the likelihood of students accepting first-phase offers or the likelihood of waiting for the next phase. At this stage, applicants observe offers and choose whether to accept a given offer, wait for the next phase, or exit the mechanism unmatched. Choices to enroll in programs in this phase offer the main source of identification for student preferences for

[5]The only paper to model a multi-stage mechanism that resembles the one in my setting is Waldinger (2021). In that problem, a two-stage mechanism of housing development choice is modeled in which the first stage consists of applications of size up to 3 among 18 possible choices and each portfolio generates a distribution of possible waitlist positions in the second stage, each of which generates a distribution waiting times for chosen developments. By contrast, students in my setting choose 10 options among over 500 in the first stage which induces a distribution of waitlist positions in the first admission stage, and from there, each preceding vector of waitlist states induces a distribution of waitlist states in the following phase. Accounting for these dynamic considerations becomes quickly intractable. The platform authority itself avoids distributing information about intermediate program cutoffs at the time of application with the goal of discouraging students from considering intermediate stages.

programs.

I estimate the model by Simulated Maximum Likelihood. The model does not generate a closed form expression for the likelihood of the sequence of observed student actions. In particular, a major challenge in estimation is computing the conditional likelihood of the chosen portfolio being optimal. Given the large number of available options, it is infeasible to compute the probability of an observed portfolio being optimal among those in the vast choice set. I use a method developed in Larroucau and Rios (2020) which derives, for the portfolio problem of the Chade and Smith (2006) type, a small sufficient set of deviations from the observed portfolio that need to be checked for optimality. This allows for both the tractable simulation of choice sets at the application stage and an estimation routine that maximizes a likelihood function that is not prohibitively flat.

Estimates of the model imply that students face significant uncertainty about their admission probabilities in this market. Both high and lower-SES students perceive on average a lower probability of being accepted to each program than they would have with perceived distributions centered at the previous year's cutoffs. The mean of the belief distribution for high-SES students is shifted further up the range of cutoffs implying they are slightly more optimistic for lower-cutoff programs, but the slope of the mean is less steep than for the lower-SES group. These estimates suggest that student information is far from perfect. It is crucial then to evaluate both the role of application constraints and how students' outside options interact with the platform application constraints.

With estimated taste and belief parameters, I conduct counterfactual analyses that evaluate the role of market structure on applicant behavior, allocations of students to programs and welfare. I assess a centralizing policy change (an "all-in" structure) from a market where all outside alternatives are private (a "partitioned" structure). In the partitioned structure, both the centralized and decentralized admissions operate simultaneously but separately in a single-stage application with DA assignment. In reality, decentralized markets suffer from congestion and matching frictions which can affect assignments in both the public match and the private market (Abdulkadiroğlu et al., 2017; Kapor et al., 2022), but the scope of this paper is not to assess such frictions.[6] The assumption of an unrestricted-list DA mechanism for the private market serves to abstract away from these frictions to focus on the effect of the strategic channel for platform applications.

I find that centralization with list-size restrictions reduces enrollment by 4.2pp (1.7pp) for lower (high)-SES students relative to partial centralization. The allocation into private and public programs changes too. For the students that go to college, assignments worsen for some and improve for others with net losses for 2.9% of high-SES students and net wins for 1.1% of lower-SES students. Welfare calculations also imply a net loss in the market, which accumulates mostly in the high-SES group, leading to slightly improved equity at a high cost of efficiency. I measure welfare relative to the gains possible under an unrestricted DA mechanism and find that an all-in policy increases the welfare gap relative to the unrestricted DA by €160 for high-SES students and €98 for lower-SES students.

I investigate the channels next. First, the direct response from a less valuable outside option induces some students to exclude a public program they would have been marginally admitted to and would have preferred relative to the assignment they get. This accounts for a minority of those with worse outcomes. Second, with outside options incorporated in the mechanism, there are more programs to choose from in addition to a lower-valued outside option for many, which induces some students to reduce the number of public programs in their list in order to accommodate private programs. This channel accounts for the majority of the losers. Third, programs' capacity constraints generate spillovers from those who change their application through the first two channels. Some students do not change their applications and get pushed to less preferred programs because more have applied to their otherwise feasible program. The first two channels are much stronger for high-SES students, while the third affects everyone. The lessons from this decomposition imply that the

[6] Other work has studied these frictions. Most recently, Kapor et al. (2022) find frictions that come from chains of on-platform offer rejections in favor of off-platform options that lead to vacancies in platform programs or mismatches as platform programs try to contact individuals that were initially rejected.

constraint channel is far stronger than the outside-option channel in determining the effects of the market structure on applications and assignments of students. The inequity in observed outcomes with the partitioned structure is due to the fact that high-SES students behave almost as though unconstrained while constraints are more binding for lower-SES students. The all-in market structure forces binding list-size constraints on high-SES students, reducing inequity, but also efficiency.

Since the effect of market structure through the strategic channel is largely determined by constraints rather than outside options, I consider market designs that keep the all-in structure, but alleviate list-size restrictions. I find that an increase in list size of just 4 additional slots recovers more than half the losses from market structure change for lower-SES.

This paper contributes to the empirical literature on implementations of centralized assignment. A small body of work documents welfare implications of different aspects of common implementations of matching (Abdulkadiroğlu et al., 2017; Calsamiglia et al., 2020; Fack et al., 2019). Luflade (2017) uses a setting in Tunisia to show that imperfect information on probabilities of admission affects strategic applications. In the same vein, Ajayi and Sidibe (2020) use a setting in Ghana to estimate welfare effects of changing the allowed number of applications students can submit to the centralized mechanism. Most similarly to my setting, Kapor et al. (2022) use a centralized platform expansion in Chile to evaluate the welfare consequences of matching aftermarkets. My paper provides the first empirical evidence of the interaction between market structure and strategic applications.

My paper also contributes to the theoretical market design literature studying incentives generated by outside options in manipulable mechanisms as well as market structure. Akbarpour et al. (2021) formalize theoretical predictions for the effect of outside options on manipulable centralized mechanisms with an application to elementary school matching. Andersson et al. (2019) study the implications of a sequential public and private school matching mechanism. I add empirics to this largely theoretical literature to assess the practical importance of outside options in partially centralized settings.

Finally, I add to the literature studying educational decisions under imperfect information. Kapor et al. (2020) use surveys to elicit beliefs about chances of admission to schools and find that students and parents have heterogeneous beliefs that depart from rational expectations. Luflade (2017) estimates beliefs that rationalize untruthful applications in a DA mechanism while extrapolating preferences from a small subset of plausibly truthful applicants. My paper is the first to jointly estimate preferences and beliefs in a centralized mechanism setting.

This paper is organized as follows: Section 1.2 describes the Albanian college admissions, policy variation, and data. Section 1.3 provides descriptive facts on application patterns that are consistent with predictions from a simple model of length-restricted college applications in which higher-SES students have better outside options. Section 1.4 analyzes the effects of a centralizing policy. Section 1.5 then provides a full model of strategic application behavior and estimation details of its primitives, and Section 1.6 presents model results. With estimated model parameters, Section 1.7 analyzes the effect of market structure on application behavior, on the allocation of students to colleges and majors and welfare and assesses the relevance of strategy in determining outcomes for students under different market configurations. Finally, Section 1.8 concludes.

1.2 Background and Data

1.2.1 Higher Education in Albania

Higher education in Albania is delivered by 12 public and 26 private universities.[7] While public universities have always served the majority of students (73% of college enrollees in 2019), private universities have enrolled an increasing share of students over the past two

[7]These counts reflect the market in the period 2016-2019. More detail on public and private programs in the system can be found in Table 1.1.

decades.[8] The quality of the degrees varies for both types of universities with significant overlap. Programs offered by public universities in the capital are on average the most competitive and most likely to be oversubscribed. Figure A.1 displays the distribution of average scores for enrollees in public and private institutions. Tirana campuses of public universities enroll better performing students than private programs on average and several of the most competitive programs in the country are public. Public regional universities on the other hand tend to offer programs that are less competitive than private universities. Geographically, all private university campuses are located in the capital, whereas public college campuses are spread across most regions of the country. Public colleges are generally tuition-free, and impose a small fee for students to attend, whereas tuition to private colleges varies by college and major and ranges between \$500 and \$5000 per year with little need-based financial aid in the private system. While scholarships are offered in private colleges, they are merit-based and generally subject to strict score cutoffs for qualification.[9] Finally, as in many other countries, higher education is immediately specialized, with students making application and enrollment decisions to college-major pairs rather than just institutions.

1.2.2 Admissions procedures in the partitioned pre-2016 market

Before 2016, a national clearinghouse managed public college admissions alone. Admissions proceeded as follows. At the end of high school in June of each year, students took national exams (called Matura Exams) that included mandatory tests in math and Albanian language and elective subject tests in two subjects among those covered in the high school curriculum. After results of the national exams were announced, students began their appli-

[8]Private provision of higher education only became possible after the end of the communist regime in 1991. In particular since the early 2000's there was a proliferation of private for-profit higher education institutions. Concerns over the quality of these institutions led to a government crackdown on private for-profit colleges and the closure of 18 private universities. With more quality oversight, there have been 26 private universities that have operated between 2014 and today. See the closure order here: https://arsimi.gov.al/wp-content/uploads/2018/07/VKM_per_heqjen_e_licences.pdf.

[9]Scholarships are also offered for four additional categories independently of their academic performance: children of policemen killed in the line of duty, athletes with high achievement at the national level in their respective Olympic sport, members of the Roma/Egyptian community, and orphaned children from low-income families. These, however, are a very small number of scholarships for each program.

cation process to public programs through the national clearinghouse. Applicants submitted a rank-ordered list of up to ten public college-major pairs to the mechanism. On the colleges' side, programs ranked students through predetermined formulas that computed a weighted average of the high school GPA and scores in the national exams. These weighted averages would weight each component depending on the course content of the program.[10] Because the majority of public programs, in particular those in the capital, were oversubscribed, a Deferred Acceptance algorithm was run by the clearinghouse to allocate students to a single program.

Admissions to private programs, on the other hand, were decentralized and spanned a period beginning before the centralized public match and ending after it. Students were able to apply and enroll in a private program at any point during the public match. Since a significant portion of the students assigned to public university seats rejected their assignment for private off-match programs, the main phase of the match left universities with vacant seats that could be filled by students that would prefer those seats relative to their assignments.[11] Therefore, a supplemental assignment round was conducted in which all participants still present in the public match would be reallocated to a choice at least as highly ranked in their initial list as the one they were assigned to in the first round. Appendix A.2.1 describes in detail the timeline of admissions and allocation mechanism for the public match before 2016.

1.2.3 Admissions policy change

A higher education reform, which was signed into law in 2015, changed the configuration and operation of the college admissions market for the graduating high school class of 2016 and all following cohorts. Below I describe the features of the policy change:

[10]For example, the Mathematics degree at the University of Tirana would give a weight of 1.4 to a Math subject score in the national exam, and only a weight of 1 to the History subject score.

[11]This externality in the match generated by the partitioned nature of the market and private off-platform options is empirically evaluated in Chile by Kapor et al. (2022). I instead focus on externalities from off-platform options at the application stage rather than during the match process.

(1) All private universities joined the centralized platform. The policy change incorporated all private programs into the centralized application platform such that no students could gain admission to any program in the country without going through the centralized admissions process. Private colleges were required to produce their own criteria and formulas for admission, which would be made public on the clearinghouse website.[12] Figure 1.1 shows the expansion of the platform in 2016 from around 300 programs to over 500 programs.

The incorporation of all private programs into the platform was immediate and complete and the jump to 430 available platform programs in 2016 reflects this implementation. In the years that followed the reform, the number of on-platform programs continued to increase as universities introduced new programs, but there was no movement of programs into or out of the platform.

(2) List size restrictions remained the same. Despite the expansion of the platform, the number of programs students were allowed to apply to remained restricted to 10, the same as in the pre-reform period when the platform only had about 60% of the post-reform programs on it. Figure 1.2 shows a histogram of submitted application sizes before and after the reform. The list size restriction became more binding after the policy change with over 80% of students filling their lists relative to the 62% before the reform. Not only did students apply to more programs after the reform, but a larger share of national exam takers applied to college through the centralized platform among both high and lower-SES students (Figure A.2).

(3) The assignment procedure changed from DA to a multi-offer dynamic procedure. The reform was accompanied with a change in the mechanism that allocated students to programs. Instead of DA, the new mechanism is a multi-offer dynamic mechanism with 7 phases in its main round. As before, students submit their application lists of at most 10 programs to the clearinghouse and programs submit their priorities. Then, as a program-proposing Gale-Shapley algorithm would begin, in the first phase, programs make initial proposals to

[12]This website is also where college applications would be submitted and information about previous years' program cutoffs would be posted (www.ualbania.al).

students ranked at the top. In a college-proposing DA, the first step would produce matches between proposing programs and students that rank these programs first. Any offer from programs that students don't rank first would be rejected. Then in following steps any programs with empty seats would propose in order of priority to students without a match and matches would be produced in the same way as the first step until either all students have been assigned, or there are no seats available in programs with unmatched applicants. Under the new procedures, in the first phase, students observe all initial offers and decide within 48 hours whether to enroll in any of the proposing programs or forgo all first phase offers and wait for a better offer in the next phase. Matches are then produced with active participation by the students. Differently from a college-proposing DA, the first stage matches include not only matches between programs and qualified students that ranked them first, but also matches between qualified students and programs preferred relative to expected payoff from the remainder of the portfolio. All unmatched students and empty seats are then carried to the next phase of the mechanism and colleges propose in the same way as in the first phase. The multi-offer phases continue until the 7th phase of the main round or until each school has filled its seats. In Appendix A.2.2 I describe in detail the post-reform admissions procedures.

1.2.4 Data

I bring together application, assignment, offer, and enrollment data from several sources for three years just before the reform (2013-2015) and four years after the reform (2016-2019).

Pre-reform applications: For applications on the centralized system for the three years before the reform, I use applicant-level data from the Center for Educational Services of Albania (*Qendra e Shërbimeve Arsimore, QSHA*), which was the agency that administered the national high school exams and managed college applications before the reform. The data contain the rank order lists of programs in each application, as well as information on each applicant's district, high school, GPA, national exams taken and exam scores. The data do not include any information on applications to private programs, a limitation common to all application data in centralized assignment systems that do not include all available options. Despite this limitation, the data allow me to compare application behavior within the centralized system, in particular for applications to public programs before and after the reform.

Pre-reform assignments: I have data on the assignments of students to public programs for the years 2013-2015, which can be linked uniquely to application data through the Matura ID, the unique ID assigned to each student at the time of national exams. Assignment data are available from QSHA for both the initial placement and the final student placement after the reassignment round once those with better off-platform options have rejected their public program assignment.

Post-reform applications: For applications in the post-reform period, I combine data from a number of sources. College application portfolios from the early post-reform years 2016-2017 are publicly available as part of a transparency effort. For the year 2018, I obtain individual application and admissions datasets directly from Albanian universities. While I was not able to obtain application data for all private universities in 2018, I collected applications to all public universities, which is sufficient to analyze strategic behavior in applying to public programs. For the year 2019, I obtained all applications from the Academic Network of Albania (*Rrjeti Akademik Shqiptar, RASH*), the agency formed after the reform to manage the college application process. Unlike application data before the reform, these data do not contain any details on applicant district or high school. I extract applicant district and high school from applicant IDs.[13] I then supplement applicant data with public information on exam scores from the national exams.

[13] I observe a regularity in the structure of IDs in all years of my data. IDs in 2019 have the following structure: 19ABcdeVWXYZ. The first two digits of the ID correspond to the cohort of the student. In the example above, the student is applying to college in 2019. I infer that the second two digits reflect the district where the student went to high school, and the next three digits correspond to the high school from which the student is graduating. Finally, the last 5 digits are unique to the individual. I then map the district code and high school code to the corresponding district and high school using data from 2013-2015, in which I have district and high school information.

Dynamic mechanism offers and enrollments: Initial (phase I) priority rankings from each program are published on the clearinghouse website every year. In addition, the clearinghouse provides initial seat counts available for each program. I obtain enrollment decisions in each of the phases of the multi-offer dynamic mechanism from RASH and combine them with priority rankings of applicants and seat counts for each program to generate offers the students receive in each of the phases of the main round.

I present summary statistics on applications and market characteristics in Table 1.1. The sample contains over 84,000 applicants in the three years before the reform and around 98,000 in the four years after the reform. Public high schools educate the majority of the applicants, about 85% in both periods, and they perform slightly worse on average than private schools in the national exams. The consolidation of private universities into the centralized application system increased the choices available through the same application from the 12 public universities to all 38 higher education institutions and increased the available number of programs through the platform from 289 to 517 over a period of five years. After the reform, students from private high schools have slightly longer application lists than students from public high schools and 21% of their application lists are private programs. In contrast, only 9% of public high school students' post-reform application lists are private programs.

1.3 Application patterns with a partitioned market

1.3.1 Do students of different SES groups have different applications and outcomes in the public-only match?

In this section I show descriptive evidence of differences in applications and assignment outcomes between high and lower-SES students in the centralized system when only public university seats are assigned centrally. This evidence is consistent with high-SES students applying to more selective colleges conditional on exam and high school performance. This behavior is rewarded for some through enrollment in more selective degrees. I formally test differences in application and assignments by regressing measures of selectivity of application lists and assignment outcomes (y_{idt}) on an indicator for SES, and exam and high school performance for the sample of applicants in the period with a public-only match:

$$y_{idt} = \delta_1 \text{lowSES}_{it} + \delta_2 score_{it} + \delta_{dt} + \varepsilon_{idt} \tag{1.1}$$

Score is the average of high school GPA and end-of-high school exam scores. The regression includes district-by-year (δ_{dt}) fixed effects such that the comparison is between high and lower-SES students within the same district and year. Outcomes include the assigned program's rank in the student's rank order list, the assigned program's selectivity and its selectivity rank among all programs.

First, I show that high-SES students apply to more selective programs and conditional on enrollment, also enroll in more selective programs. Table 1.2 shows the coefficient estimates $\hat{\delta}_1$ for the selectivity of the top three listed programs in students' applications. For the selectivity of a program is measured as its score cutoff, the lowest average score for a student who was assigned a seat in the program. Columns 1-3 show that each of the top three choices in the applications of public high school students is less selective than each of the top three choices of private high school applicants by 0.2 grade points. This difference is statistically significant and equivalent to 0.23 standard deviations of the distribution of top choices across the full set of applicants.[14] Columns 4 and 5 in table Table 1.2 show that the differences in selectivity are not limited to programs to which students applied, but also to those where they enrolled. While the DA mechanism narrows the selectivity difference between schools to which students are assigned relative to the schools to which they apply, there remains a difference of 0.054 grade points and 4 ranks in the programs that high and lower-SES students enroll in conditioning on their scores.

[14]This magnitude is also four times the size of the difference in average selectivity between the top and the third listed option for high-SES students, as shown in the third row of Table 1.2.

Second, the rate of assignment to a program on the platform overall, and to top-listed programs in particular differs by SES. Table 1.3 shows the coefficient estimates $\hat{\delta}_1$ for the rate of assignment to the top-one and one of the top-three listed choices as well as the rate of remaining unassigned. For the same average score, lower-SES students are more likely than high-SES students to be assigned to a higher listed choice. They are roughly 20% (6pp) more likely than high-SES students in the same district to be assigned to their first listed choice and 18% (9pp) more likely to be assigned to one of their top three choices. These results show that lower-SES students get assigned more frequently to programs higher in their lists, which implies, consistently with the theoretical prediction, that they apply to relatively less risky choices at the top of their lists. Lower-SES students are also about 12% (2pp) less likely to remain unassigned after conditioning on scores, which indicates that the difference in aggressiveness of applications is not only present among the top choices, but also for the entire portfolio.

Specifications with additional controls confirm that the differences in assignments are robust to comparing applications within high school track and type of high school. In addition, not only do the differences appear once reassignment requests are accommodated, but also for initial offers (Table A.1) and assignment of only students who did not reject their platform offers (Table A.4).

The evidence from both differences in the rate of assignment to top choices and any choices, and the differences in the measures of of selectivity for programs in the rank order lists and those in which students enrolled are consistent with differential risk-taking behavior that may arise when the appeal of outside options is higher for the high-SES group.

Alternative explanations: There are several possible alternative explanations for these observed patterns that may not be related to outside options.

Unequal choice sets: Higher SES students may have different choice sets even within the public match that may explain the patterns. For example, consider two students from different SES backgrounds applying to college from the same district. The lower SES student may not be able to afford relocating to a different city to attend college, so her choice set may be limited to programs available at the local university, which may be less selective than programs further away. It may appear then as though this student is exhibiting less risky application behavior than the high-SES student, though her choices are not reflective of strategic misrepresentation due to preferences for outside option, but are a result of her having a different choice set than the high-SES student even within the set of available public programs. When focusing on applicants with similar choice sets within the options offered on the centralized system, the differences in assignments and applications remain, suggesting that the disparities are not explained by differential access to public programs. Columns 4-6 of Table 1.3 show final assignment outcome comparisons by type of school for applicants from the capital. The restriction to students in the capital allows comparisons of students with very similar choice sets. The differences in applications and assignments are even more stark here than at the national level. Public HS students are assigned to their first choice 6pp more frequently than private HS students, who get assigned to their first choice an average of 17% of the time. The difference is 11pp for likelihood of assignment to the top three listed choices.

Differential grade inflation in high schools: Higher grade inflation in public high schools relative to private high schools may explain application patterns. The formulas for admission to public programs are a function of both high school GPA and national exam scores. If private high schools exhibit lower grade inflation than public schools, it may be that students of the same ability level have different high school GPA components in their score, but they apply according to their privately perceived ability rather than the observed scores. To alleviate this concern, I regress national exam scores on indicators for public status of high school, high school GPA, and an interaction term between high school grades and public status. I find evidence of the opposite: while distributions of within-high school performance differences overlap across private and public high schools, private high schools on average have higher high school grades in math and language than their exams, so that their high school GPA on average overstates their ability to the extent that ability is correctly reflected in their exam grades.

Informational differences: Lower-SES students may have better information about their admission chances to programs or alternatively high-SES students may be overconfident about their chances. This cannot be ruled out at this stage, but the policy analysis in Section 1.4 shows that systematic differences in beliefs cannot explain the full difference in applications in the pre-reform period. In addition, the model I specify in section **??** will allow for students from different backgrounds to have different beliefs about their chances of admission in order to assess the importance of informational gaps in explaining different applications by SES groups.

Different preferences for selectivity: One may be concerned that the observed data patterns are due to high-SES applicants having stronger preferences for more selective programs. While this cannot be ruled out, the model presented in section **??** will allow the separate identification of preferences for selectivity for high and lower-SES students in order to assess the relative importance of preferences for selectivity in students' applications.

As a whole, this descriptive evidence documents disparities along SES lines in assignments through a centralized process to free public colleges. The disparities arise from differences in application behavior that is consistent with more aggressive applications for high-SES students that is detectable both directly, and through assignment patterns. This does not imply that better outside options are responsible for more selective applications by high-SES students. Importantly, differences in beliefs about admission chances or strength of preferences for selectivity could explain this behavior. In the next section, I provide evidence from a policy that incorporated outside options into the centralized application to show additional evidence that the availability of private programs in the choice set, rather than preferences or information, is an important factor contributing to applications in the public program match.

1.4 Effects of the Reform on Applications

In this section, I turn to providing evidence that a shock that differentially impacted the outside and on-platform options of high-SES students relative to lower-SES students closed much of the gap in the selectivity of applications for the two groups. I will analyze the 2016 reform and show that the collapse of the off-platform choice set of private programs and the expansion of the within-platform choices decreases the selectivity gap between high and lower-SES students observed in the pre-reform period. This provides additional evidence that the presence of private programs outside of the centralized system offers high SES students an advantage in public program applications.

1.4.1 Event Study Specification

Difference-in-differences event study: In an ideal experiment, to estimate the effect of outside options on strategic misreporting in applications, I compare two groups that are identical, except one has access to outside options and the other does not, such that when outside options are completely removed, the changes in the affected group's applications relative to the unaffected group reflect the effect of the shock to outside options. Instead, the Albanian policy environment calls for a comparison between the application behavior of high-SES students to lower-SES students before and after the reform.. This is an imperfect comparison. In the Albanian setting, private high schools are a good proxy for high-SES students, whereas public high school students are a worse proxy for lower-SES students as there are many high-SES students among those that attend public high schools. The comparison between these two groups in the data relies on the fact that all private high school (high-SES) students will be treated by the policy (the value for the outside option will shift for all of them), whereas public high school students (lower-SES) have a lower treatment rate. The event study specification is:

$$y_{idt} = \sum_{k=2013}^{2019} \beta_k \mathbb{1}\left[t = k, t \neq 2015\right] + \sum_{k=2013}^{2019} \beta_k^{\text{LowSES}} \mathbb{1}\left[t = k\right] \times \mathbb{1}\left[\text{LowSES}\right] + \beta_s \text{score}_i + +\delta_d + \varepsilon_{it} \tag{1.2}$$

where the omitted group is high SES students in the year 2015 and controls include district fixed effects, and scores. Because the policy change will impact both the outside options of high and lower-SES students, the effect of interest would be the net effect of the reform, the extent to which the reform differentially changed the applications of high SES students compared to lower-SES students. This is captured by the difference $(\beta_k - \beta_k^{\text{LowSES}})$ for each $k \in \{2013, ..., 2019\}$.

Even though this event study specification will allow me to trace out the time path of the estimated effects of the reform, it is susceptible to time-varying confounds. For example, the reform of 2016 also changed the mechanism that allocated students to programs from an algorithmic DA to a live-DA-type assignment system with exploding offers[15]. This may have changed students' beliefs about their probability of admission differentially for high and lower-SES students, which would confound the estimates and would not appear in the pre-trends.

Triple differences event study: To address any time-varying effect of the broader changes to the system that would have systematic effects for all students of the same background but would differ across backgrounds, I use an additional source of cross-sectional variation that would absorb such time-varying effects. This source of variation is merit eligibility for scholarships to private colleges. As discussed in Section 1.2.1, private schools do not generally offer scholarships based on need, but they do offer scholarships based on merit. A common rule of thumb private colleges use to offer scholarships, which is advertised widely, is to give full rides or scholarships for the majority of the tuition to students above a grade threshold.[16] These scholarship policies imply identical, or near identical access to

[15]Details of this mechanism change are discussed in Section 1.2.3 extensively Appendix A.2.2.

[16]For many years, that threshold has been 9 for most schools. This corresponds to roughly the top decile

private university options for top students in both private and public high schools. However, if there are any changes in beliefs about chances of admission that are systematically different across backgrounds after the reform, lower-SES merit-eligible students would be differentially affected by these changes, which would isolate the effect of these changes on lower-SES students relative to high-SES students.

To formalize this empirical strategy, the event study specification is the following:

$$
\begin{aligned}
y_{idt} = & \sum_{k=2013}^{2019} \beta_k \ \mathbb{1}\big(t = k, t \neq 2015\big) + \sum_{k=2013}^{2019} \beta_k^{\text{LowSES}} \ \mathbb{1}\big(t = k\big) \times \mathbb{1}\big(\text{LowSES}\big) \\
& + \sum_{k=2013}^{2019} \beta_k^{\text{NM}} \ \mathbb{1}\big(t = k\big) \times \mathbb{1}\big(\text{NM}\big) \ + \sum_{k=2013}^{2019} \beta_k^{\text{LowSES,NM}} \ \mathbb{1}\big(t = k\big) \times \mathbb{1}\big(\text{NM}\big) \times \mathbb{1}\big(\text{LowSES}\big) \\
& + \gamma_t \text{score}_i + \delta_d + \varepsilon_{idt}
\end{aligned}
$$

where the omitted category is applicants from private high schools in the year just before the reform. In this specification, β_k are the coefficients for merit-elegible high-SES students, β_k^{LowSES} are coefficients for merit-eligible lower-SES students, β_k^{NM} are coefficients for high-SES non-merit eligible students, and $\beta_k^{\text{LowSES, NM}}$ are the coefficients for lower-SES non-merit students. The net effect of the reform is then the difference $(\beta_k - \beta_k^{\text{LowSES}}) - (\beta_k^{\text{NM}} - \beta_k^{\text{LowSES, NM}})$ which reflects the outcome changes for high-SES students relative to lower-SES students after differencing out other time-varying confounds that affect students differently across backgrounds but identically within backgrounds.

In addition, the effect of the reform will be the combined effect of a shock that differentially affects outside options *and* expands choice within the platform.

of applicants, or about 2,000 students. Further details can be found in Appendix A.3.1.

1.4.2 Triple Differences

Finally, the triple difference specification is:

$$
\begin{aligned}
y_{iht} = {} & \beta_1 LowSES_{iht} \times NM_{iht} \times Post_{iht} \\
& + \beta_2 LowSES_{iht} \times NM_{iht} + \beta_3 LowSES_{iht} \times Post_{iht} + \beta_4 NM_{iht} \times Post_{iht} \\
& + \beta_5 LowSES_{iht} + \beta_6 LowSES_{iht} + \beta_7 LowSES_{iht} + \gamma score_{iht} + \delta_d + \varepsilon_{iht}.
\end{aligned}
$$

For student i, attending a high school with status $h \in \{\text{private}, \text{public}\}$ and applying in year t, outcome variables y_{iht} capture application characteristics such as the selectivity of the most selective program or the top two and three most selective public programs in the portfolio. $LowSES_{iht}$ is a dummy for lower-SES type, NM_{iht} is a dummy for non-top student type, and $Post_{iht}$ is a dummy for the post-reform period. The coefficient of interest is β_1 and it captures the effect of the reform on application behavior.

The outcome y_{iht} for both the event study and triple differences is the most selective *public* program choice on the platform. This is chosen as the main measure because the upward gamble induced as more options are chosen (Chade and Smith, 2006), indicates that some of the higher listed and more selective choices will be the last to be chosen, and the first to be removed when the list size restriction becomes more binding.

I measure selectivity of each program as its historical cutoff score, set at year 2013. This decision is made to avoid changes in cutoff scores year-to-year from affecting the estimates when they may be a result of more or less difficult national exams, or years with more or fewer applicants. Importantly, this decision sidesteps changes in equilibrium cutoffs that may be as a result of changes in the capacity of private programs at the time of incorporation into the public match. The historical cutoff score serves as a measure of reputation of a program, which, unlike cutoff scores, takes longer to update.

1.4.3 Results and Discussion

Application sizes and list-filling

The first step to evaluating the effects of the reform on the strategic incentives of students is to establish the extent to which list-filling behavior changed after the reform and the extent of crowd-out of public programs by private programs in application lists. Figure 1.1 provides a visual description of the change and Table 1.4 formally tests the change. Before the reform, the average number of programs to which both high and lower-SES students applied was 8.7 (table 1.4) with approximately 65% of students filling their lists. After the reform, over 80% of students from both high and lower-SES backgrounds filled the application lists, not statistically differently from each other (first column of Table 1.4), with the average number of applications submitted at 9.4 per student. The crowd-out effect of the platform expansion, however, induced more students from high SES background to apply to more private programs than did students from lower-SES backgrounds. The second and third columns of Table 1.4 indicate that after the reform, lower-SES students had, on average, one 1.2 more public programs listed than did higher SES students (or about 13% more of their average application list). This is true even when restricting to applicants who filled their lists (fourth and fifth columns of Table 1.4), indicating that a larger share of the portfolio for those of higher SES that filled their lists was made up of private programs than for lower-SES students. The post-reform changes in portfolio composition is not merely due to students adding private programs to their application lists, but rather replacing some of the public programs they would otherwise include with private programs. This can be seen in the fact that the average count of public programs declines from 8.7 before the reform to 7.6 after for high-SES students.

One potential reason for this could be that the composition of students changed. For example, students could have been induced to apply that prefer private programs to any public ones and would only or mostly apply to private programs through the platform. As shown in Figure A.2, the rate of application through the platform increased for both types of SES backgrounds by only 5pp, and even if all the marginal applicants had applied to only private programs, int could not explain the decline in the average number of public programs in the lists.

Even if changes in composition do not explain the changes in application counts, it is possible that these changes in applications are merely a result of the ability to express preferences over a larger number of programs on the platform. That is, replacing a public program with a private program may be consistent with truth-telling if the least preferred public programs are the ones being replaced with the private programs. In what follows, I will show evidence that it is in fact the most selective programs, and those highest ranked that are most likely to be replaced by private programs in the portfolio.

Application selectivity

The first exercise of this section is to measure the effect of the reform on the selectivity of highly ranked options. I use the most selective option included in the list as a proxy for the most preferred option. This is because in the pre-reform period, where lists are ranked, the top ranked option is the most selective in 94% of the applications. I estimate the event study specification and show the double differences for each period in Figure 1.3.

After the reform, the gap between the selectivities of the most selective choices of higher and lower-SES students closes. Figure A.5 shows the results of the event study triple differences, or the difference between the selectivity gap for top-performing high and lower-SES students and non-top high and lower-SES students. The application gap for top students between SES groups is small before the reform due to targeted scholarships for high performing lower-SES students, which leads both these groups to have similar access to outside options. This gap remains small after the reform. On the other hand, for lower performing students, the gap between high and lower-SES students is high before the reform, but declines after. This exercise suggests that after the reform students' top choices are not as selective as they were before the reform, and the gap in most selective choices between SES group declines, suggesting that the closure in the gap is not due to lower-SES students increasing the selectivity of their top option after the reform, but rather it is due to higher SES students decreasing the selectivity of their top option by more than the lower-SES students do. While the gap and its closure are most pronounced for most selective public choices in students' portfolios, I present event study results for average portfolio selectivity for public programs in Figure A.6. Overall, average selectivity of portfolios for higher SES students declines more than for lower-SES students, consistent with theoretical predictions.

In additional regressions, I check the robustness of the above results to two alternative measures of selectivity. First, I measure the selectivity of a program as the cutoff score of the program and center and scale the distribution of selectivities to have mean 0 and standard deviation 1 in each year[17]. This choice circumvents issues with year-to-year changes in cutoffs that reflect changes in difficulty of end-of-high school exams or changes in the average performance of the pool of candidates. The results of these robustness checks are presented in Figure A.8 and are qualitatively the same as those with the main selectivity measurement. In a second set of alternative results I measure the selectivity of a program as its rank in the given year and estimate the event study specification. Figure A.9 displays results similar to all specifications above.

Finally, in the triple differences analysis, I formally quantify the effect of the reform on program selectivity. The double differences pass tests of parallel trends in several specifications as shown in Table A.5. Results from the triple difference specification for most selective public program are shown in **??**. The removal of outside options decreases the selectivity of public programs at the top of applications by 0.1 standard deviations. On average, portfolio selectivity declines by 0.04 standard deviations. These estimates are robust to alternative specifications and alternative measurements of program selectivity.

Next, I investigate the mechanisms that drive the closure in the application gap. Figure A.4 shows the event study estimates for the four groups of students. Estimates show that overall the reform lead to a reduction in the selectivity of the most selective choice

[17]For programs with empty seats at the end of the admissions process, I set the cutoff score to the lowest average score a student can achieve in the national exams.

listed in the applications of all types of students, but the reduction is largest for high-SES students that are lower performing. This is precisely what is expected to be the effect of the reform: higher SES students in the non-top performing group see their outside options restricted more than lower-SES students and reduce the selectivity of their applications to public schools more in response to this change.

1.5 Model of Application and Decisions on Waitlists

1.5.1 Model Primitives

1.5.1.1 Timing and Sequence of Decisions

1. *National Exams and Decision to Apply*—Students finishing their senior year of high school take end-of-high school exams and learn their scores from the exams. After learning their scores, those who pass the national exams decide whether to apply to colleges through the common application.

2. *Portfolio Selection*—If student i decides to apply, she must select an application portfolio R of ten programs (college-major pairs) from the available set of over 500 such that $R \subset \mathcal{J} = \{1, ..., J\}$, $|R| = 10$. All programs accept applications only through the common application and there are no institutions, public or private that conduct their admissions outside of the common app. Students are not required or encouraged to rank programs in any particular order and the ordering within application will not matter for admissions.

3. *Student Priority Ranking*—Applications are received by each program, and applicants are ranked according to their scores and a pre-determined and pre-announced formula. The first round of acceptances is made in each program for as many students at the top of the program's list as there are seats available. The platform forms waitlists for every program and everyone observes the state of all waitlists in $\mathcal{J}$. Formally, in round

1 of admissions, program j makes offers to the top $q_{1,j}$ students on its waitlist, where $q_{1,j}$ is the number of seats at program j.

4. *Offers and Enrollment Decisions*—At the beginning of each round $t \in \{1, ..., T\}$ of admissions (in the data, $T = 7$), program j makes offers to the top $q_{t,j}$ students on its waitlist, where $q_{t,j}$ is the number of available seats at program j at the beginning of wave t. The state of all waitlists is common knowledge at the beginning and end of each round. Each student with offer set $A_{t,i} \subseteq R_i$ makes a decision to accept a single offer from the set $A_{t,i}$ or reject all round t offers and remain on the waitlists of all programs in R_i to which she have has been offered admission yet, $R_i \backslash \bigcup_{s \leq t} A_{s,i}$.

5. *Waitlist Evolution*—If i receives and accepts an offer from program j in round t, then a j seat is allocated to i and i forgoes all potential future offers to programs she has not yet been admitted to as of round t. Program j has one fewer seats available in round $t+1$ and student i is removed from all waitlists. If i rejects all round t offers, she is removed from all waitlists in $A_{t,i}$, but remains in all waitlists of $j' \in R_i$ from which she have not yet received an offer. Once the final offers are made in round $T = 7$, any remaining seats are allocated on a first-come first-served basis.

1.5.1.2 Student Preferences for Programs

I model the indirect utility that is realized from attending a program as a function of observed and unobserved student characteristics, and observed program characteristics. The utility of student i from attending a program j is given by:

$$v_{ij} = u(z_i, x_j, \omega_{ij}, \varepsilon_{ij}; \theta) \tag{1.3}$$

where z_i is a vector of characteristics for student i, x_j is a vector of characteristics for program j and ω_{ij} is a vector of pair-specific characteristics, and ε_{ij} is an idiosyncratic taste shock for program j unobserved to the econometrician, but observed by the student at the time of deciding whether to apply to college. I assume that the distribution of the idiosyncratic taste shocks is known to the econometrician and each is drawn i.i.d. from a type-1 extreme value distribution. The distributional assumption of the idiosyncratic shock normalizes the scale and location of the utility. In addition, a key restriction imposed by the independence assumption is that the ε_{ij} shocks are independent from student characteristics, in particular distance to programs. This rules out location choices that are correlated with preferences for program. I further parameterize the utility function as follows:

$$\begin{aligned} v_{ij} &= \beta_{c(z_i)} x_j + \gamma^d_{c(z_i)} d_{ij} + \gamma^p_{c(z_i)} \text{price}_{ij} + \sum_k \lambda_{c(z_i)} x_{j,k} z_i + \varepsilon_{ij} \\ v_{i0} &= \varepsilon_{i0} \end{aligned} \tag{1.4}$$

where d_{ij} denotes distance to program j and price_{ij} is the student-specific out-of-pocket price, calculated by subtracting the scholarship a student is eligible for from the list price of program j. Scholarship eligibility and amount is primarily determined by the weighted average score for each student.[18]

Preference parameters are specific to each of four mutually exclusive groups of students in cells $c(z_i) \in \{\text{High-SES, Lower-SES}\}$ and allow for heterogeneity in preferences for observed program characteristics for students from different socio-economic backgrounds and high school subject path. Heterogeneity along the SES dimension for all program characteristics will be crucial in capturing the distributional consequences of alternative market designs, as strategic portfolio choices and enrollments will depend in part on preferences for programs inside and outside the centralized system. In addition, preferences for all program characteristics are allowed to be heterogeneous along high school academic path. In particular, students that chose the social science path at the beginning of tenth grade might care differentially about characteristics of programs such as selectivity and field of study compared

[18]There are very few scholarships offered in each private program for non-merit categories (no more than one or two per program) and I cannot identify the students eligible for these scholarships in the data. The number of students whose choices would be affected by scholarship eligibility through the above categories is very small and its impact on parameter estimates negligible.

to students who chose the science path.

Student characteristics z_i include average score on the end-of-high-school exams and urban/rural location. Program characteristics x_j include selectivity, private/public status, field of study in one of four categories (science and applied science, health, social science and humanities, business and economics), an indicator variable for whether j is located in the capital.

Finally, the value from the outside option is given by v_{i0} and represents the value of not enrolling in any college in the current year and is known to students at the time of application. This may include the value from entering the labor force without a college degree or waiting to apply the following year. This is without loss of generality because the choice of portfolio and decision to enroll will depend on differences with the non-college option and not on the value of the non-college option itself. Alternative admissions designs are assumed not to affect the value of the non-college option. This may not hold if changes in the way this market operates affect the expected utility from applying the following year, which is included in the non-college option.

This formulation does not allow for systematically different value from the outside option for high-SES and lower-SES students. Because preferences for programs are defined separately for each SES group and for application and enrollment decisions, only the differences between the value of programs and the value of outside options matter and if high-SES students have better non-college outside options, this will be captured by less strong preferences for any college alternative. Since the value of the outside option does not change in alternative market designs, this specification choice will not affect the distributional consequences of counterfactual policies.

1.5.1.3 Information and Beliefs Over Program Cutoffs

Before applying, students learn their score and form beliefs about their probability of admission to each program. For each student entering the application stage, each possible application portfolio is a lottery over entrance to one of the programs in the portfolio and not enrolling in college. In their portfolio selection, students take into account not only their preferences for each available program, but also the probabilities over possible outcomes induced by their choices. Because the admissions process involves multiple waves, the initial portfolio choices contain information not only about student preferences, but also about beliefs over the possible outcome paths generated by the portfolio and choices to enroll or wait in each of the admissions waves. A student may form beliefs about the state of waitlists in the each of the waves and distribution of outcomes in the next wave for each possible choice made in each possible induced state of the waitlists. The interdependence of waitlists down the admissions process generates an extremely large number of possible states and a different distribution of perceived possible outcomes for following rounds for each possible state, which makes estimation of all of the belief objects infeasible.

To circumvent this issue, I rely on the specification of the perceived payoff function, which only takes in the probabilities of clearing the final cutoff of each program. That is, the only relevant object in determining the perceived uncertainty of each outcome is $P(c_{Tj} < \text{score}_i)$ for all programs j, where c_{Tj} is the cutoff of program j at final wave T. To support this assumption, it helps to recall the institutional details of the information that students receive from the central admissions authority. At the time of application, students have access only to information about the previous year's *final* score cutoffs for each program (c_{Tj}). This feature is important as students are lacking any public information about the program cutoffs in intermediate waves.[19]

In addition to providing support for the objects over which students form beliefs, the information provided by the admissions authority supports a model in which students form beliefs over the distribution of cutoffs for programs that arise form the previous year's cutoffs.[20] Therefore, I model below the perceived distribution of possible realizations of the

[19]In fact, according to the admissions agency, this decision is intentional so as to simplify the information given to applicants and to focus their attention on final cutoffs rather than intermediate ones.

[20]This approach to modeling perceived admission probabilities as arising from beliefs over the distribution of cutoffs is widely used in the school choice literature. See for example Agarwal and Somaini (2020).

cutoff c_j, $\tilde{c}_{c(i),j}$ as normally distributed and parameterized as follows:

$$\tilde{c}_j \sim N(\bar{c}_{j,c(i)}, \sigma_j^2)$$

where

$$\bar{c}_{j,c(i)} = \text{cutoff}_j^{y-1} + \mu_{0,c(i)} + \mu_{1,c(i)}\text{cutoff}_j^{y-1}$$

and

$$\sigma_j = \log(1 + \exp(\sigma_0 + \sigma_1 \text{cutoff}_j^{y-1})).$$

The specification above assumes that the belief object is centered around the shifted previous year's cutoff with cutoff_j^{y-1} denoting the previous year's $(y-1)$ cutoff, $\mu_{0,c(i)}$ denoting the shift intercept and $\mu_{1,c(i)}$ denoting the shift slope on program cutoff. This reflects the fact that the mean of the perceived possible distribution of cutoffs may depart systematically as students may believe that admissions in the year they are applying are more or less competitive in general than admissions in the year before and perhaps differentially so for more competitive programs. In addition, I model the standard deviation of the distribution of beliefs around the cutoff as a function of cutoff. This allows for the dispersion perceived distribution of cutoffs to be heterogeneous among students with different academic performance facing a set of choices that are more or less selective. Finally, all belief parameters to be estimated that are indexed by $c(i)$ are allowed to vary by SES. In practice, this parameterization with heterogeneous beliefs by SES, in addition to the preference specification, allow for separate estimation of both taste and belief parameters for the two groups of students.

With this model of perceived distributions of possible program cutoffs, we can derive the perceived probability of student i to clear the final cutoff of program j as:

$$\begin{aligned} P(c_{Tj} < s_i) &= \Phi\Big(\frac{1}{\sigma_j}(\text{score}_i - \tilde{c}_j)\Big) \\ &= \Phi\Big(\frac{1}{\log(1 + \exp(\sigma_0 + \sigma_1 \text{cutoff}_j^{y-1}))}(\text{score}_i - \text{cutoff}_j^{y-1} - \mu_{0,c(i)} - \mu_{1,c(i)}\text{cutoff}_j^{y-1})\Big) \end{aligned}$$

The probability of admission at program j will be 50% for students with scores at exactly the value $(\bar{c}_{j,c(i)})$. For anyone with scores to the right of that point, the perceived probability of being admitted to program j will be greater than 50%, but will be declining with increasing variance in perceived possible cutoff distribution. The opposite is true of students on the left side of $(\bar{c}_{j,c(i)})$. Their perceived probability of admission to program j is less than 50% but is increasing with increasing variance σ_j^2. To see the implications of allowing heterogeneity in variance for students with different performance, consider a higher scoring student (A) and a lower scoring student (B) who are looking at schools with past cutoffs equally far from each of them $(\text{score}_A - \text{cutoff}_j = \text{score}_B - \text{cutoff}_{j'})$. If they both fall on the $< 50\%$ probability of admission, A will think admission to j is less likely than B thinks is his admission probability to j' if A has less uncertainty than B. On the other side, if they are quite likely to be admitted, student A will be more certain of that than student B.

1.5.2 Choice Problem

1.5.2.1 Portfolio Choice

Applicants in the centralized system take in the vector of utilities v_i and cutoff clearance probabilities p_i and submit application portfolios of size $|R| = 10$ to maximize the following objective function:

$$\begin{aligned} V(R') &= v_{R'(1)} \cdot p_{R'(1)} + (1 - p_{R'(1)}) \cdot v_{R'(2)} \cdot p_{R'(2)} + \ldots + \prod_{l=1}^{|R|-1} (1 - p_{R'(l)}) \cdot v_{R'(|R'|)} \cdot p_{R'(|R'|)} \\ &= \sum_{l=1}^{|R'|} v_{R'(l)} \cdot p_{R'(l)} \prod_{k=1}^{|R'|-1} (1 - p_{R'(k)}). \end{aligned}$$

where $v_{R'(l)}$ denotes the utility from the lth most preferred program in the portfolio and $p_{R'(l)} = P(c_{R'(l)j} < \text{score}_i)$ denotes the probability of clearing the final cutoff of the lth most preferred program in the portfolio. The restriction on students filling their lists completely is a reasonable approximation in my empirical context as it is founded on the fact that the vast majority of students fill their lists ($> 80\%$) in the estimation period and less than 10% of students submit fewer than 7 programs in their application. In addition, the restriction of list-filling implies that some students may apply to programs they will not choose over the outside option which will help explain waitlist decisions to exit the system altogether than enroll in some available option. The probability of individual i choosing portfolio R_i is then:

$$\ell_i(R_i|\theta, z_i, \omega_i) = \mathbb{P}(V_i(R) > V_i(R')\forall R' \neq R).$$

1.5.2.2 Probability of Observed Choice in the First Round

At the first wave of admissions, students learn the set of programs A_{1i} they have been initially admitted to and their waitlist position in the programs to which they have not been admitted yet. They decide whether to enroll in one of the options to which they are admitted (a_j, s.t. $j \in A_{1i} \subseteq R_i$), forgo all such options and wait for a more preferred program in their list (w) or exit the admissions system altogether without enrolling in a program (e). In the offer acceptance and rejection stage of the first wave W_1, applicants take one of the possible actions $d_{1i} = \{a_j, w, e\}$ such that

$$d_{1i} = \underset{a_j, w, e}{\operatorname{argmax}}\ V(W_1) = \begin{cases} v_{ij} & \text{if } d_{1i} = a_j \text{ s.t. } j \in A_{1i} \\ v_{i0} & \text{if } d_{1i} = e \\ V(R_i \backslash A_{1i}) & \text{if } d_{1i} = w \end{cases}$$

where

$$V(R_i \backslash A_{1i}) = \sum_{l=1}^{|R_i \backslash A_{1i}|} v_{R_i \backslash A_{1i}(l)} \cdot p_{R_i \backslash A_{1i}(l)} \prod_{k=1}^{|R_i \backslash A_{1i}|-1} (1 - p_{R_i \backslash A_{1i}(k)})$$

is the value of forgoing a current offer and waiting for future rounds. The expression for the value of future rounds comes directly from the perceived probabilities of clearing the admissions cutoffs for the programs that students had at the time of application. This expression makes the assumption that applicants' beliefs about the final cutoffs of programs they are waitlisted for do not change once they observe the cutoff for the first round. Two facts support this assumption: first, there are six additional waves of admission after the first and cutoffs in practice move dramatically between the first and the seventh wave for most programs.

Then the probability of observing d_{1i} conditional on the portfolio choice R_i and round 1 admission set A_{1i} is:

$$\ell_i(d_{1i}|A_{1i}, R_i, \theta, z_i, \omega_i) = P(d_{1i} = \text{argmax } V(W_1))$$

1.5.3 Identification argument

The identification challenge this paper faces is that of separating preferences for programs from beliefs about probabilities of admission to various programs. My strategy utilizes the three stages of decision-making modeled above: (1) decision to apply, (2) portfolio choice, and (3) waitlist decision to enroll or wait. On its own, none of the three stages allow separate identification of beliefs and preferences. In particular, portfolio choices arise from a combination of both beliefs and preferences and can be rationalized with many such combinations.

Separating preferences and beliefs: Identification of preferences and beliefs is done jointly using both portfolio choices and waitlist decisions. Waitlist decisions, in particular, help identification in two key ways. First, enrollment choices among students with multiple offers provide crucial information on preferences over programs by a simple revealed preference argument. It is important to clarify, however, that this is not sufficient for understanding whether and by how much the program that the student enrolled in is more preferred among all other programs given that each matriculation choice is made among a small and selected set of programs.[21] Incorporating the portfolio selection stage helps dis-

[21] The information obtained from matriculation choices is similar in nature to that obtained from observing the specific rankings of programs in a rank-order list under common DA implementations. Under the common assumption that students list the included programs in order of preference, an observed list provides

cipline the nature of selection into the waitlist stage choice sets. Second, it is helpful to observe waitlist decisions when students face both programs they've been admitted to and programs for which they remain on the waitlist. To illustrate why such decisions aid the separate identification of beliefs and preferences, consider a case in which (1) some students face a choice between program j, from which they have an offer among others, and waiting for program j', for which they are on the waitlist, and (2) other students face a choice between enrolling in either j or j', having been admitted to both. Differences in the extent to which students facing uncertainty for program j' are more likely to accept j (conditional on the application portfolio and observables) is determined by the probability of admission to j', which helps pin down $p_{j'}$.

In addition to waitlist choices, portfolio choices are used to identify substitution patterns across programs and heterogeneity in preferences. With relevance to heterogeneity by SES, observing more private university choices in the portfolios of higher-SES students than is predicted by lower-SES students' tastes implies a stronger taste for private education. More generally, the covariance between student and program characteristics identifies heterogeneous tastes for observable program characteristics. While the specification of my model does not include unobserved heterogeneity in preferences for program characteristics, it is quite possible to include such heterogeneity, which would be identified from seeing portfolios from observably similar students that have a concentration of programs with certain characteristics, or a starker absence of programs of other characteristics to an extent greater than can be predicted by a model without unobserved preference heterogeneity. Finally, it is worth noting the identification of the price coefficient. As discussed in Section 1.4, the policies on scholarships to private programs available to students with scores above a certain threshold provide exogenous variation in price for students on either side of the threshold,[22]

information about the relative preference ordering of programs on the list, but not those programs among all in the choice set. This is the crux of the demand estimation challenge in school choice (see Agarwal and Somaini (2020) for a review of the evolution of approaches the school choice literature has taken to estimate demand).

[22]In order to take advantage of this variation, I compute the price that would need to be paid by each student that clears the scholarship cutoffs for various scholarship amounts for each school. Using this

which helps identify the price coefficient in equation (1.4).

Separating slope terms from intercepts in belief parameters: The main source of identification of slope parameters will come from choices across students of different performance. The variation in the distribution of selectivity of programs that are selected will be informative as the only way perceived admission probabilities will change across applicants facing choices in different ranges of school selectivities is through the slopes.[23] The extent to which higher performing students choose portfolios that are more selective relative to their own performance than lower performing students will be telling of the steepness of slopes for both mean shift and uncertainty parameters. To this end, choices on the waitlist across students of different performance will similarly aid identification. For example, the choice of higher performing students to wait for more selective schools relative to own performance than lower performing students would, pins down the differences in perceptions for higher cutoff schools.[24]

Separating shift parameters from variance parameters: The final challenge in identification is separating the mean shift in perceived distribution of potential cutoffs from the variance of this distribution. To help illustrate the identification point, Figure 1.4 shows how features of the selectivity of the chosen portfolio would vary with varying μ_0 and σ_0 holding slope parameters of the belief distribution constant. The key insight of the figure is

constructed price as the price the students face for each private program assumes that they know they would be able to obtain these scholarships. This may not be an innocuous assumption if applicants are distrusting of the offer viability or exact amount.

[23] My estimation strategy does not involve computing the optimal portfolio. Instead this insight would be reflected as a higher likelihood of observing a deviation from the chosen portfolio toward a less selective reach program. See more details on estimation in Section 1.5.4.

[24] Variation in portfolio entries within student is also useful here. In particular, the extent of expansion on either side of the cutoff distribution relative to own score helps pin down how much admissions probabilities change along the cutoff line relative to own score. The addition of more selective reach choices than would be predicted optimal under a given set of beliefs indicates that students have either a more optimistic mean shift or a greater belief variance (assuming their scores fall below the shifted mean of the reach program included) about such choices. On the other end, if students are adding less selective safeties than would be predicted by a set of beliefs, then beliefs around less selective programs have a less optimistic mean or a higher variance (assuming student score falls above shifted mean of such programs). While it's not clear in principle whether the likelihood contributions of choices at the extremes of the portfolio are significant, the general argument holds for less extreme portfolio choices with cutoffs closer around the student's own performance: the distribution of selectivities of portfolio choices relative to own performance is a source of variation that helps identify the extent of decline in probability of admission along the cutoff line.

that the most selective program, average selectivity, and least selective programs respond at different rates to changing σ_0 for the same mean shift and at different rates to changing μ_0 for the same variance. Even though there are two variance parameters for each mean shift that would yield the same *reach* program selectivity (panel (a)), only one variance guess of the two would be closer to yielding the data-observed selectivity of the *safety* school (panel (b)). In addition, across mean shifts some can be ruled out as fitting the data and among those that cannot, the likelihood-maximizing shift-variance pair will be one best-fitting all choices in the portfolio according to their likelihood contributions.

1.5.4 Estimation

Bringing together the likelihood of observing student i making the decision of whether to apply, the likelihood of observing student i choosing the observed portfolio and the likelihood of student i making the observed wave 1 decision yields the following expression of the likelihood of observing the sequence of choices in the data for each student i:

$$\ell_i(\theta) = \int \ell_i(\text{apply}_i|\theta, z_i, \omega_i)\ell_i(R_i|\theta, z_i, \omega_i, \text{apply}_i)\ell_i(d_1|A_1, R_i, \theta, z_i, \omega_i, \text{apply}_i)dG_i(\epsilon)$$

Preferences and beliefs are jointly estimated via Simulated Maximum Likelihood (SMLE). The expressions for the likelihood function is not closed-form and an estimation method of this kind would call for drawing utilities many more times than there are portfolio alternatives given a guess of the parameter vector, computing the optimal portfolio and the probability that this portfolio is chosen among all possible candidates, and then computing the conditional probability of choice in the second stage. An obvious issue arises here. Given the number of possible programs to choose from, there are 10^{20} possible portfolio choices making computation of probabilities infeasible. I address it using a simplification introduced in Larroucau and Rios (2020). For rank order lists, the insight implies that it is sufficient for optimality that a rank order list is preferred to all portfolios created by a single-shot replacement of each ranked program in the list with a program not ranked. I adapt this insight to unordered portfolios and show that it holds for unordered portfolios too. Formally, using notation from Larroucau and Rios (2020):

Let $C = \{j_1, ..., j_k\}$ be an unordered application list of length at most K, i.e. k $\leq K$. Without loss of generality, let $u_{j_1} \geq u_{j_2} \geq ... \geq u_{j_k}$ so that the utility from submitting application portfolio C is

$$V(C) = p_{j_1} u_{j_1} + (1 - p_{j_1}) p_{j_2} u_{j_2} + ... + \left(\prod_{l=1}^{l=k-1} (1 - p_{j_l}) \right) p_{j_k} u_{j_1}$$

If $\mathcal{S}(C)$ is the set of one-shot swaps of portfolio C and

$$V(C) \geq V(C') \ \forall \ C' \in \mathcal{S}(C),$$

then

$$V(C) \geq V(C') \ \forall \ \ C' \ \ s.t. \ \ |C'| = K.$$

Proof. See Appendix A.4.1 $\square$

Estimation sample: Computation of the likelihood function remains challenging even after drastically reducing the choice set. Given that there are over 500 programs in the market, the likelihood of a portfolio being optimal even among its one-shot deviations must be calculated among more than 5000 alternatives.[25] Some simplifications are in order. I restrict the set of programs in each person's choice set to those in the region that the student went to high school and the capital. For example, I restrict the choice set of students from the north to the programs offered by the regional university in the north and the capital. This consists of 5 public universities (1 regional and 4 in the capital) and all of the private ones. For students in the south of the country, this implies 6 public universities (2 regional

[25] The count of alternatives in the "choice set" is $|R'| \times (|\mathcal{J}| - |R'|)$, which in context would imply $10 \times (517 - 10)$ alternatives.

and 4 in the capital) and for students in the bigger central districts, this implies 6 public universities.[26]

1.6 Results from the model

Predicted beliefs using model estimates are shown in Figure 1.5 and the remainder of the parameter estimates can be found in Table A.7. I find a substantially dispersed distribution of possible cutoffs. In fact, the distribution of cutoffs falls outside the range of possible grades, $[4, 10]$ for programs with cutoffs at the extremes of the range. Naturally, no student would believe that the cutoff of a program could be above the range of possible scores with any non-zero probability. I interpret the existence of probability mass outside of the extremes of possible scores as simply and indication that the highest scoring students do not believe they have a probability 1 of being admitted to all programs, and similarly, that the lowest scoring students do not believe they have 0 probability of getting into any programs. In addition, I find that lower-SES students are more optimistic about programs with low cutoffs and more pessimistic about programs with high cutoffs. The variance of the distributions is held constant across groups as estimates were performing poorly for the high-SES group when estimated separately.

[26]This reflects the fact that in the Albanian context, students from the north will almost never apply to regional universities in the south of the country and vice-versa. In addition, the choice sets within the relevant universities are restricted to reflect the differential choice sets relevant for students of different high school tracks. Those of the science track will never apply to certain humanities degrees in any of the years in the data, both before and after the reform. In addition, students from a social science high school track will never be observed applying to certain science degrees.

1.7 Counterfactual results

1.7.1 Evaluating trade-offs of partially and fully centralized admissions

Using the estimates for the structural model, I simulate choice and allocations in counterfactual designs and evaluate the role of private outside options on the choices, assignment outcomes and welfare of students graduating high school in Albania. First, I describe the setup for the counterfactual simulations and the measure of welfare I use and then present results from the simulations.

1.7.1.1 Counterfactual setup

I conduct counterfactual simulations using the sample of students who graduate high school in 2019. The preference parameters, including the distribution of the random taste parameters are held fixed in the simulations. Similarly, the distribution of beliefs is also held invariant to policy changes. The beliefs in estimation are captured in reduced form and aim to isolate the level of uncertainty and bias carried year-to-year in this market. I compare the performance of two market structures. The baseline structure is one in which students are allocated to college seats in two procedures simultaneously: the centralized assignment to public programs with restricted lists and a centralized assignment to private programs where there is no constraint on the number of applications one can submit. The alternative market structure is one in which students are allocated to college seats in a single procedure where their list sizes are constrained and they apply to both public programs and private programs in the same platform. I describe the two market structures in more detail below.

Centralized Public Match with Private Outside Options (baseline configuration): The baseline market is one that assigns students to seats according to a Deferred Acceptance mechanism in two parallel assignments. The mechanism proceeds as follows:

1. Students apply to ten programs in the public match and to all available private programs in a simultaneous private match.

2. Applicants to the public match are ranked according to the pre-determined formulas of each program and an initial placement to public programs is made through a DA algorithm.

3. Private programs simultaneously rank all students according to their formulas and determine student priorities

4. Once students have their initial placement in a public program, the private options begin proposing to students in order of score priority and the outside option proposes to everyone. Once private programs and outside options begin proposing, the allocation evolves as follows

    - If the initial public placement is a student's first choice among the set of all private programs, the outside option, and the public programs to which the student has applied then this initial placement is the student's final assignment.

    - If one of the private programs proposing is a student's first choice in the set above, then the student enrolls in that program, foregoing her placement in the public match. Similarly, if the outside option is the best option in the set then the student exits the mechanism foregoing all inside options. If the private proposal is better than the public proposal, the student temporarily holds the private offer and forgoes the public offer.

    - The public match fills vacancies created in the first stage of private program proposal by reassigning all students except those with final assignments in the first stage.

    - Private programs similarly reassign all temporarily assigned or unassigned students

- The second stage of outside option proposal proceeds the same way as the first. Private programs and outside options propose anew and students accept only if the program proposing is their favorite program. Otherwise they hold their best offer temporarily.
- The rounds of assignments proceed until either each student is enrolled in their best option that they applied to or private option conditional on clearing the program's cutoff, or the student has exited the mechanism, or has remained unassigned and no program that they applied to or any private program they prefer to the outside option would admit them.

A few notes are worth making about the above mechanism. First, the assignment procedure above is frictionless. That is, I assume a centralized market for outside options that is coordinated with the public match. This excludes the possibility that a student may receive an offer that they accept from a private program and there is a vacancy they leave behind in the public match that never gets filled. It also excludes the possibility of congestion in the private market and abstracts away from a situation in which offers may be made in the private market to students who apply first and may not be as qualified as students who may apply later do not get accepted due to the timing of their application to a private program. These types of aftermarket frictions are outside of the scope of this paper. The goal of the counterfactual exercises presented is to understand the effect of *strategic applications* that result from the quality of outside options and list size constraints.[27] In fact, the mechanism described above is equivalent to a one-stage deferred acceptance mechanism in which students submit a single application list to the platform in which they are allowed to include up to 10 public programs and all private programs.

A parallel mechanism for public programs with a maximum list size of 10 and unrestricted

[27] A different study, Kapor et al. (2022) studies the role of aftermarket frictions on the allocation of students in partially centralized admissions. They find that outside options generate frictions in the centralized match through creating vacancies that need to be filled by creating other vacancies through extracting students from their assigned seats. My paper abstracts away from these aftermarket frictions and assumes that the chains of reassignment happen through a whole-market reassignment with no frictions.

list size for private options is equivalent to a DA algorithm with 10 slots reserved for public programs and an unrestricted number reserved for private programs.

The only difference of this mechanism with an unrestricted DA is what arises at the application stage due to list size restrictions, incorrect beliefs, and preferences for private options.

The final note that needs to be made about the counterfactual simulations is the assumption that beliefs are invariant to policy changes. In this counterfactual, I assume that both beliefs about the cutoffs of public programs and private programs are the same as those estimated. This has implications for applications not only directly through how students weigh the probability of admission to programs in the restricted-list application, but also through the value from the private options. Results are qualitatively the same in two alternative counterfactuals: (1) one in which the students' value from the private outside options is the utility from their favorite private option assuming that they do not consider the probabilities of admission to private programs, and (2) a counterfactual in which students carry rational expectations beliefs about the outside options. Next, I describe the second main counterfactual, the "all-in" configuration:

The All-In Match (no-private outside options configuration): This no-private outside options configuration allows students to apply to college only through the single restricted-list match. The match assigns students to seats via a Deferred Acceptance mechanism. The only non-platform option is the no-college outside options. This procedure is more straightforward than the one with outside options as all options are on-platform and no assumptions have to be made about the allocations to programs outside the platform. The assignment proceeds as follows:

1. Students apply to ten programs only among both public and private programs in a single application on the "all-in" platform.

2. Applicants are ranked according to the pre-determined formulas within each program

and an initial placement is made through a DA algorithm.

3. Once students have their initial placement in a public program, the outside option proposes and students who prefer the outside option exit the mechanism altogether

With the details of the two counterfactual assignment mechanisms established, it is worth emphasizing that the only difference in assignment generated by differences in applications, rather than any differences in the assignment process. This is unlike assignment processes with centralized public and decentralized private markets, but I make this choice in order to isolate the changes in assignments only through changes in application. In the end, final assignments may change in the case with outside options if congestion effects cause unstable matches.

1.7.2 Effects of incorporating all programs into the same platform

In this section, I report the descriptives of application patterns under the two alternative market structures, the final assignments with the assignment algorithms described above and the final welfare results.

1.7.2.1 Changes in the allocation of students

The assignment results show that changing the market structure to an all-in system reduces matching efficiency relative to the counterfactual with a partitioned market, but more so for high-SES students, thus improving equity. Overall, the assignment results show that there is a decrease of 4.2pp in enrollment to college for lower-SES students and 1.7pp for high-SES students. In addition a net of 2.9% more high-SES students had worse assignments than had better ones whereas 1.1% more lower-SES students had better assignments than had worse assignments. The assignment results show that more people lose than gain in both groups, and this happens through two channels: (1) students whose applications became less selective and who excluded programs they would have been admitted to and that they preferred relative to the one they were assigned to; and (2) students who were displaced because more people are applying to less selective colleges. The reduction in enrollment is more prevalent for lower-SES students while the worsening of matches in an all-in system happens on net for the group of high-SES students alone, who have more inside options they are willing to accept inside the platform, but who are more constrained in their applications.

Figure 1.6 shows more detail on the above results and carefully documents differences in matches when going from a partitioned system to an all-in system. When the system changes, there are both winners and losers that result directly from changes in applications as well as indirectly from changes in the applications of others. Among lower-SES students, 4.6% are induced to not enroll in college and among high-SES students, 1.7% are induced to end up unenrolled. Even though applications changed for this group, the spillover effects from others dominate the conservativeness of their applications and they end up not being enrolled anywhere, whereas they would have attended college in a partitioned system where others would have applied to and gone to different, potentially more selective colleges allowing them to not get pushed out of going to college. The spillover effects serve not only to induce people to end up not enrolling in college, but also to induce some people who would have otherwise not gone to college to enroll. These benefits accrue more to lower-SES students who are more likely to keep applying the same way and are only exposed to the spillover effects from others: 0.4% of lower-SES students go to college in an all-in configuration that otherwise would have not. The final net effects of an all-in policy are to reduce college enrollment through constraints that induce students to change their applications as well as spillovers onto others of such changes.

By similar arguments, even the set of people who would enroll in college in both configurations observe both winners and losers. Of particular note in Figure 1.6 is the set of students who go from enrolling in a public program to enrolling in a private program. All the students in this group experience a worse assignment from the policy change. In the partitioned market they would have rejected their best possible private option for the offered public option, but in the all-in case they either never applied to this public option, or were pushed out of it. 1.6% of high-SES students and 1.5% of lower-SES students experience such a change, which is due to constrained applications–high-SES students must substitute away from public programs in their applications and toward private programs more frequently than lower-SES students and as such forgo a public assignment more frequently.

Finally, I discuss the outcomes of those that enroll in the public system in both structures. Spillover effects and application effects net out for lower-SES students, but a net of 0.7% high-SES students lose. The losses here come from reducing the size of the public-only subset of one's application and changing its composition. The constraints on this subset of the application are more binding for high-SES students who also prefer private programs more and will have to incorporate them to a larger extent in their applications.

In summary, strategic applications cause a net loss when incorporating outside option colleges in a central system. A higher share of high-SES students lose because their application behavior is the most affected. Lower-SES lose mostly through lower enrollment, but gain conditional on matching due to spillovers from others. Three forces are at play. First, applications to public programs become less selective due to both the worsening of outside options and a more binding list-size constraint. This results in some assignments that are worse because students have removed from their applications programs they would have preferred and been admitted to. Second, the new restriction on private applications induces misrepresentation of preferences among private options. Third, students are crowding their applications more toward lower selectivity programs causing spillover effects that push out of admission students that would otherwise have been admitted to certain programs. In all, these forces end up causing more losers than winners, on net 3.1% for the lower-SES group and 4.6% for the high-SES group.

1.7.2.2 Changes in welfare

I compute the average student welfare as the average ex-post utility from assignment under each regime. Formally, welfare is computed as:

$$W(M) = \frac{1}{N}\sum_{i=1}^{N} E\left[v_{i,f(i)}\right]$$

where M is the market structure, $M \in \{\text{partitioned}, \text{all-in}\}$ and $f(i)$ is student i's final assignment $f(i) \in \varnothing \cup \mathcal{J}$.

Figure 1.7 presents welfare differences relative to welfare realized under an unrestricted DA algorithm. In the partitioned market, high-SES students achieve more of the gains possible than lower-SES students, but this difference reverses when moving to an all-in system. This reversal in relative gains, comes at an efficiency cost: the gap in realized gains relative to the unrestricted DA for both groups increases. For high-SES students, in particular, the welfare gap increases by €160, while for lower-SES students it increases by €98. The findings from the welfare calculations imply that a policy that aims to improve equity through disallowing a market outside of the centralized match to exist comes at an efficiency cost due to the application and information constraints.

A note of caution should be added here about interpreting these results. They do not reflect equilibrium behavior because in equilibrium students may update their beliefs about admission probabilities and reoptimize in a way that may change the direction of these qualitative results. Nevertheless, given the heterogeneity of beliefs by SES group and substantial uncertainty about cutoffs, it is hard to imagine that expectations are rational. Second, the counterfactual exercise abstracts away from improvements in efficiency from the reductions in matching frictions that come from centralization. These improvements are well-documented in the literature (Abdulkadiroğlu et al., 2017; Kapor et al., 2022). The lesson is to highlight a channel of behavior that market designers must take into account when choosing their school choice or college admissions system. In the Albanian context, despite substantial uncertainty about cutoffs, outside options have only a small role in choice.

Welfare under alternative list sizes: I use the estimated model to evaluate alternative market structures that may improve student equity at a lower efficiency cost for the market. I evaluate alternative list sizes the size constraint with an all-in configuration. In principle, strategic effects coming from list size restrictions can be completely neutralized with an unrestricted list, but policymakers may not want to give unlimited choice to students.[28] Figure 1.8 shows the welfare effects of adding choices to the application list incrementally. Allowing just four more choices in the all-in system closes the welfare gap with the unrestricted DA system for lower-SES students by more than half relative to a partitioned system. The gains for high-SES students need more options to converge to a partitioned system because they value unlimited choice of private options more than lower-SES students do.

1.8 Conclusion

In this paper I evaluate the role of market structure in strategic choice and allocation in partially centralized college admissions with list size restrictions and private schools as outside options. The insight is that for applications within the match, it may matter strategically what outside options a student has. Students with better outside options can apply more ambitiously within the centralized platform and ultimately be assigned to more preferred programs than students with worse outside options. This strategic response may have significant efficiency and equity consequences, given that private outside options are expensive and offer higher value to students with higher socioeconomic status (SES). To evaluate the effects of outside options in strategic applications, I use novel data from Albania and a policy that incorporated all private colleges into the public centralized assignment while maintain-

[28]There could be many reasons for this, which are not modeled in this paper. Allowing an unlimited number of choices may impose a cognitive burden on students searching for programs and may advantage those who have more resources and help and are able to obtain more information and better able to rank many choices. These forces are not modeled in the paper and I assume that allowing a few more marginal choices does not impose additional cost on students searching.

ing the same list size restriction. I combine a reduced-form analysis and structural model to evaluate strategic behavior, strategic advantages of students with better outside options, and their effect in welfare and equity.

I first provide descriptive evidence of differences in the applications of high and lower-SES students and the quality of public institution that they attend in the centralized system with restricted application sizes when private programs are excluded from the platform. Higher SES students apply to and enroll in more selective free public institutions than their lower-SES peers with the same high school performance. This striking fact motivates the rest of my analysis.

I analyze a policy change implemented in the centralized match in Albania that enforced participation by all colleges, public and private, in the centralized platform. I use an event study design to measure the effects of centralizing all available alternatives on application behavior and match outcomes. For this event study analysis, I compare applications of high-SES students to lower-SES students before and after the reform. In addition, to reduce concerns over confounding effects of other market-wide changes that may affect different SES groups differently, I take advantage of merit-based scholarships that ensure that top students of all backgrounds have equal access to private college seats to argue that top students from high- and lower-SES backgrounds apply to similarly competitive programs before and after the match expansion and they are not differentially affected by other aspects of the reform. For all lower-performing students, the desirability of private decentralized options depends on SES status, so the theoretical framework predicts that the removal of private colleges as outside alternatives to the match has differential effects on application behavior of high-SES and lower-SES students.

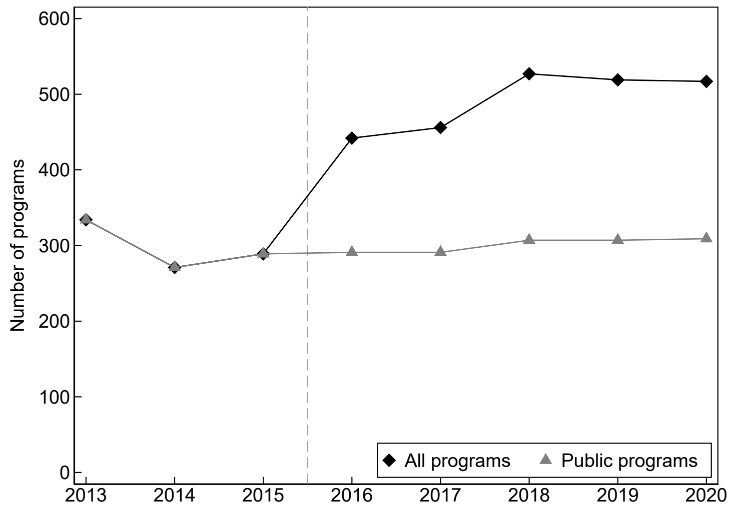
The event study captures the differential changes in application behavior of high- and lower-SES students that come from changes in the desirability of outside alternatives. This exercise offers several suggestive findings. First, outside alternatives affect application behavior and matches within the centralized portion of the market. Second, when outside alternatives are private and costly, their desirability will break along SES lines, yielding different strategic application behavior for high- and lower-SES students. Outside alternatives are more desirable for high-SES students, giving them not only higher direct value from choosing these options, but also a strategic advantage over lower-SES students within the match to public colleges and majors. Third, participation in the match by all colleges reduces differences in application behavior, primarily driven by a reduction in aggressiveness of applications of high-SES students.

In the second part of the paper, I quantify the welfare and distributional impacts of the existence of private outside options and list size restrictions. I build a structural model of student applications and matriculation choices and estimate it using data and institutional features from both before the policy change and after it. My model captures student choices that balance preferences for college-major pairs and beliefs about probability of admission to each option. I advance the literature by relaxing assumptions of truth-telling or rational expectations, and instead rely on institutional features to separately identify expectation formation and preferences. Specifically, I develop an estimation procedure to identify preferences for college-majors using post-reform data, when market-clearing procedures change, and students clear the market in rounds of observing multiple offers and choosing to enroll or wait for a better offer. The setting allows me to observe direct choice between options and estimate preferences using standard revealed preference methods.

The estimated structural model enables me to conduct counterfactual analyses. First, I quantify the heterogeneous welfare and distributional effects of a centralizing policy change when all outside alternatives are private. Even further, extrapolating outside of my project's empirical setting, I simulate counterfactual policies that bring lower-SES students closer to the first best without sacrificing the welfare of high-SES students.

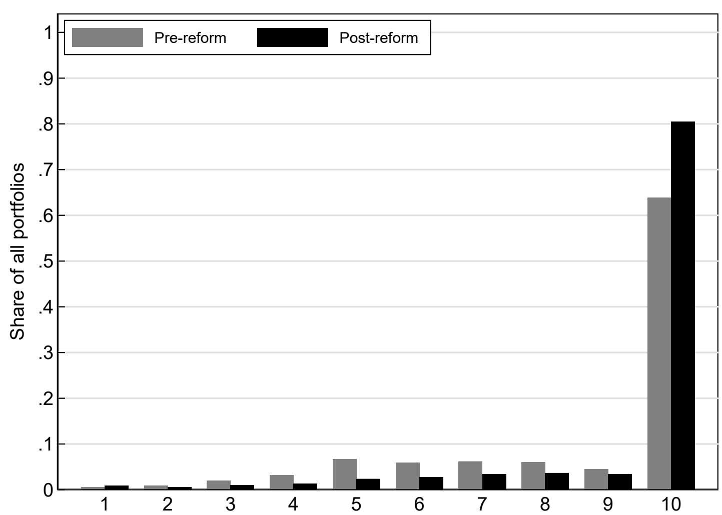
1.9 Figures

Figure 1.1: Number of Programs on the Centralized System



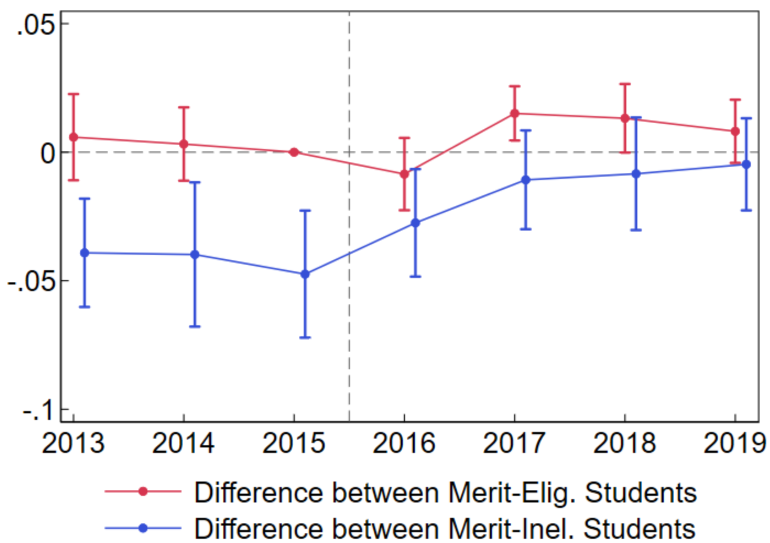
Note: Chart shows the number of participating programs in the centralized system. In 2016, all of the private universities in the country joined the centralized system. The increasing number of programs after 2016 reflects private universities in the system increasing their program offerings.

Figure 1.2: Application Sizes Before and After the Reform



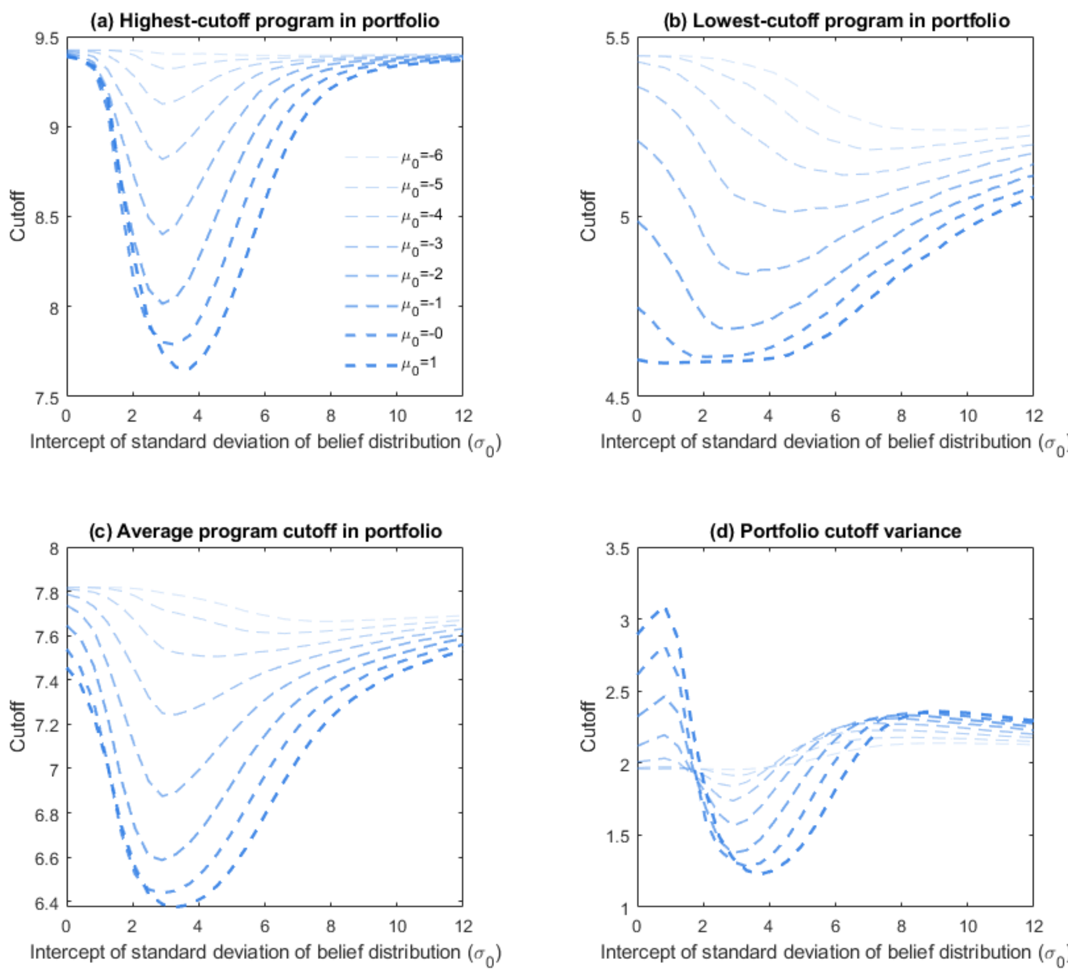
Note: Chart shows the number of programs submitted in an application portfolio before and after the reform.

Figure 1.3: Event Study of Double Differences in Selectivity of Most Selective Public Programs
Top and Non-Top Students



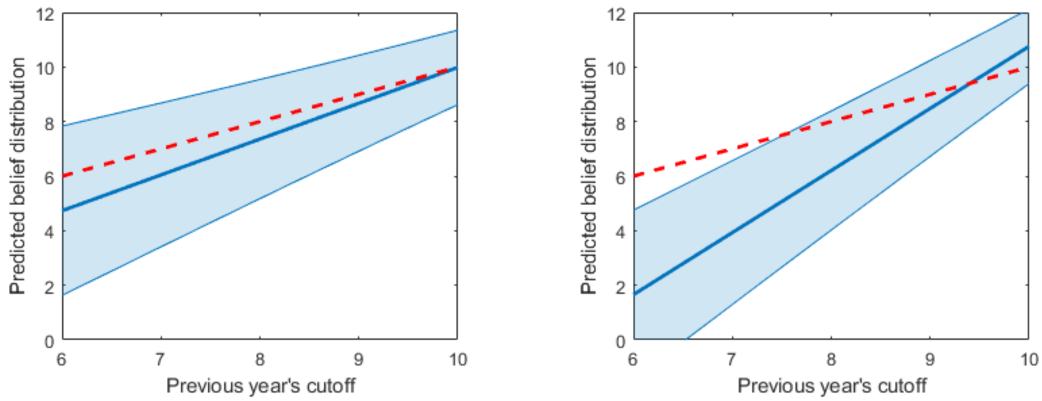
Note: This chart plots the differences in selectivity of most selective public programs on the centralized application between private high school and public high school students. Negative differences reflect less selective choices at the top of portfolio for public high school students. Regressions control for average exam score and include district FE. Standard errors clustered at district level.

Figure 1.4: Portfolio selectivity features under varying belief parameters



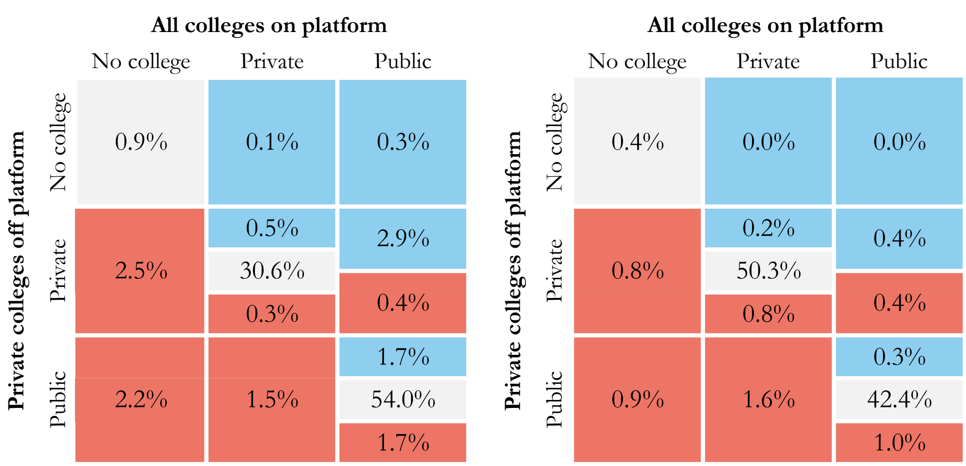
Note: The figure shows portfolio selectivity features when varying belief parameters μ_0 and σ_0. The remainder of both preference and belief parameters are held constant. The x-axis varies the intercept of the standard deviation of the belief distribution (σ_0) while each line corresponds to a different value for the mean shift intercept (μ_0). Legend for all panels is in panel (a).

Figure 1.5: Prediction of perceived program cutoff distribution using model estimates



Note: Plot shows the model-predicted distribution of perceived program cutoffs as a function of the previous year's cutoff. High-SES students' beliefs are plotted on the left panel and lower-SES students' beliefs on the right. The blue line in the middle of the shaded area is the mean of the distribution calculated as $\bar{c} = \text{cutoff}^{y-1} + \hat{\mu_0} + \hat{\mu_1} \times \text{cutoff}^{y-1}$. The shaded area reflects the scores that are within a standard deviation of the mean for the estimated standard deviation of the beliefs. I compute the bounds of the shaded area as $\bar{c} \pm \log(1 + \exp(\hat{\sigma}_0 + \hat{\sigma}_1 \times \text{cutoff}^{y-1}))$. I add a y=x line as a reference.

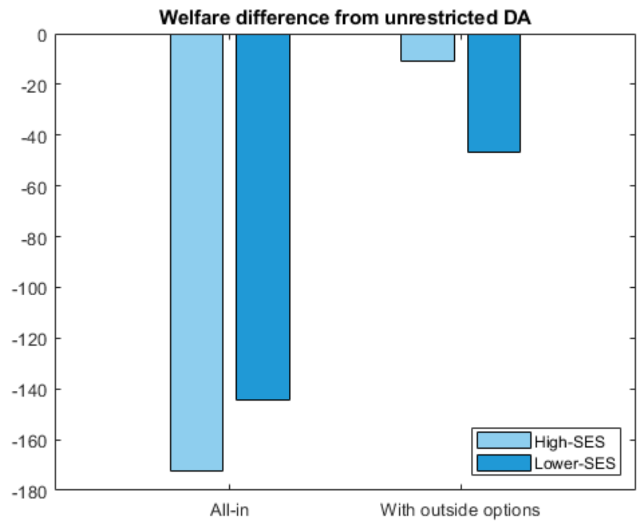
Figure 1.6: Assignment results changing market structure from partitioned to all-in



(a) Lower-SES Students **(b) High-SES Students**

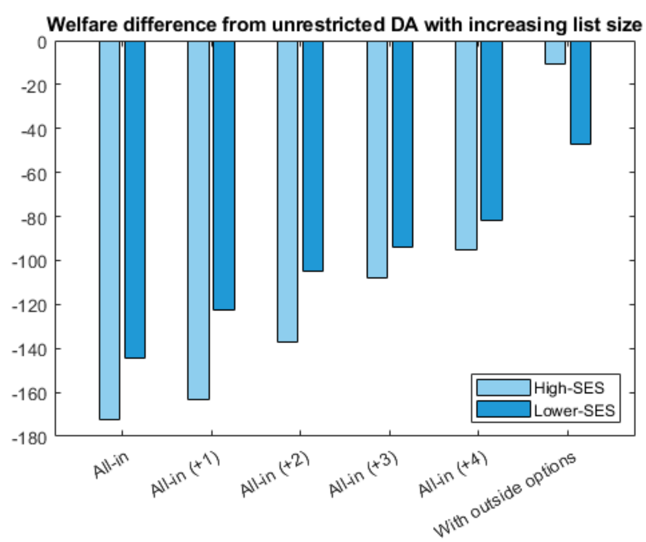
Note: This figure shows the share of winners (blue) and losers (red) among graduating high school students when moving from a partitioned market structure to an all-in one. The share of students with an identical outcome under both regimes are shown in gray.

Figure 1.7: Welfare results under two market structures by SES



Note: The figure shows welfare differences in Euros between outcomes from an unrestricted DA mechanism and each of the two counterfactual market structures by SES group.

Figure 1.8: Welfare results by SES under varying list sizes and an all-in structure



Note: This chart shows welfare differences from welfare achieved under an unrestricted DA with an all-in structure for assignments under varying list sizes for high and lower-SES groups.

1.10 Tables

Table 1.1: Summary statistics on applicants and programs

	Pre-Reform (2013-2015)			Post-Reform (2016-2019)		
	All	Public HS	Private HS	All	Public HS	Private HS
a. Applications on the platform						
Number of applicants	84,931	72,766	12,165	98,459	82,499	15,527
Share of applicants		0.86	0.14		0.84	0.16
Share from capital	0.20	0.20	0.25	0.23	0.22	0.28
Exam score average	6.99	6.92	7.41	7.30	7.22	7.70
Application portfolio size	8.60	8.61	8.60	9.19	9.18	9.26
Public share of portfolio	1.00	1.00	1.00	0.89	0.91	0.79
	All	Public Uni.	Private Uni.	All	Public Uni.	Private Uni.
b. Programs on the platform						
Number of colleges	12	12	0	38	12	26
Number of programs	289	289	0	517	309	208

Notes: Application data come from the Center for Educational Services for 2013-2015, publications of the Ministry of Education for 2016-2017, the Academic Network of Albania for 2019, and individual colleges for 2018. Exam score data come from publications of the Center for Educational Services. The averages of exam scores and portfolio sizes exclude 2018 as application data are missing for some colleges for that year.

Table 1.2: SES differences in selectivity of application and enrollment

	Application Selectivity			Enrollment Selectivity	
	Top ranked	Second ranked	Third ranked	Cutoff	Rank
Public HS	-0.210***	-0.192***	-0.196***	-0.054***	3.896***
	(0.011)	(0.010)	(0.010)	(0.010)	(1.040)
Priv. HS Mean	9.118	9.078	9.056	8.682	121.993
Adjusted R^2	0.243	0.282	0.289	0.406	0.445
Obs.	50,947	50,753	50,456	45,595	45,595
Condition on Score	✓	✓	✓	✓	✓
District-by-year FEs	✓	✓	✓	✓	✓

Notes: Sample includes years 2014-2015 as application selectivity is calculated using the previous year's cutoff. Data from the year 2013 are only used to calculate selectivities, but are excluded from the sample. Standard errors are clustered at the high school level.

Table 1.3: Relationship Between Attending a Public HS and Final Assignment Outcomes

	National Outcomes			Outcomes in Capital		
	Assigned to Top Choice	Assigned to a Top-Three Choice	Unassigned	Assigned to Top Choice	Assigned to a Top-Three Choice	Unassigned
Public HS	0.056***	0.091***	-0.019**	0.062***	0.111***	-0.009
	(0.012)	(0.014)	(0.009)	(0.019)	(0.025)	(0.018)
Private HS Mean	0.288	0.498	0.134	0.173	0.346	0.212
Adjusted R^2	0.109	0.154	0.221	0.123	0.209	0.268
Observations	84,931	84,931	84,931	17,336	17,336	17,336
Condition on Score	✓	✓	✓	✓	✓	✓
District-by-year FEs	✓	✓	✓			
Year FEs	✓	✓	✓	✓	✓	✓

Notes: Sample includes years 2013-2015. Private HS Mean is the unconditional mean of the outcome variable for students attending private high schools. Standard errors are clustered at the district level for the national sample and are robust for the capital-only sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 1.4: Effect of reform on application counts

	All applicants			Applicants who filled lists	
	Count of programs listed	Count public	Share public	Count public	Share public
Lower SES	0.010	0.040	0.004	0.048	0.005
	(0.134)	(0.126)	(0.003)	(0.030)	(0.003)
Lower SES x Post-Reform	-0.014	1.165***	0.133***	1.191***	0.119***
	(0.087)	(0.204)	(0.019)	(0.161)	(0.016)
High SES Mean Pre	8.680	8.680	1.000	10.000	1.000
High SES Mean Post	9.360	7.610	0.802	8.214	0.821
Adjusted R^2	0.087	0.064	0.194	0.195	0.195
Observations	109,092	109,092	109,092	81,252	81,252
Condition on Score	✓	✓	✓	✓	✓
District FEs	✓	✓	✓	✓	✓
Year FEs	✓	✓	✓	✓	✓

Notes: Count public is the number of public programs listed in the application. Share public is the share of programs listed that are in public universities. The first three columns show the difference in list sizes, count public and share public for all applicants. The last two columns restrict the sample to those applicants who filled their lists. Sample includes all applicants for years 2013-2019, excluding 2018. Data from year 2018 are excluded from this table as I only have applications to public programs for that year. Standard errors are clustered at the district level.

CHAPTER 2

How Costs Limit Contraceptive Use among Low-Income Women in the U.S.: A Randomized Control Trial

2.1 Introduction

For over half a century, two schools of thought have debated the determinants of childbearing. One emphasizes *demand* factors, arguing that pregnancies and contraceptive use are mainly determined by preferences, wages, and income (Blake, 1969; Easterlin, 1980; Becker, 1981; Pritchett, 1994a,b). The other stresses the primacy of factors impeding *access* to contraception, suggesting that barriers such as costs or information play a critical role in preventing women from achieving their desired childbearing (Harkavy, Jaffe, and Wishik, 1969; Ryder and Westoff, 1971; Knowles, Akin, and Guilkey, 1994).[1]

The resolution to this debate has important implications for public policy. Whereas policy can have immediate and direct effects on access to contraception, it is less clear whether policy makers could (or should attempt to) shape preferences, wages and incomes. The resolution could also affect the lives of many Americans. Around 40% of pregnancies in the U.S. occurred either sooner than desired or when no pregnancy was desired at any

[1]We understand and acknowledge that people of all genders give birth. For parsimony, this paper uses the word "woman," "mother," and female pronouns when discussing individuals who become pregnant or give birth.

point in the future (Kost, Zolna, and Murro, 2023). Prior to the U.S. Supreme Court's July 2022 *Dobbs* decision, which has allowed 15 states to restrict or eliminate abortion access (McCann et al., 2023), about 40% of undesired pregnancies ended in abortion (Kost and Lindberg, 2015). Children born from undesired pregnancies are more likely to have low birth weight and other birth complications (Mohllajee et al., 2007; Kost and Lindberg, 2015). Undesired births contribute to the cycle of poverty by decreasing women's educational attainment, employment, and family resources (Bailey, 2006; Bailey, Hershbein, and Miller, 2012; Bailey, Malkova, and McLaren, 2018; Miller, Wherry, and Foster, 2020) and limiting the life opportunities of children (Ananat and Hungerman, 2012; Bailey, 2013).

This paper presents evidence from the Michigan Contraceptive Access, Research, and Evaluation Study (M-CARES), which uses a randomized control trial to estimate the role of cost as a barrier to contraceptive choice among uninsured women. Although the Affordable Care Act (ACA) eliminated cost-sharing for contraception for those with health insurance, it did not affect costs for uninsured individuals seeking reproductive care through Title X—a national family planning program that offers patient-centered, subsidized contraception and reproductive health services to low-income individuals. In 2018, around 1.4 million Title X clients (or 40% of all 2018 Title X clients) were uninsured and faced substantial out-of-pocket costs for contraceptives after applying Title X discounts. Importantly, no market mechanism or public program allows individuals wishing to delay or avoid pregnancy to finance these out-of-pocket costs; these costs are paid upfront.

Between August 2018 and November 2019, M-CARES recruited 1,597 uninsured women aged 18 to 35 at twelve Planned Parenthood of Michigan (PPMI) health centers. Half of the participants were randomized to receive vouchers that could be used toward their out-of-pocket costs for contraception, and the other half (the control group) received the usual clinical care at usual costs. The study did not nudge, advocate for, or compel individuals to use any method of contraception. The premise of the study was that individuals, in consultation with their physicians, know best which method of contraception is best suited for them. Consequently, vouchers could be used for *any* contraceptive method at PPMI for up to 100 days after enrollment and varied by phase. In the first study phase, vouchers covered costs up to 50% of a name-brand intrauterine device (IUD). In the second phase, vouchers covered costs up to 100% of a name-brand IUD. With participants' consent, the study then collected information on participants' use of contraception in a follow-up survey and PPMI medical records over the next two years.

If financial access posed little barrier to using a preferred contraceptive method, the study would find that vouchers have no effect on women's use of contraceptives or choice of method. But the results show otherwise. Participants receiving 100% vouchers were 40% more likely to use any birth control, nearly doubled the value of the birth control they purchased, purchased contraceptives covering around 328 more days, and switched to more effective methods. Over one-third of 100%-voucher recipients switched to a more effective method versus one-quarter in the control group. Among 100% voucher recipients, the likelihood of choosing a long-acting, reversible method (LARC, either an IUD or implant) increased by 324%. These effects persist two years after study enrollment, which implies that the voucher resolved a binding, long-term constraint limiting women's ability to use their desired method of contraception.

A comparison of the effects of the 50% and 100% vouchers also sheds light on a highly relevant public policy choice: how generous of a subsidy is required to enable women to use their preferred method? Doubling the voucher subsidy from 50% to 100% more than quadrupled the relative effect size for LARCs, raising the treatment effect from 77% to 324%. This large increase in the relative effect shows that even 50% of the already discounted Title X price remains prohibitive for uninsured women. Eliminating cost-sharing—as with the 100% voucher—allows many more to use their preferred contraceptive method.

We also explore heterogeneity in these effects to better understand which groups were more financially constrained in their choice of contraceptive methods. Subsidizing contraceptives has large and similar effects on contraceptive efficacy for a broad set of pre-specified subgroups, including stratifications by race and ethnicity, education, relationship/marital status, religiosity, and having a usual place for reproductive health care. For women who were not planning to get a LARC at the time of enrolling, receiving a voucher that eliminated cost-sharing increased more than eight times the likelihood that they changed course and elected to use a LARC. Women with children appear to be one of the most financially constrained groups. One fifth of mothers receiving the 100% voucher chose to use a LARC versus just three percent in the control group. These findings allow a simple cost-benefit calculation of scaling this RCT's intervention to make all contraceptives free through Title X. The findings imply that a U.S. policy eliminating out-of-pocket costs for all Title X patients would reduce pregnancies by 5.3%, birth rates by 3.9%, and abortions by 8.3%. We also find that the increased costs of such a policy would be offset by reductions in federal health care spending through Medicaid, resulting in a net savings of $1.43 billion in the first year alone.

This study contributes to the literature by using the gold-standard of causal inference—a randomized control trial—to address a highly relevant policy question that has been almost exclusively studied in observational and quasi-experimental settings. One of the most influential studies to date, the St. Louis Contraceptive Choice Project (CHOICE), found that giving no-cost LARCs to study participants affects birth rates (Secura et al., 2010; Mestad et al., 2011; McNicholas et al., 2014; Birgisson et al., 2015; Broughton et al., 2016). Because CHOICE had no control group, this study's design makes it difficult to interpret these findings as reflecting costs alone (Bailey and Lindo, 2018).[2] In addition, this study contributes evidence collected in the current policy environment. Previous studies with credible quasi-experimental designs consider contexts prior to the Affordable Care Act's (ACA) expansion of health insurance coverage and mandate that insurance policies cover the costs of contraception for millions of American women (Becker, 2018; Carlin et al., 2016; Dalton et al., 2020;

[2]CHOICE's research design compares outcomes for women who enrolled in the study (who wanted to start a new contraceptive method) to similarly aged women in the greater St. Louis area, who differed from study participants in that they were seeking reproductive care. Perhaps unsurprisingly, the group of women in CHOICE were less likely to give birth than the broader population. Other RCTs in the U.S. have been limited to adolescents and have not included the broader population of women facing high costs of contraception (Kirby, 1997; DiCenso et al., 2002).

Heisel et al., 2018).[3] Third, no study to date is able to link out-of-pocket costs to women's choice of contraceptive methods or identify individual characteristics that mediate or moderate the effects of costs. This study's individual-level randomization allows the consideration of differences in sensitivity to cost across groups as well as predictors of this sensitivity, allowing a novel characterization of who is most affected by the cost of contraception.

2.2 Theoretical Framework and Hypotheses

A simple theoretical framework by Michael and Willis (1976) combines elements of the neoclassical model of demand for children (Becker, 1960, 1965; Becker and Lewis, 1973; Willis, 1973) with a model of contraceptive use (Sheps, 1964; Sheps and Perrin, 1966)—an innovation that relaxes the neoclassical assumption that fertility regulation is costless. Each contraceptive method j is associated with a fixed and marginal price per birth prevented, and pregnancy occurs probabilistically rather than deterministically as in the standard neoclassical model. The number of children is a random variable, and women choose a method j to reduce the monthly probability of conception, which is equivalent to choosing an expected distribution of the number of pregnancies, summarized by the first, μ_j, and second moment, σ_j. Women maximize their utility by weighing the marginal costs of preventing pregnancy using different contraceptive methods against the marginal benefit of attaining different expected distributions.

Closely related to the M-CARES intervention, the model distinguishes between the fixed and marginal costs of a contraceptive method. The total cost of using method j to attain an expected pregnancy distribution, μ, is given by $\pi_j = \alpha_j + \beta_j(\mu_N - \mu)$, where $_N$ is the mean of the distribution of pregnancies in the absence of any contraception. The fixed cost of using a method j is α_j, which includes any out-of-pocket costs, the fixed cost of going to the doctor,

[3]Previous quasi-experimental studies consider how expansion in federally funded family planning programs in the 1960s and 1970s reduced birth rates (Bailey, 2012); state-level expansions in Medicaid eligibility for family planning services in the 1990s and 2000s increased the use of contraception and reduced childbearing (Kearney and Levine, 2009); and Colorado's Family Planning Initiative (CFPI), which made LARCs free in 2009, reduced teen birth rates (Packham, 2017; Lindo and Packham, 2017).

and the cost of learning about a particular method (e.g., overcoming misinformation, personal circumstances, or other external factors). β_j is the marginal cost of preventing a pregnancy using method j. The marginal cost reflects behaviors (e.g., abstinence), inconvenience or discomfort at the time of intercourse (e.g., withdrawal or barrier methods like diaphragms or condoms), and the necessity of returning to fill a prescription (e.g., the pill or injections).

Figure 2.1 A plots an example of total costs for different contraceptive methods and pregnancies prevented. Different methods are optimal for women wishing to avoid different numbers of pregnancies. For instance, if a woman wishes to prevent two pregnancies, then method 1, which entails a small fixed cost but a high marginal cost (like condoms or withdrawal, represented by line Π_1), would be her lowest cost option. One wishing to prevent three births would choose method 2, paying a higher fixed cost but gaining a lower marginal cost. The high fixed but near zero marginal cost of method 4 is similar to LARC methods, which require an upfront, fixed investment of time and out-of-pocket payments, but have the lowest total cost for women seeking to prevent eleven or more pregnancies. The lowest-cost function for achieving an expected number of births before the M-CARES intervention is given by the dashed, lower envelope, or $C(\mu) = \min_j[\alpha_j + \beta_j(\mu_N - \mu)]$.

This model does not include behavioral biases and optimization missteps in the behavioral hazard literature (Baicker, Mullainathan, and Schwartzstein, 2015). Yet it clarifies the endogeneity of method choice to prices and suggests several testable hypotheses for M-CARES. First, the use of contraception is endogenous to both the demand for children (preferences, wages, income) and the costs of different contraceptives. Method use itself does not indicate that women are constrained by costs or motivated by other factors. M-CARES uses random assignment to circumvent this complication and compares women expected to have identical demand for children who face different fixed, out-of-pocket costs for contraceptives. Second, reducing the fixed costs of contraception would lead many women to adopt more effective, and lower marginal cost, methods, because many women seeking care are highly financially constrained. M-CARES vouchers reduce the fixed costs of contraception and shift the lowest-cost function downward as shown in Figure 2.1B. For instance, the 50% voucher reduces the fixed costs of contraceptive method 2 from Π_2 to Π_2'; the fixed cost of method 3 from Π_3 to Π_3'; and so forth. The lowest-cost envelope would, therefore, shift such that women seeking to prevent three to nine pregnancies would choose method 3, and women seeking to prevent nine or more pregnancies would choose method 4. Third, receiving a higher valued voucher should have larger effects on take-up of higher fixed cost methods such as LARCs, as shown in Figure 2.1C. For instance, the 100% voucher would reduce the lowest-cost envelope such that women seeking to prevent one to five pregnancies would choose method 3 and women wishing to prevent five or more pregnancies method 4.

2.3 M-CARES Methods

M-CARES recruited women at 12 PPMI health centers to participate in a randomized control trial.[4] PPMI is Michigan's largest Title X service provider, and Planned Parenthood affiliates served 40% of the 4 million Title X clients in the U.S. in 2018, making this study's context and focus on the costs of contraception highly policy relevant to Title X providers today. M-CARES's goal is to support reproductive autonomy by eliminating cost barriers: the vouchers should make any desired method of contraception more financially accessible or free. The analysis covers the period between August 20, 2018, and November 3, 2019, before Planned Parenthood withdrew from the Title X program, increasing its prices and altering other operations and the trial.

2.3.1 Study Enrollment, Randomization, and Sample Inclusion

Study eligibility required that participants were (1) females ages 18 to 35, (2) at risk of unintended pregnancy, (3) facing out-of-pocket costs for contraceptives, and (4) at PPMI for

[4]The trial protocol is approved by the University of Michigan's Health Sciences and Behavioral Sciences Institutional Review Board (HUM00132909) and registered at clinicaltrials.gov (NCT03673007). A pre-analysis plan for the first year is available at Open Science Framework and the American Economic Association RCT Registry.

a clinician visit. Criteria (1) and (2) ensure that participants are legal adults, biologically capable of pregnancy, and are not pregnant at the time of enrollment or wishing to become pregnant in the next year. Out-of-pocket costs for criterion (3) are determined using PPMI's income assessment during check-in. PPMI does not charge patients with incomes below the poverty line for services, so this group had no out-of-pocket costs and were excluded from the study. Criterion (4) was logistically necessary, because few patients without clinician visits remained in the waiting room long enough to complete the screening and enrollment process. The study did not require that participants be visiting PPMI to obtain contraception.

Professionally trained NORC field interviewers recruited patients in the waiting room to complete a 5-minute screening survey, which was compensated with $10.[5] If a patient met the inclusion criteria and was willing to participate, a tablet led her through the informed consent with optional assistance from the NORC interviewer. Consenting participants were randomly assigned in a 1:1 ratio to receive a voucher. After the appointment, interviewers invited participants to complete a baseline survey, and the participants were compensated with $60 for taking the survey in the health center on the same day. Participants unable to complete the survey in the health center received a link by email/text to complete the survey later for $40.

Figure 2.2 documents the enrollment, randomization, and inclusion in the analysis samples. 2,561 participants met eligibility criteria (1), (3), and (4) and agreed to take the 5-minute screening survey on a tablet. 1,603 patients met all inclusion criteria, were able to enroll before their appointment began, and elected to participate. 819 received vouchers, and 784 were assigned to the control group. After randomization, two participants withdrew from the voucher group, and four withdrew from the control group. All but 16 participants were linked to PPMI billing records (10 in the control and six in the treated group). The baseline survey, which contains information for subgroup analyses, achieved a response rate

[5]NORC is a non-partisan research organization at the University of Chicago that specializes in survey research.

of 79%, which did not differ between the treatment and control groups.[6]

Table 2.1 compares M-CARES participants (column 1) to a nationally representative sample of women ages 18-35 from the 2017-2019 National Survey of Family Growth (NSFG, column 2) and to the characteristics of 2018 Title X clients reported in the Health and Human Services (HHS) Family Planning Annual Report (column 3) {Fowler, 2019 #1803}. Relative to the NSFG, M-CARES participants are slightly younger, less likely to be a racial or ethnic minority, and significantly more likely to have lower incomes. They are also less likely to use contraception than the national sample of women. The M-CARES sample also differs in expected ways from the national population of Title X patients. While similar in the use of birth control, the M-CARES sample contains no one with income below than the federal poverty line (per the study inclusion criteria). In addition, the M-CARES sample is less likely to be Hispanic, owing to this group's underrepresentation in Michigan, and less likely to be Black, owing to this group's underrepresentation in the areas served by Planned Parenthood health centers participating in M-CARES.

Table 2.1 also documents balance in the intervention (column 4) and control groups (column 5) in pre-specified patient characteristics, including contraceptive methods used before enrollment, age, race/ethnicity, marital/cohabitation status, income as percent of federal poverty line, and previous childbearing. Consistent with randomization, these characteristics do not jointly predict voucher receipt (F-statistic of 0.97, p-value=0.50). Our main specifications include indicator variables for race and education to account for the slight imbalance between these groups that occurred by chance.

2.3.2 The Intervention and Voucher Amounts

By design, this RCT sets aside other aspects of access to contraception and focuses on the role of costs as a barrier to use. Consequently, the study altered as little as possible

[6] To evaluate systematic non-response, we regress a binary variable equal to 1 if a participant completed the baseline survey/0 otherwise on voucher receipt and correct standard errors for heteroskedasticity. The estimate of the effect of receiving a voucher on response is 0.0072 (se: 0.021).

relating to the health center operations. Following randomization, recipients were handed an M-CARES card with their study number, an email address to contact the study, and voucher amount. They were also sent a text and email with the same information in case they lost the M-CARES card. Recipients were told that vouchers could be used to pay for any contraceptive and related services at PPMI *for 100 days after enrollment*.[7] The voucher could not be used for an abortion, because Title X funds do not cover abortion. The 100-day time limit allowed recipients to return to PPMI to use their vouchers, which was enough time to get two shots of Depo-Provera (each lasts 90 days) or have an IUD inserted (which often requires a return visit). We used a deadline to help minimize procrastination, which could lead participants to forget about or lose the voucher (Ariely and Wertenbroch, 2002; O'Donoghue and Rabin, 1999). Surveyors informed voucher recipients that M-CARES would pay for removal of any device funded by the voucher within one year of enrollment.

Participants assigned to the control group were also handed an M-CARES card with their study number and an email address to contact the study team with any questions or concerns. The card had no voucher amount filled in, and these participants received the usual clinical care with costs determined by the standard PPMI sliding scale as described below.

Participants in both the voucher and control groups were handed a standard information sheet about the effectiveness of different contraceptive methods. Following enrollment, participants proceeded with their pre-scheduled appointments with PPMI clinicians with no involvement from M-CARES.

Voucher amounts were initially chosen to make any contraceptive up to the cost of the lowest-cost LARC free of charge after applying the PPMI sliding scale in the first stage of the study. Vouchers were applied at check-out by PPMI, similar to a gift card. When the study started, PPMI indicated that the lowest cost LARC was a Liletta IUD, which cost half as much as name-brand devices (e.g., Skyla, Paragard, and Mirena). During the first

[7] "Related services" are those medically required to use a contraceptive. For example, inserting an IUD requires a pregnancy test.

study phase from August 20, 2018, to March 3, 2019, all contraceptives at PPMI up to the out-of-pocket costs of a Liletta insertion were free for voucher recipients. PPMI charges patients with incomes at 101-150% of the poverty line 25% of the total costs for services; those with 151-200% of the poverty line 50%; 201-250% of the poverty line 75%; and 251% or above the poverty line 100% for the services they receive. Voucher amounts were, therefore, $123, $246, $369, and $492 for the respective income categories (Appendix Table B.1). The voucher could be used for less expensive methods, such as birth control pills, injections, rings, and hormonal patches, or more expensive methods, such as name-brand IUDs or an implant. However, participants had to pay any costs above the voucher value out of pocket. PPMI's sliding fee scale means that out-of-pocket costs depend on a woman's income relative to the federal poverty line.

In early 2019, the M-CARES team learned that Liletta was only rarely stocked or inserted by PPMI. This meant that—although the voucher was intended to make the lowest cost, available LARC free—the voucher had only covered 50% of the cost for available IUDs. The study team subsequently increased voucher amounts to cover the costs of the available, name-brand IUDs as of March 4, 2019. The cost of insertion and related services did not change, so the amount of the voucher almost doubled in the second study phase. Voucher amounts were $223, $446, $669, and $892 for women with incomes at 101-150%, 151-200%, 201-250%, and 251% or above of the poverty line, respectively. Our analysis refers to the period before March 4, 2019, as the 50% phase, and the period on or after March 4, 2019, as the 100% phase.

On November 4, 2019, Planned Parenthood withdrew from Title X due to new Trump Administration requirements that organizations providing both family planning and abortion services must physically separate these services in order to receive federal funding, affecting both PPMI pricing and operations. This paper, therefore, analyzes the period from August 20, 2018, when recruitment started, to November 3, 2019, which informs the causal effect of providing a 50% and 100% voucher for contraceptives to low-income women with out-of-pocket costs.

2.4 Outcomes and Research Design

M-CARES combines survey and administrative data to create five pre-specified primary outcomes capturing different dimensions of contraceptive efficacy: (1) the dollar value of services purchased; (2) a binary measure for whether any contraceptives were purchased; (3) a binary measure of LARC insertion; (4) the likelihood of a pregnancy within one year based on the CDC failure rate with typical use of the most effective method purchased (Trussell 2011); and (5) the days covered by the most effective contraceptive method purchased.[8] Following Kling, Liebman, and Katz (2007), we also create an index of contraceptive efficacy that combines these five outcomes to summarize the *overall* effect of receiving a voucher and limit the number of statistical tests. The index is constructed as the arithmetic mean of its component z-scores,

$$ContraceptiveEfficacy_i = \frac{1}{5}\sum_{o=1}^{5}\frac{y_i^o - \bar{y}^{o,c}}{\sigma^{o,c}}.$$

y_i^o is the value of outcome o for individual i, $\bar{y}^{o,c}$ is the mean of outcome o and $\sigma^{o,c}$ is the standard deviation of outcome o in the control group by study phase. Note that we reverse code outcome (4) as one minus the failure rate, so that a positive value indicates a higher efficacy contraceptive.

We estimate the reduced-form effects of receiving a voucher for contraceptives using the following linear specification separately by phase,

$$Y_i = \tau_1 Voucher_i + \mathbf{X}_i'\beta + \varepsilon_i, \tag{2.1}$$

[8] Days of coverage is the number of days that a purchased unit covers multiplied by the number of units purchased. Unit coverage is 1095 days (3 years) for implants, 2190 days (6 years) for Liletta, 1825 days (5 years) for Mirena, 3650 days (10 years) for Paragard, 1095 days (3 years) for Skyla, 28 days for birth control pills, 90 days for Depo-Provera injections, 1 day for diaphragm, and 28 days for rings.

where Y_i is one of the five measures of contraceptive efficacy above or the index of contraceptive efficacy, which combines them; $Voucher_i$ is a binary variable equal to 1 if an individual i was randomly selected to receive a voucher and 0 otherwise; $\mathbf{X}_i$ is a vector of exogenous covariates, including indicator variables for race and education account for slight imbalance in these characteristics in Table 2.1, and indicators for the patient's income relative to the poverty line, which determine the PPMI sliding scale and level of the voucher.[9] Standard errors are corrected for heteroskedasticity (Huber, 1967; White, 1980). This "intention-to-treat" (ITT) estimate captures the net, causal effect of providing a voucher to women seeking reproductive health care, which could be used for 100 days toward the purchase of any contraceptive.[10]

Another relevant policy question is: what is the causal effect of the voucher among women who used it? To answer this question, we estimate the local average treatment effect (LATE) of receiving a voucher using two-stage least squares (2SLS), which also allows us to explore heterogeneity in the causal effect of treatment on the treated across the two study phases and pre-specified subgroups. The first-stage equation is,

$$1(\text{Used Voucher} = 1) = \pi_1 Voucher_i + \mathbf{X}_i' \pi_2 + \varepsilon_{2i}, \tag{2.2}$$

and the second-stage is,

$$Y_i = \delta_1 Voucher_i + \mathbf{X}_i' \delta + \varepsilon_i. \tag{2.3}$$

The estimate of δ_1 is given by the ratio of the reduced form and first stage coefficients (τ/π_1).

The causal interpretation of the 2SLS estimate as the treatment effect of reducing out-of-

[9]Our pre-analysis plan explained that the inclusion of covariates "is intended to increase precision by accounting for differences in characteristics between the treatment and control groups that occur by chance" (p. 12). Slight imbalance in race and education characteristics in Table 2.1 led us to include indicators for race and education. Results without covariates are available upon request.

[10]Our Appendix A3 presents alternative estimates using dollars spent as the first stage outcome.

pocket costs for contraception on the treated requires several assumptions: (1) exogeneity, (2) excludability, and (3) monotonicity. Randomization ensures that exogeneity holds. Excludability requires that receiving a voucher only affects contraceptive efficacy only by reducing out-of-pocket costs. This assumption seems plausible as the voucher can only be used for purchasing contraceptives at PPMI. Moreover, women in both the treatment and control groups receive cash benefits for completing the surveys, implying that any effects of these cash benefits should be the same in the two groups. It is possible that receiving a voucher may have other effects on outcomes (e.g., a voucher can imbue a recipient with a positive or optimistic feeling), but it seems unlikely that this indirect effect would have a large effect on contraceptive efficacy over two years. Monotonicity requires that, if participants were moved from the control to the treatment group, their contraceptive efficacy would not decrease (or vice versa). While it is not possible to test this directly, there is little reason to believe that receiving a voucher would induce participants to reduce their contraceptive efficacy.

Under these assumptions, we interpret the 2SLS estimate as the local average treatment effect, or LATE (Imbens and Angrist, 1994), of reducing out-of-pocket costs on contraceptive efficacy. The 2SLS estimate, δ_1, identifies the causal effect of receiving a voucher among the women who shift their use of contraceptives after receiving a voucher and who would not have shifted their use without the voucher.

2.5 The Effect of Subsidizing Contraception on Use

A central question of the study is whether out-of-pocket costs affect patients' use of contraception or their choice of method. If patients' choice of method is not driven by financial constraints, voucher dollars may simply crowd out money that patients would have spent in the absence of the intervention. Figure 2.3 shows decisively that financial constraints play a large role in patients' choice of contraceptives, both in the immediate term of the first 100 days after enrollment (panel A) and longer-term over the next two years (panel B). Our discussion focuses on the LATE, but Figure 2.3 also presents ITT effects for the interested reader.[11]

The 50% voucher increased PPMI charges by 89% (LATE, \$268 over a control mean of \$300) and, in the 100% phase, by 144% (\$413 over a control mean of \$287). The "+++" symbol to the right of the estimates in the 100% phase indicates that the 100%-voucher effect was also significantly larger than in the 50% phase at the 1% level. The length of the bar on the right side of the figure indicates the percent increase in the LATE over the control mean along with the 95% confidence interval. (Note that the LATE abstracts from changes in take-up in the two periods, as indicated by the first stage, and the percent change in the LATE over the control mean accounts for time-varying factors affecting the patient population and health center operations). Recipients of the 50% and 100% vouchers were 56% and 69% more likely to purchase contraception relative to the control group, and they also purchased more effective methods and more days of coverage. The 50% voucher more than doubled LARC use (0.09 relative to a control mean of 0.07), and the 100% voucher increased LARC use by roughly five times the control group (0.22 over a control mean of 0.04). Doubling the value of the subsidy more than doubles the percent increase in LARC take-up, which indicates that, even at half price relative to the subsidized Title X sliding scale, out-of-pocket costs dissuade many women from choosing their preferred methods. Voucher recipients also chose methods or purchased more of their preferred methods. The period covered by contraceptive purchases increased by 152% (280 days) in the 50% phase and 346% (503 days) in the 100% phase, minimizing the need to return to the health center. Altogether, the voucher allowed women to shift to more effective methods, reducing the expected one-year incidence of pregnancy by 0.32 in the 100% phase and 0.27 in the 50% phase.[12] Summarizing over the five primary outcomes, the index of contraceptive efficacy

[11] Note that percent changes in the text may differ from what is implied in the figure due to rounding.

[12] Appendix Table B.2 summarizes these method changes. 36% of 100%-voucher recipients switched to a more effective method versus 25% in the control group. 61% of women in the voucher group stayed on the same method or did not purchase any contraceptives at PPMI compared to 72% in the control group. Only 3% switched to less effective methods in both the control and treatment groups.

increased by 0.60 of a standard deviation in the 50% phase and 1.0 of a standard deviation in the 100% phase.

Figure 2.3B examines the long-term effect of the intervention using data at two years after the participant enrolled in M-CARES, which sheds light on whether the voucher hastened contraceptive purchases by a few months or resolved a binding, long-term credit constraint. The results point to the latter. While the effects relative to the control mean fall over time as more individuals in both the treatment and control group purchase more contraceptives, the gaps between the treatment and control group remain large and highly statistically significant after two years. The voucher's effect on the value of contraceptive purchased (PPMI charges) remained at $200 after two years for the 50% phase participants (vs. $268 at 100 days) and at $319 for the 100% phase participants (vs. $413 at 100 days). Its effect on the use of a LARC remained 0.06 higher after two years for the 50% phase participants (vs. 0.09 at 100 days) and 0.19 for the 100% phase participants (vs. 0.22 at 100 days). The effect on the expected one-year incidence of pregnancy remained 0.23 lower after two years for the 50% phase participants (vs. 0.27 at 100 days) and 0.30 lower for the 100% phase participants (vs. 0.32 at 100 days). The "set-it-and-forget-it" nature of LARCs along with the near zero failure rate suggests that the voucher's effect on pregnancy could last from 3 years (e.g., implants) up to 10 years (e.g., Paragard IUD). Summarizing over the five primary outcomes, the index of contraceptive efficacy remained 0.38 standard deviations higher after two years for the 50% phase and 0.68 standard deviations higher after two years for the 100% phase. These modest reductions indicate that vouchers hastened some contraceptive purchases, but the treatment effects are highly persistent.

We also explore how making LARCs free allowed women to follow through on their plans or induced others to change them (Appendix Figure B.1.c). To study this, a screening survey prior to randomization asked respondents what they planned to do during their PPMI appointment that day. If they answered, "get family planning services," the survey asked what methods they "*planned* to get" [emphasis added] that day. In the 100% phase control group, 29% of the women who indicated that they planned to get a LARC that day followed through within the next 100 days. In treatment group, this number was more than three times higher at 73%. This indicates that just over two thirds of the gap between women's plans and their follow-through is explained by the high price of LARCs, with the remaining gap reflecting factors not considered in this study (e.g., medical contraindications). For the 86% of participants who did not plan to get a LARC in the 100% phase, less than 1% got a LARC within 100 days. In the treatment group, more than 8% did—a statistically significant effect over eight times as large as in the control group. In short, making LARCs free allowed more women to follow through on their plans and allowed others to change plans to get their preferred method.

Figure 2.4a summarizes treatment-effect heterogeneity by pre-specified demographic subgroups and Figure 2.4.b by baseline survey question answers. Because the index is constructed separately by phase, but not separately for each sub-group, the control mean for subgroups differs from zero. Per our pre-analysis plan, we use the index of contraceptive efficacy to increase statistical power to detect same-signed changes in efficacy across the five outcomes for these smaller subgroups. (Appendix Figure B.1a —Figure B.1.e present estimates for each of the primary outcomes separately for the interested reader.) For the 100% phase, the treatment effect among women who used the voucher on the index of contraceptive efficacy is highly statistically significant and exceeds 0.78 standard deviations for all demographic subgroups in Figure 2.4a. The effect is 0.94 standard deviations for White, Non-Hispanic women; 1.09 for Hispanic/Latina women; and 0.78 for Black women (statistically different from White, Non-Hispanic women, p=0.043). The magnitudes of the effects differ somewhat across demographic groups, but the differences are not statistically different: 0.89 standard deviations for women below the median age of 26 versus 1.13 for women above the median age; 0.94 standard deviations for women with less than a Bachelor's degree versus 1.04 for women with a Bachelor's degree or more; 0.84 standard deviations for married or cohabiting women versus 1.13 among those who are single. The effects are also slightly larger among higher income women than lower income women.

Figure 2.4.b stratifies on other pre-specified baseline characteristics. The treatment effects are large and significantly different from zero in most of the pre-specified sub-groups but not significantly different among these groups. Financial constraints appear similarly binding for women with and without a usual place to obtain birth control, those using highly effective and less effective methods at enrollment, and women with different beliefs that they will achieve their career aspirations. The effects are statistically larger for women in the 100% phase who delayed getting birth control in the previous year relative to those who did not (p-value=0.022), which is consistent with financial barriers playing a role in their choices to delay getting contraception before the trial. In addition, the effects of the voucher were significantly larger for women with more positive desires to avoid childbearing relative to women with less strong feelings (pvalue=0.019).

2.6 Conclusion: Subsidizing Contraception Facilitates Take-Up of More Effective Methods

This study shows that the choice of contraceptive methods is highly sensitive to out-of-pocket costs. Reducing high out-of-pocket costs has large and persistent effects on women's ability to choose their preferred methods, especially when that method is more effective (and more expensive). These effects attenuate only slightly over two years, suggesting that the intervention did not simply hasten the use of a preferred method. Making contraception free eliminated a binding, long-term financial constraint that significantly limited women's reproductive autonomy. This finding held in all age, race, and demographic groups we considered as well as across women with different aspirations, desires, and access to reproductive care. This study sets aside the consideration of other barriers to reproductive autonomy to focus on the role of out-of-pocket costs, and these other barriers remain important and fruitful directions for future research. A key finding is that making contraception more affordable could have large effects on women's take-up of their preferred methods and, ultimately, undesired pregnancies.

The study also sheds light on how the generosity of subsidies for contraception matter. Making LARCs half price increased take-up by 78% (ITT effect), whereas doubling the value of the voucher to 100% (making them free) increased take-up by 324%. That is, many contraceptive methods remain prohibitively expensive even with a 50% discount from the already subsidized Title X sliding scale.

As with any RCT, it is important to consider external validity, which in this case will be limited to populations seeking reproductive health care. To understand how a national policy making any contraceptive free at Title X providers could affect outcomes, we reweight the M-CARES sample to reflect the age, race/ethnicity, and income of the national Title X population (see Table 2.1, col 3)—all of whom were seeking care at Title X providers (Hainmueller, 2012; Fowler et al., 2019). If every Title X patient in the U.S. received free contraception up to the price of the lowest-cost LARC, the reweighted results indicate that pregnancies would fall by 21 pp (versus 19 pp unweighted, Figure 2.3, ITT estimate).

An important caveat is that reweighting does not account for treatment effect heterogeneity due to unobserved factors. For instance, treatment effects for Title X patients nationally may differ due to different state reproductive health care programs or policies or states' decisions to expand Medicaid coverage under the ACA (as Michigan did). In addition, these results may misstate the intervention's true effects on pregnancies if (1) low-income, uninsured women obtain contraception from other providers not observed in our data; (2) women do not use the most effective method purchased for one year (we use the one-year method failure rate as a summary metric); or (3) women make other adjustments in their sexual behavior to accommodate their contraceptive method. The first issue is not likely important in practice, because PPMI served 70% of all Michigan Title X patients in 2018, and Title X clients have few other options for affordable care. The quantitative importance of the second and third issues is harder to gauge, so they remain important caveats to the external validity of the results.

With these caveats in mind, we use the reweighted estimates to evaluate the implications of scaling the M-CARES intervention to the entire U.S.—implementing a federal policy making all contraception free for low-income Title X patients up to the cost of the lowest price IUD. Based on the income distribution of Title X patients historically and costs based on voucher use from M-CARES, a national policy making all contraceptives free would cost \$178 million annually—an increase of around 62% over current funding levels for the program.[13] Assuming the demand for children remained constant, we apply the reweighted estimate of the reduction in pregnancies with the 100% voucher (0.21) to the 1.4 million Title X patients nationally with out-of-pocket costs.[14] As with the M-CARES sample, all of the Title X patients sought reproductive health care. The results imply that the policy should reduce pregnancies by 301,000, or 5.3% relative to the estimated 2018 level. Using previously published estimates of the share of pregnancies that result in childbirth, these numbers imply a reduction in births of 144,000, or 3.9% from the 2021 level (Bailey, Bart, and Lang, 2022). Another consequence of eliminating cost-sharing for contraception for Title X clients is that the number of abortions would fall by around 77,000, or 8.3% relative to the 2020 level (Diamant and Mohamed, 2023). The number of births and abortions would continue to be reduced to some degree in later years, although these reductions in later years are less certain and are not included in these calculations.

The reduction in unplanned pregnancies resulting from the policy would also have immediate budgetary implications. Assuming that around 62% of the 144,000 births would be funded through Medicaid implies a reduction in Medicaid costs of more than \$1.61 billion in the first year of the policy.[15] That is, eliminating Title X cost-sharing for contraception for

[13]Estimates assume no increases in the use of Title X services due to an increase in funding generosity.

[14]Around 1.4 million individuals—36% of Title X clients nationally who are female, are uninsured, and have out-of-pocket costs—would be immediately affected by eliminating cost-sharing for contraceptives (Fowler, Gable, Wang, Lasater, and Wilson, 2019).

[15]5 This calculation uses Guttmacher's estimate of \$12,770 in 2010, which includes the costs of delaying prenatal care, labor and delivery, postpartum care, and 12 months of infant care and inflates this estimate using the health care inflation index (Sonfield et al., 2011). This inflation yields \$17,987 in 2022 dollars. Fowler et al. (2019) show that around 38% of Title X clients have private health insurance, implying that

low-income women would cost the federal government around \$178 million per year and reduce federal and state government spending by \$1.61 billion in the first year of the program, for a net savings to taxpayers of around \$1.43 billion in the first year of the program. Around 50% of the \$1.6 billion in total savings, or \$804 million, less \$178 million in additional Title X appropriations would accrue to the federal government under the FY 2024 FMAP rates (Kaiser Family Foundation, 2023). While the actual reduction in childbirth in the first year of the program could be more or less than what we estimate, this estimate would have to be too high by an order of magnitude to change the conclusion that a policy making contraception free to Title X clients would pay for itself. In addition, state governments would save the remaining \$812 million.

These estimates of cost savings are conservative because they do not account for the fact that some unplanned pregnancies will be deferred for more than one year and that some unplanned pregnancies are undesired, meaning that they may never occur in the future. This calculation also ignores likely revenue gains from more women remaining in the labor force (and paying taxes). Moreover, given the significant increase in unplanned childbirth expected in the aftermath of the *Dobbs* decision, free contraception could reduce births resulting from unplanned pregnancies by more than we estimate. Thus, the reduction in costs by expanding access to contraception could be more substantial.

62% of births to Title X clients will be paid by public insurance (i.e., Medicaid). We obtain \$1.61 billion by multiplying \$17,987 per birth by the reduction of 62% of the 142,000 unplanned births.

2.7 Figures

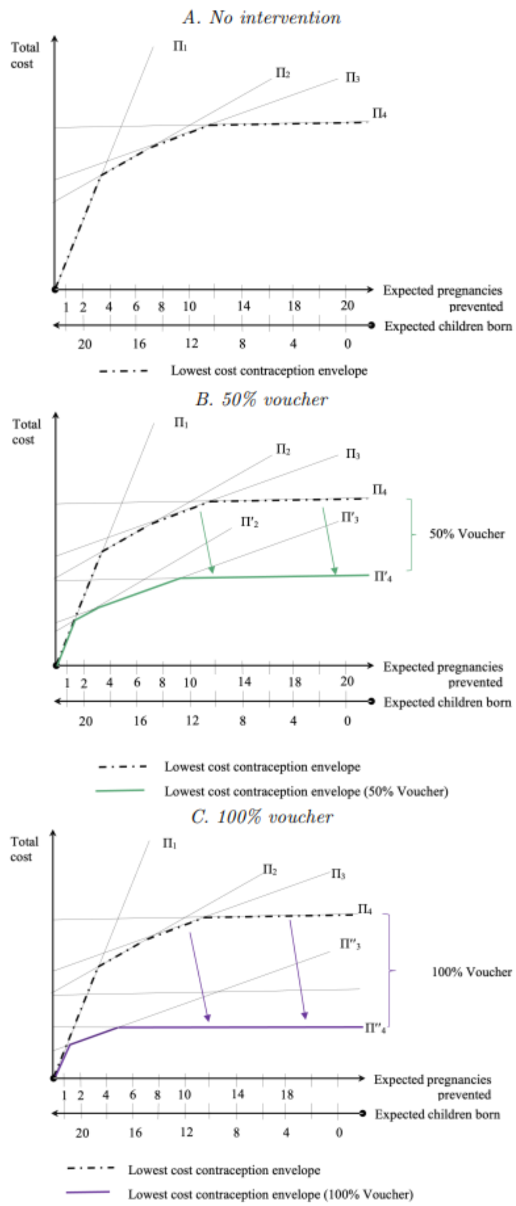


Figure 2.1: How Out-of-Pocket Costs and Vouchers Affect the Choice of Contraceptives

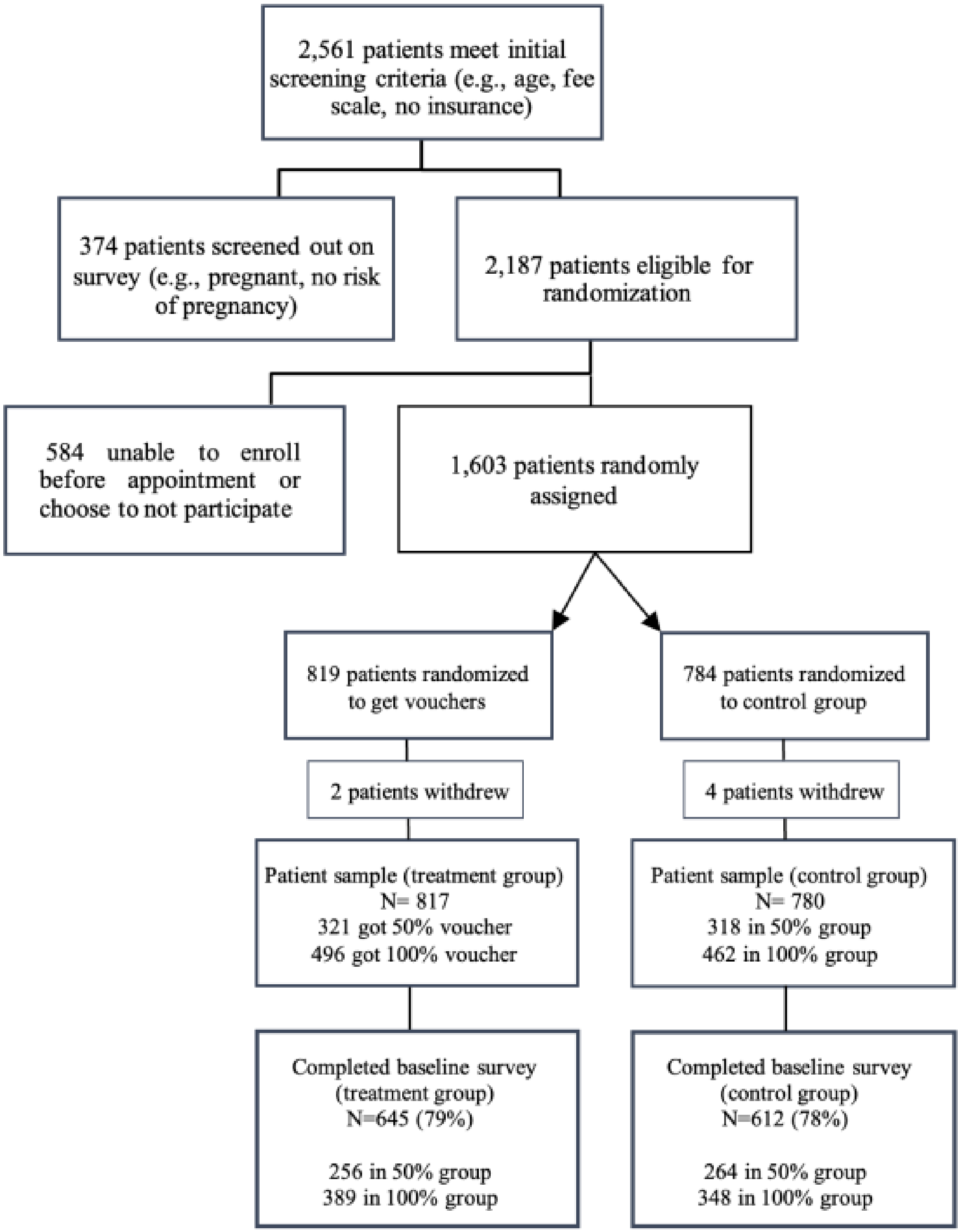


Figure 2.2: M-CARES Enrollment and Randomization of Patients

Notes: Participants in the 50% phase received vouchers between August 20, 2018, and March 3, 2019, valued at 50% of the cost of receiving a name-brand IUD. Participants in the 100% phase received vouchers between March 4, 2019, and November 3, 2019, valued at 100% of the cost of receiving a name-brand IUD.

	50% Voucher Group			100% Voucher Group		
	Control mean	ITT effect	LATE	Control mean	ITT effect	LATE
A. Contraceptive efficacy within 100 days of enrollment						
First stage: 1=Used voucher		0.72 (0.03)			0.66 (0.02)	
PPMI charges in dollars	300	181 (25)	268 (33)	287	270 +++ (22)	413 +++ (29)
Any birth control purchase	0.52	0.18 (0.04)	0.29 (0.05)	0.48	0.19 (0.03)	0.33 (0.04)
LARC insertion	0.07	0.05 (0.02)	0.09 (0.03)	0.04	0.14 +++ (0.02)	0.22 +++ (0.03)
Expected annual pregnancies	0.52	-0.17 (0.03)	-0.27 (0.04)	0.55	-0.19 (0.03)	-0.32 (0.04)
Temporal coverage	184	186 (48)	280 (66)	145	328 ++ (41)	503 ++ (57)
Index of contraceptive efficacy	0.00	0.39 (0.07)	0.60 (0.09)	0.00	0.65 +++ (0.06)	1.02 +++ (0.08)
B. Contraceptive efficacy within two years of enrollment						
First stage: 1=Used voucher		0.72 (0.03)			0.66 (0.02)	
PPMI charges in dollars	596	120 (41)	200 (55)	561	191 (33)	319 (46)
Any birth control purchase	0.57	0.15 (0.04)	0.25 (0.05)	0.52	0.18 (0.03)	0.31 (0.04)
LARC insertion	0.11	0.03 (0.03)	0.06 (0.04)	0.08	0.12 ++ (0.02)	0.19 +++ (0.03)
Expected annual pregnancies	0.47	-0.14 (0.03)	-0.23 (0.04)	0.52	-0.17 (0.03)	-0.30 (0.04)
Temporal coverage	350	135 (56)	220 (76)	283	265 + (46)	427 ++ (64)
Index of contraceptive efficacy	0.00	0.23 (0.06)	0.38 (0.08)	0.00	0.41 ++ (0.06)	0.68 ++ (0.08)

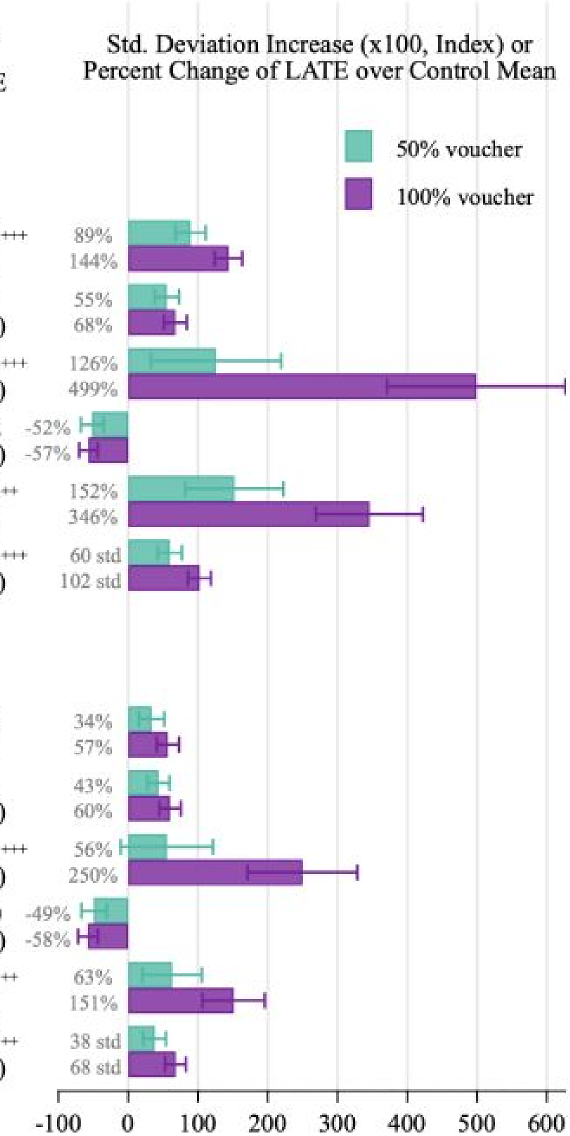
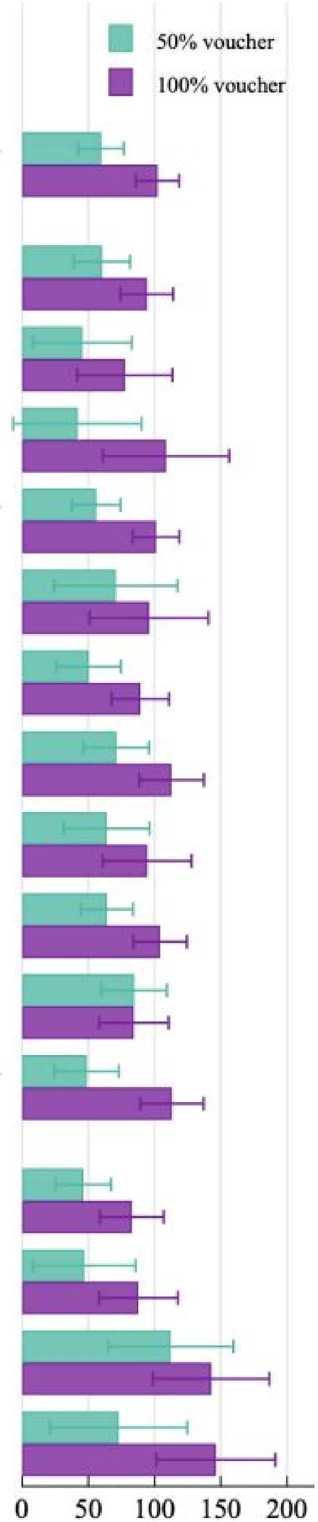


Figure 2.3: Treatment Effects of Receiving a Voucher on Contraceptive Efficacy

Notes: Panel A presents the estimated treatment effects using equation 1 for participants up to 100 days after enrollment when the voucher expired; panel B presents the estimated treatment effects for participants at two years after enrollment. +++, ++ and + indicate that the 100% effect is statistically different from the 50% effect at the 1, 5, and 10% levels, respectively.

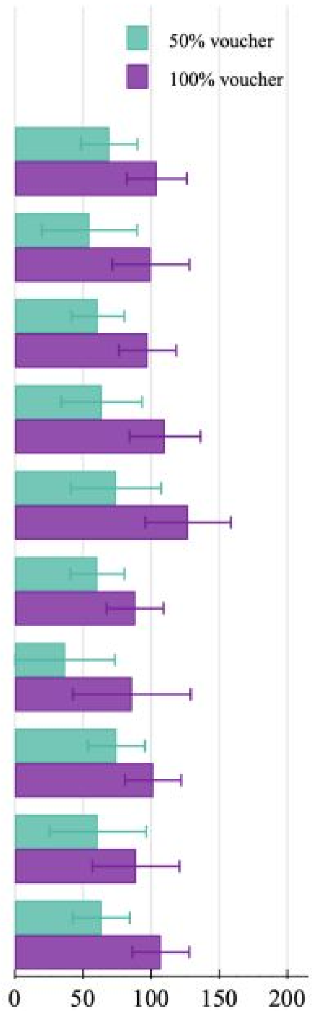
Figure 2.4: Heterogeneity in Treatment Effects of Receiving a Voucher on Contraceptive Efficacy



	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on index	0.00	0.72	0.39	0.60	0.00	0.66	0.65+++	1.02+++
		(0.03)	(0.07)	(0.09)		(0.02)	(0.06)	(0.08)
Pre-specified demographic groups								
Non-Hispanic White	0.03	0.73	0.42	0.60	0.07	0.69	0.65+	0.94++
		(0.03)	(0.09)	(0.11)		(0.03)	(0.08)	(0.10)
Non-Hispanic Black	-0.14	0.60	0.22	0.45	-0.19	0.49	0.32	0.78
		(0.09)	(0.15)	(0.19)		(0.07)	(0.15)	(0.18)
Hispanic any race	0.11	0.72	0.23	0.42	-0.01	0.63	0.80+	1.09+
		(0.08)	(0.20)	(0.25)		(0.07)	(0.20)	(0.24)
Women without children	0.04	0.74	0.38	0.56	0.00	0.66	0.65+++	1.01+++
		(0.03)	(0.07)	(0.09)		(0.02)	(0.07)	(0.09)
Mothers	-0.17	0.62	0.36	0.71	0.04	0.70	0.69	0.96
		(0.07)	(0.19)	(0.24)		(0.06)	(0.17)	(0.23)
Age<26	0.06	0.71	0.32	0.50	0.00	0.63	0.57++	0.89++
		(0.04)	(0.09)	(0.12)		(0.03)	(0.09)	(0.11)
Age≥26	-0.07	0.72	0.46	0.71	0.00	0.69	0.73+	1.13++
		(0.04)	(0.10)	(0.13)		(0.03)	(0.10)	(0.12)
Below associate's degree	-0.06	0.77	0.45	0.64	-0.05	0.63	0.53	0.94
		(0.06)	(0.15)	(0.16)		(0.05)	(0.13)	(0.17)
Associate's degree or higher	0.00	0.74	0.44	0.64	0.04	0.73	0.75+++	1.04++
		(0.03)	(0.08)	(0.10)		(0.03)	(0.08)	(0.10)
Married or cohabitating	-0.07	0.80	0.67	0.85	0.11	0.75	0.63	0.84
		(0.04)	(0.11)	(0.13)		(0.03)	(0.11)	(0.13)
Single	0.03	0.70	0.30	0.49	-0.03	0.69	0.75+++	1.13+++
		(0.04)	(0.09)	(0.12)		(0.03)	(0.10)	(0.12)
Fee scale								
101-150% FPL	0.06	0.75	0.32	0.46	0.12	0.66	0.53	0.83++
		(0.04)	(0.09)	(0.11)		(0.03)	(0.09)	(0.12)
151-200% FPL	0.12	0.71	0.28	0.47	0.01	0.69	0.59	0.88
		(0.05)	(0.15)	(0.20)		(0.04)	(0.12)	(0.15)
201-250% FPL	-0.13	0.76	0.76	1.12	-0.14	0.68	0.96	1.43
		(0.07)	(0.19)	(0.24)		(0.06)	(0.20)	(0.22)
250+% FPL	-0.31	0.59	0.40	0.73	-0.23	0.62	0.84+	1.46+
		(0.08)	(0.17)	(0.26)		(0.06)	(0.17)	(0.23)

(a) Overall effect on index and pre-specified demographic groups

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Have a usual place for birth control	-0.03	0.75	0.49	0.69	0.01	0.74	0.76++	1.04++
		(0.03)	(0.09)	(0.11)		(0.03)	(0.10)	(0.11)
Do not have a usual place for birth control	0.06	0.71	0.38	0.55	0.03	0.67	0.64	1.00+
		(0.05)	(0.13)	(0.18)		(0.04)	(0.11)	(0.14)
Using Tier 1 or 2 method	-0.06	0.74	0.43	0.61	0.01	0.70	0.68++	0.97++
		(0.03)	(0.08)	(0.10)		(0.03)	(0.09)	(0.11)
Not using Tier 1 or 2 method	0.08	0.70	0.35	0.64	-0.02	0.62	0.64+	1.10++
		(0.04)	(0.12)	(0.15)		(0.03)	(0.10)	(0.13)
Delayed getting birth control	0.19	0.90	0.67	0.74	-0.02	0.77	0.97	1.27++
		(0.04)	(0.16)	(0.17)		(0.04)	(0.14)	(0.16)
Did not delay getting birth control	-0.08	0.68	0.40	0.61	0.04	0.68	0.59	0.88+
		(0.03)	(0.08)	(0.10)		(0.03)	(0.08)	(0.11)
Positive desire to have a baby	-0.01	0.68	0.19	0.37	-0.06	0.56	0.42	0.86
		(0.06)	(0.14)	(0.19)		(0.05)	(0.14)	(0.22)
Negative desire to have a baby	-0.01	0.75	0.51	0.75	0.07	0.75	0.73+	1.01
		(0.03)	(0.09)	(0.11)		(0.02)	(0.09)	(0.11)
More likely to meet career aspirations	-0.04	0.80	0.45	0.61	0.03	0.66	0.56	0.89
		(0.05)	(0.15)	(0.18)		(0.04)	(0.13)	(0.16)
Less likely to meet career aspirations	0.01	0.73	0.43	0.63	0.02	0.71	0.76+++	1.07+++
		(0.03)	(0.09)	(0.11)		(0.03)	(0.09)	(0.11)



(b) Pre-specified pre-randomization categories

Notes: Standard errors are presented in parentheses below the coefficients. The figure on the right plots the standard deviation increase of LATE multiplied by 100 with the 95% confidence intervals. +++, ++ and + indicate that the 100% effect is statistically different from the 50% effect at the 1, 5, and 10% levels, respectively.

2.8 Tables

Table 2.1: Representation of M-CARES Participants and Balance in Characteristics

	(1) M-CARES participants	(2) NSFG 2018 Population	(3) Title X	(4) Voucher	(5) Control	(6) Significance of diff (p-values)
Observations	1,597	2,768	-	817	780	-
Birth control use						
Any birth control	0.79	0.90	0.79	0.79	0.79	0.99
Birth control pills	0.31	0.23	0.24	0.32	0.30	0.53
LARC (IUD, implant)	0.14	0.17	0.16	0.13	0.14	0.82
Withdrawal	0.02	0.06	0.02	0.01	0.02	0.36
Condoms	0.22	0.13	0.15	0.22	0.21	0.76
Other method	0.09	0.32	0.18	0.09	0.09	0.98
Age						
Age 18-19	0.10	0.12	0.10	0.10	0.10	0.76
Age 20-24	0.38	0.28	0.25	0.36	0.40	0.08
Age 25-29	0.31	0.29	0.21	0.32	0.30	0.53
Age 30-34	0.19	0.26	0.15	0.20	0.17	0.27
Age 35+	0.02	0.05	0.22	0.03	0.01	0.10
Race						
Non-Hispanic White	0.69	0.55	0.33	0.69	0.69	0.89
Non-Hispanic Black	0.11	0.16	0.19	0.10	0.13	0.11
Hispanic any race	0.11	0.21	0.34	0.11	0.11	0.94
Other	0.09	0.08	0.13	0.10	0.07	0.08
Marital status						
Single	0.51	0.26	-	0.50	0.51	0.49
Cohabiting	0.26	0.17	-	0.27	0.25	0.27
Married	0.08	0.56	-	0.08	0.07	0.25
Education						
Less than high school	0.02	0.07	-	0.02	0.02	0.90
High school degree	0.15	0.24	-	0.17	0.13	0.03
Some college	0.45	0.37	-	0.43	0.47	0.09
College degree or more	0.22	0.31	-	0.23	0.21	0.30
Previous childbearing						
0 births	0.85	0.63	-	0.85	0.85	0.87
1 birth	0.09	0.17	-	0.09	0.09	0.97
2 births	0.04	0.13	-	0.04	0.04	0.65
3+ births	0.01	0.08	-	0.01	0.01	0.43
Income as % of federal poverty line (FPL)						
Up to 100%	0.00	0.22	0.65	0.00	0.00	
101-150%	0.45	0.13	0.14	0.46	0.45	0.54
151-200%	0.27	0.11	0.07	0.27	0.27	0.74
201-250%	0.13	0.08	0.03	0.13	0.14	0.56
251%+	0.14	0.46	0.07	0.15	0.14	0.91

Notes: Column 1 presents the M-CARES sample, column 2 the population-weighted means from the 2015-17 NSFG, and column 3 the selected characteristics of the Title X population reported in the Family Planning Annual Report (Fowler et al. 2018). 1For M-CARES participants, birth control use refers to the month before enrollment and is asked during screening before enrollment. Age and fee scale are also derived from the screening survey. Race, marital status, education, and previous childbearing come from both survey and, when missing, PPMI data.

CHAPTER 3

The Gender Gap in Perceived Talent: Evidence from the Putnam Competition

Abstract: Top-talent labor and education markets are unique in the importance that skill ordering within top students has on the distribution of rewards (Stephan, 2012; Hill, 2020). How well talent ordering is observed by academic supervisors, in particular early in one's career, may have important consequences for future opportunities, but little is known about the information quality that academic supervisors have about their students. I digitize archival records on all participants of the Putnam Mathematical Competition over three decades to make progress on the following questions: (1) Are supervisor perceptions of talent different for women and men of highest ability? (2) Do any such differences affect the education and career trajectories of top talent women? A unique feature of the Putnam offers insight. Anyone from each college can participate individually, but each college's team of three is pre-selected by a supervisor who ranks students ahead of the competition in order of expected score. I find that ex-ante, for women and men with the same ex-post competition scores, supervisors expect women to do worse than their male peers. Women are less likely than men to have been pre-selected in the top three, even when they obtain a score that places them in the top three performers of their college. Female supervisors are no less biased than male supervisors. I find evidence of supervisor learning about individual women, but not about the group over time. I then link individuals to their later outcomes to evaluate long-run effects of talent misperception for women.

3.1 Introduction

Top-talent labor and education markets are unique: reward is frequently not based on the level of skill, but the *ordering* of skill within top students. How well talent ordering is observed, in particular in the early careers of highly talented students, may have important consequences on the opportunities and recognition they receive, and in later career output. Supervisors and their beliefs are important to talent discovery and promotion, and their effect on the career paths of top-talent youth is not well understood. This effect is important to understand given the outsize role that top talent has in innovation and economic growth.

This project assembles new confidential data on all participants of the Putnam Mathematical Competition over four decades and utilizes unique features of the competition to make progress on the following questions: (1) Are supervisor perceptions of talent different for women and men of highest ability? (2) Do any such differences affect the career outcomes and output of top talent youth?

To answer my research questions (the first in this version of the paper, and the second in a future version), I have obtained access records from the Putnam Competition, the preeminent math competition for undergraduates in North America, which sits thousands of students from hundreds of colleges each year for an olympiad-style problem-solving contest. The data contain individual and team-level information for all participating students and colleges, including names, declared gender, college team members, all scores and team supervisors.

I use the combined individual and team structure of the competition to measure supervisors' beliefs about student talent and test in the data the extent to which beliefs are heterogeneous along gender lines of both participants and supervisors. The following feature of the Putnam allows measurement of heterogeneity in beliefs: the competition is individual and anyone can participate, but there is a team component in which each college's Putnam supervisor pre-selects its representing team of three members before the competition. Colleges are then ranked by the performance of the pre-selected team. With few exceptions, colleges allow supervisors complete discretion in the selection of the team and encourage them to list the students who they believe to be of highest talent to represent the college. Infamously, supervisors are often wrong about who they pick. The first goal of the project is to test in the data whether supervisors are differentially wrong about the abilities of women relative to men and whether this heterogeneity varies by gender of the supervisor.

I find that ex-ante, for women and men with the same ex-post competition scores, supervisors expect women to do worse than their male peers. Women are less likely than men to have been pre-selected in the top three, even when they obtain a score that places them in the top three performers of their college. Female supervisors are no less biased than male supervisors. I find evidence of supervisor learning about individual women, but not about the group over time. I then link individuals to their later outcomes to evaluate long-run effects of talent misperception for women.

This paper contributes to the literature on gender biases of teachers and its importance on short and long-term outcomes. Several papers have developed ways to measure gender bias in an educational context. Lavy and Sand (2018) study the effect of primary school teachers' gender bias on later outcomes for boys and girls. They measure bias using differences in scores between exams graded by teachers and blind examiners in national exams to show that teachers who favor a gender have positive effects on that group's achievements. Carlana (2019) measures teachers' gender stereotypes in the Italian middle school context using the Gender Science Implicit Association Test and shows that teachers with stronger gender stereotypes exacerbate the gender gap in math performance and affect longer-term choices and education paths of girls. Other work uses self-reported measures of gender bias (Alan, Ertac, Mumcu, 2018). This paper contributes to this literature by (1) providing a new measure of gender bias using supervisor-predicted rankings relative to rankings achieved through competition; (2) studying gender bias in a higher-education context for the very right tail of talent distribution. This subset of the population is particularly important because bias in skill perception for this group may have implications for innovation and growth.

To this end, a goal of this project is to contribute to the literature studying talent as well as the institutions that drive talent discovery and innovation. Relative to a few existing studies using mathematical competitions as context (Ellison and Swanson (2016), Agarwal and Gaulé (2019), Buser and Yuan (2016), Moreira (2019)), in a future iteration, this project will be able to link competition participants with their career outcomes and provide causal evidence of any effect of early supervisor perception on career paths. In addition, this project would contribute to a strand of literature studying the science of innovators. Some existing evidence suggests that early setbacks of scientists that are based on pure chance may matter in their later careers (Hill and Stein (2020)). This project would advance this literature by focusing on the role of supervisor bias in the later careers of top math talent.

3.2 Context and Empirical Strategy

3.2.1 The Putnam Competition

The William Lowell Putnam Mathematical Competition, (henceforth the Putnam Competition), is a prestigious university-level mathematics competition held annually in the United States and Canada. Participants in the Putnam Competition are undergraduate students from a wide range of educational institutions, including both major research universities and smaller liberal arts colleges. Participation over the years has varied, but in the decade 1991-2000, between 2000 and 3000 students from over 600 institutions every year took the exam.

The competition itself comprises of two three-hour sessions on the same day, typically the first Saturday in December. Each session presents participants with six mathematical problems in algebra, combinatorics, geometry, and calculus. Each problem is worth 10 points for a maximum possible score of 120 points. It is not uncommon for the median score to be extremely low, often zero or one.

The competition has an individual component and a team component. Anyone that attends a US or Canadian university can participate, but a team from each college is pre-selected through a highly subjective process that is left, in most institutions, at the discretion of the Putnam supervisor for the college. The team is selected among the registered participants, and alternates are listed in order in the event that team members miss the competition. While there is variation in the preparation process for the Putnam, it is often rigorous, with many students participating in university-sponsored training sessions and problem-solving seminars that revolve around solving problems from previous competitions.

3.2.2 Strategy

It is hard to find contexts in which one can measure the early signals received by supervisors and mentors so as to compare them with actual performance and make progress on the questions of whether advisors and supervisors early in one's career perceive talent differently for men and women of the same ability. The Putnam competition is an exception.

The key to signal measurement is the fact that until 2019, the teams of three that represented each school were selected by the supervisor by varying methods. Some schools choose their teams on the basis of a pre-selection exams, while other schools allow more subjectivity in the selection of the team. It is precisely this subjectivity that allows the identification of the signal of talent received by the supervisor. Since the supervisor selects the team of three that will represent the college, it can be assumed that the supervisor believes each of the members of the team to be of possibly higher talent (or likelihood to perform well) than participants from the same school not selected to be on the team. The actual performance of members of the team compared to participants from that same college that were not on the team will be a measure of actual ability. While the signals received by the faculty supervisor may be noisy, it is important to understand if they are equally noisy for both men and women[1]. This forms the basis of the hypothesis that this paper in its current extent tests:

[1] I am aware from public data on Putnam team performances that the predictions that supervisors make are frequently incorrect. However, the key idea of the paper is to measure the gender difference in signals, rather than to assess the overall quality of signals.

that women and men, even at top levels of talent, are perceived as differentially talented even when of similar ability.

Where do supervisors acquire information about student skills? Supervisors are regular math faculty at their institution and may know participants through courses. Unfortunately, I have no information on the course content of students on the team and therefore the extent to which supervisors have interacted with participants in the past. In addition to regular classes, many institutions offer problem-solving classes or seminars which bring together Putnam participants and the supervisor for informal problem-solving sessions to prepare for the exam.[2]

Are students informed about supervisor perceptions before the exam? A central assumption that allows the interpretation of the difference in rankings as a difference between supervisor perception and skill is that the ranking that a supervisor gives a student does not impact their performance on the exam, in particular because the ranking is determined several weeks before the competition. The Pygmalion effect might bias measures of the difference between perception and performance because students chosen to be on the team may try harder and similarly students who were not chosen to be on the team may feel discouraged and not perform as well. While there is no way to rule out possible contamination from self-fulfilling prophecy effects, if supervisors tend to make systematically favorable predictions for one group, the effects should reduce the discrepancy between prediction and performance which may imply a lower bound bias.

There is anecdotal evidence that supervisors refrain from sharing information about the

[2]For example, the University of Minnesota has such a class: "Practice sessions for these competitions run weekly; students can come and go as is convenient." Source: https://www-users.cse.umn.edu/ tlawson/putnam/. Another example, Duke University, also has such a seminar: "There's no official syllabus for prepping for the Putnam. To get ready, the students practice working through problems and discussing their solutions in a weekly problem-solving seminar held each fall. Students serve as the instructors, focusing on a different topic each week ranging from calculus to number theory. 'They get a sense of what the problems are like, so it's not quite as intimidating as it might be if they went into the contest cold,' said math department chair Robert Bryant. 'Not only do they learn how to do the problems, but they also get to know each other,' said professor emeritus David Kraines, who has coached Duke Putnam participants for more than 30 years. Kraines said 8-10 students take his problem-solving seminar for credit each fall." Source: https://researchblog.duke.edu/2024/03/01/a-grueling-math-test-so-hard-almost-no-one-gets-a-perfect-score/.

team to foster competition and better preparation for the exam. For example, the supervisor for the University of Maryland uses the hope for team selection as an incentive for students to better prepare: "Selecting the team in advance creates a competitive atmosphere in the school for students to try and get into the school team."[3]

3.3 Data

The data come from Putnam registration forms for each college covering the period 1991-2000. Competition supervisors from each college must register their institution for competition in each year. To do so, they fill a form sent to them by the competition with basic information about the institution, the department that is hosting the competition, the full name of the competition supervisor for the institution, and the names of each of the participants from that college ranked in such a way that the top three ranked make up the pre-selected team for the college.

For the period covered in the analysis, the registration forms are stored in physical copies at the Archives of American Mathematics at the University of Texas at Austin.[4] I digitize all registration forms and extract from them the competition year, name of the college listed in each registration form, the name of the supervisor, names of each of the registrants and associated exam ID-s which are used for anonymous grading, and rank of each student listed in the registration form. While I have no measure of score for the students that were registered but did not take the exam, I record the names and rank of all students that were initially registered as their position in the registration forms is useful for two main reasons: first, to keep track of the number of registrants a supervisor had to rank in a given competition as that number may affect the ability of the supervisor to accurately rank students; and second, to test for alphabetic ranking. For each of the participants, I use

[3]Source: https://www.putnam.math.umd.edu/2019/09/04/putnam-rule-changes-and-what-they-mean/.

[4]Scans of registration forms were done in order of archival numbering, not in chronological order of competition year. The competition year 1992 was the last year of the initial analysis period that I digitized and was incomplete at the time of dissertation submission. Therefore the year 1992 is excluded from the analysis. A future iteration of the paper will include the recently-scanned registration forms from year 1992.

Genderize.io, an API that "predicts the gender of a person based on a first name" to assign a binary gender to each of the participants.[5] In addition, the registration forms contain the scores for each of the registered students who take the Putnam exam. I extract gender information for both participants and supervisors using their first name.

Table 3.1 presents summary statistics on the registrants and participants as well as institutions and supervisors for the competitions in the analysis period. A total of 39,843 students from 800 unique institutions across the United States and Canada were listed in registration forms by their supervisors and among them, 22,210 (56.3%) took the exam and were given a score. The vast majority of participants in the analysis period are male (75.5%) and 21.1% are female. The API was unable to categorize 3.4% of the names (756 observations) as being either male or female and these were excluded from the analysis. On average, colleges have about 6 participants per year, but they can have as man as 87 participants in a year and many often have as few as 1. Supervisors oversee the competition for an average of 2.5 year, but the median in the data is just 1. Less than half of the supervisors oversee competitions more than 1 year, but some supervisors serve for the entire analysis period. This set of supervisors will be important to understand the extent to which perceptions about talent change over time.

The average score achieved in competition in the analysis period is 9.9, whereas the median is 2. A simple comparison shows that on average, male participants score substantially higher than females (11 points for males and 5 points for females) and the median score is different as well (9 for males and 0 for females). The standard deviation of scores for males is also larger than that of females (14.6 and 9.6 respectively).

[5] Refer to https://genderize.io/documentation for more information. I acknowledge a limitation of this paper: that the measure of gender does not capture the fact that any participant may identify with a different gender than the one predicted using only information about their first name. For the purpose of this paper, "male" is the term I use to refer to participants whose first names are more likely to be male names and "female" is the term I use to describe participants whose first names are more likely to be female names.

3.4 Results

I establish four main sets of results. First, I show that there are significant differences between men and women in the likelihood of being ex ante predicted to be a top performer conditional on performance. Second, in a heterogeneity analysis, I show that these gender gaps in likelihood to correctly predict a top performer hold for both male and female supervisors. Third, I show that these differences are primarily driven by first-time participants. Among women and men for whom a participation is not their first, top performing women are just as likely as men to be predicted to be a top performer. I interpret this result as a supervisor's ability to learn about individual talent from previous participation. However, in a fourth set of results, I show that over time supervisors do not change the rate at which they underpredict top-performing women being at the top *for first-time participants.*

1. Prediction of top performers–The first choice variable that will give insight into whether supervisors have differential information about the skills of men and women is that of predicting whether a student will be in the top three of their college. Therefore, the specification that I use compares the rate at which men and women are ranked in the top three by their supervisors before the competition. The estimating equation is:

$$\begin{aligned} y_{isy} = \beta_0 + \beta_1(Female_i) + \beta_2(ExPostTopThree_{isy}) \\ + \beta_3(Female_i \times ExPostTopThree_{isy}) + \theta_s + \theta_y + \varepsilon_{isy} \end{aligned} \tag{3.1}$$

where y_{isy} is the probability of student i being predicted to be among the top three performers in one's school s in year y of competition, $Female_i$ is an indicator for whether the name of the individual corresponds to a female name, $ExPostTopThree_{isy}$ is an indicator for whether student i was in the top three performers of the school. The specification includes college fixed effects in order to absorb variation in the likelihood of being ranked in the top-3 that may come from schools tending to consistently have more or fewer participants. Results from the preferred specification are shown in column (3) of Table 3.2. Women are 9.7pp less likely to be predicted to be among the top-3 conditional on ex-post achieving a score that puts them in the top-3. Women are also less likely to be beneficiaries of a false positive, a mistaken prediction that they will earn a top-3 score when they do not. Alternative specifications show the same qualitative results and similar magnitudes. Column (1) of Table 3.2 shows that a simple linear regression with no fixed effects shows a difference of 6.7pp in the probability of deserving men and women being predicted to be at the top. This difference reflects a greater-than 13% difference in likelihood of deserving women's chances to be perceived as being at the top relative to deserving men's chances of being predicted to be at the top. Column (3) additionally controls for score. Holding score constant, men and women are differentially likely to be predicted as top-3 performers by 3pp on average if they don't end up being at the top, and an additional 3pp if they end up being at the top. Columns (4) and (5) further restrict the sample to colleges with fewer than 25 participants, and those where at least one student earns more than 0 points respectively. Column (4) aims to check the robustness of the results for colleges in which supervisors are more likely to have better information about students since there are fewer participants. Column (5) confirms that results hold for the subset of colleges where not all students got a 0.

Further, as shown in Table 3.3 female students who are the very top performer in their school are 14.6pp less likely to have been predicted to be the top performer than male students (column 3 of Table 3.3). The same specifications as in the previous table result in differences between 12pp and and 15pp for top male and female students to be listed at the top of the ranking by their supervisors. Of particular interest is specification (4), which includes score controls. Holding scores constant, women not at the top of their college have a slight but statistically significant difference in the likelihood of being ranked at the top (1.3pp). This difference increases dramatically for women at the top (a 10.3pp increase) for women and men with the same scores. Both the results in Table 3.2 and Table 3.3 provide evidence that supervisors do not have perfect information about the skills of students. In particular they are likely to underestimate top women and slightly more likely to overestimate men not at the top. This is consistent with beliefs that are in line with score distributions of the two groups: men have higher scores on average and higher variance in scores.

2. Are female supervisors better at correctly predicting top women performance? To investigate whether the gender gap might be due to the noisiness of the signals sent by the students or due to preferences/animus, I conduct a test of heterogeneity in the gender gap in prediction of excellence. I separately run the specification equation (3.1) for female and male supervisors and show the results in column (3) of Table 3.4, along with alternative specifications as in the earlier exercises. In all specifications, female supervisors are more likely than male supervisors to mistakenly rank men among the top-3 who do not achieve a score that places them in that category. The gender gap in likelihood to be overestimated has a similar magnitude for female supervisors as for male supervisors, but is not statistically significant. In column (1) of Table 3.4, the specification that compares rates of top-3 rankings by gender without college or year fixed effects shows no differential rate in correct prediction of top-3 standings for female vs. male students when the supervisor is a woman. The magnitude is half the gender gap for the male-supervised subsample and not statistically different from 0. For the preferred specification, women supervisors are overall less good at predicting the team, but the gender gap in the correctness of their predictions is also smaller. While the sample of colleges that are supervised by women is much smaller and estimates more noisy, the evidence is suggestive that the gender of supervisor matters for talent perceptions of students of different genders.

3. Do supervisors learn about individuals? Figure 3.1 provides evidence that while information is imperfect, supervisors are able to learn about individuals. This analysis repeats the specification above, but limits the sample to students who will participate numerous times. On the left panel, I show that among those that were of team-level performance, women are about 10pp less likely to have been predicted as such. This is similar to the differences observed in the full sample, although supervisors are more likely to predict correctly team-level performance even on first participation for both men and women among those that take the exam more than once. However, at second or later participation, the likelihood of correct prediction, conditional on top performance is over 77% for both groups and the difference is not statistically significant. I interpret these results as providing evidence that supervisors have better information about repeat students than those they see in competition for the first time. Supervisors show learning about students ability.

4. The gender gap in prediction does not change over time for first-time participants

While supervisors learn about individual talent and as shown above improve the quality of their prediction about the performance of individual female students once they have observed a student once, it is not clear whether they improve the quality of their predictions over time for the whole group. To test whether the quality of information changes over time, I specify the following equation

$$\begin{aligned} y_{isy} &= \beta_0 + \beta_1 \ ExPostTopThree_{isy} + \beta_2 \ Female_i + \beta_3 \ (Female_i \times ExPostTopThree_{isy}) \\ &+ \beta_4 (Female_i \times YearSup) + \beta_5 (ExPostTopThree_{isy} \times YearSup) \\ &+ \beta_6`(Female_i \times ExPostTopThree_{isy} \times YearSup) + \delta_s + \varepsilon_{ist} \end{aligned} \tag{3.2}$$

with the coefficients β_4, and β_6 being of interest. Each of those coefficients is the evolution of the gender gap in prediction of performance. The sample is restricted to all first-time participants. $YearSup$ is the number of years in the sample that a supervisor has overseen the competition for their school. In this specification, I include college fixed effects (δ_s) and the results are shown in column (1) of Table 3.5. Alternatively, in order to capture the supervisor-specific slope I include supervisor fixed effects instead. Those results are shown in column (2) of Table 3.5. The gender gap in prediction among top-3 students is 7.6pp and 6.8pp for the two specifications respectively and the slope over the tenure of the supervisor is 0.7pp and 0.5pp respectively, and not statistically significant in either case.

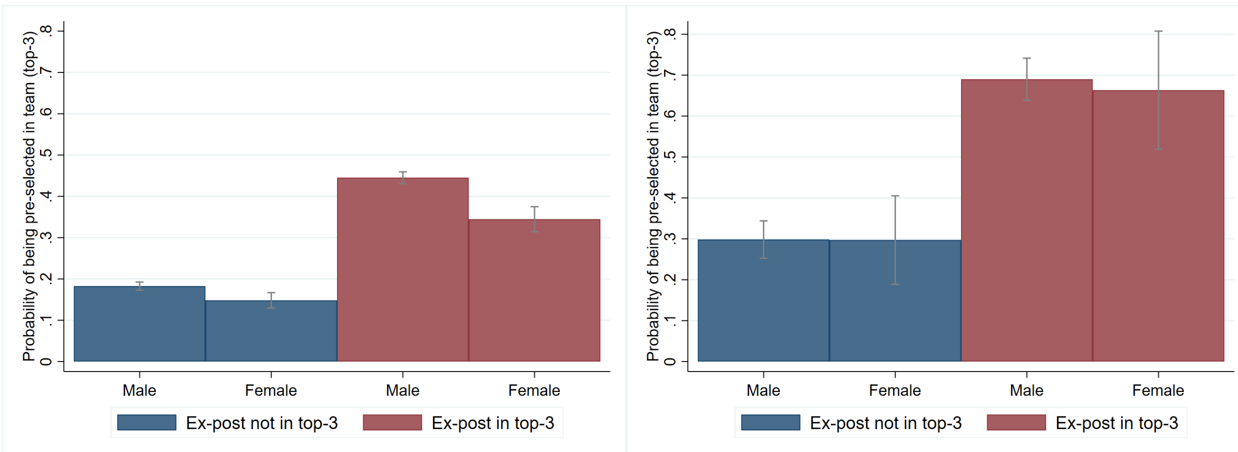
The results for both specifications show no significant slope in the gender gap in predictions for first-time participants. Even though the period covered is short 1991-1999 (and it excludes 1992), the results indicate that while supervisors learn about individuals and gender gaps in expected performance close in second participation, there is no change over time in the gender gap in expected performance for first-time participants.

3.5 Conclusion

I create a novel dataset from the Putnam Mathematical Competition to make progress on the following questions: (1) Are supervisor perceptions of talent different for women and men of highest ability? (2) Do any such differences affect the education and career trajectories of top talent women? This current version of the paper only answers the first. I aim to answer the second in continued work. A unique feature of the Putnam allows an analysis to make progress on the first question. Utilizing it, I find that ex-ante, for women and men with the same ex-post competition scores, supervisors expect women to do worse than their male peers. Women are less likely than men to have been pre-selected in the top three, even when they obtain a score that places them in the top three performers of their college. Female supervisors are no less biased than male supervisors. I find evidence of supervisor learning about individual women, but not about the group over time. As I continue this project I aim to link individuals to their later outcomes to evaluate long-run effects of talent misperception for women.

3.6 Figures

Figure 3.1: Comparison of predictions for first and >1 time participants



Note: Likelihood of correctly predicting tier of performance for the subset of students who participate more than once. The left panel shows, for this subset, the likelihood of predicting top-3 performance in the first time a student takes the exam, and the right panel shows, for this subset, the likelihood of a supervisor predicting top-3 performance in the 2nd-4th time a student takes the exam.

3.7 Tables

Table 3.1: Summary statistics on Putnam data

	Number	(fraction)
Years of competition	9	
Unique institutions	800	
Registered students	39,483	
Participants (fraction of registered)	22,210	(0.563)
Male participants (fraction of particip.)	16,758	(0.755)
Female participants (fraction of particip.)	4,691	(0.211)
No gender assign (fraction of particip.)	756	(0.034)

	Average	Min	Max	Median	St. dev.
Scores (full sample)	9.882	0	102	2	13.933
Scores (males)	11.097	0	102	9	14.569
Scores (females)	5.126	0	90	0	9.599
Particip. by college (full sample)	6.055	1	87	5	6.321
Particip. by college (males)	4.768	1	70	4	5.203
Particip. by college (females)	2.071	1	17	2	1.552
Supervisor years observed	2.473	1	9	1	2.149

Table 3.2: Prediction that a student will perform in the top three

	(1)	(2)	(3)	(4)	(5)	(6)
Female	0.001	-0.032***	-0.032***	-0.028***	-0.034***	-0.034***
	(0.010)	(0.010)	(0.010)	(0.010)	(0.011)	(0.010)
In top-3 ex-post	0.341***	0.275***	0.274***	0.157***	0.269***	0.292***
	(0.014)	(0.008)	(0.008)	(0.009)	(0.008)	(0.008)
Female $\times$ in top-3 ex-post	-0.067***	-0.065***	-0.065***	-0.029*	-0.061***	-0.058***
	(0.020)	(0.017)	(0.017)	(0.016)	(0.018)	(0.017)
Score				0.008***		
				(0.000)		
Constant	0.186***	0.218***	0.218***	0.172***	0.239***	0.207***
	(0.013)	(0.005)	(0.005)	(0.005)	(0.005)	(0.005)
Adjusted R^2	0.119	0.144	0.144	0.176	0.113	0.155
Observations	18,541	18,541	18,541	18,541	16,669	17,514
College-years with >3 participants	✓	✓	✓	✓	✓	✓
College FEs		✓	✓	✓	✓	✓
Year FEs			✓	✓	✓	✓
College-years with <25 participants					✓	
Maximum score of college is ¿0						✓

Notes:. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.3: Prediction that a student will be the top-scorer

	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.009*	-0.021***	-0.021***	-0.013**	-0.022***	-0.021***
	(0.005)	(0.006)	(0.006)	(0.006)	(0.007)	(0.006)
rf_expost	0.267***	0.249***	0.249***	0.170***	0.250***	0.267***
	(0.013)	(0.008)	(0.008)	(0.008)	(0.008)	(0.008)
femrf_expost	-0.123***	-0.125***	-0.125***	-0.103***	-0.130***	-0.112***
	(0.027)	(0.018)	(0.018)	(0.018)	(0.019)	(0.020)
Score				0.005***		
				(0.000)		
Constant	0.081***	0.086***	0.086***	0.046***	0.093***	0.082***
	(0.004)	(0.003)	(0.003)	(0.003)	(0.003)	(0.003)
Adjusted R^2	0.072	0.065	0.065	0.088	0.055	0.076
Observations	18,541	18,541	18,541	18,541	16,669	17,514
College-years with >3 participants	✓	✓	✓	✓	✓	✓
College FEs		✓	✓	✓	✓	✓
Year FEs			✓	✓	✓	✓
College-years with <25 participants					✓	
Maximum score of college is >0						✓

Notes:. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.4: Prediction accuracy by gender of supervisor

A. Subsample with female supervisor	(1)	(2)	(3)	(4)	(5)	(6)
Female	-0.016	-0.043	-0.045	-0.039	-0.045	-0.028
	(0.028)	(0.032)	(0.032)	(0.031)	(0.032)	(0.033)
In top-3 ex-post	0.258***	0.195***	0.192***	0.088***	0.190***	0.214***
	(0.038)	(0.027)	(0.027)	(0.030)	(0.027)	(0.027)
Female × in top-3 ex-post	-0.015	-0.033	-0.030	-0.003	-0.028	-0.037
	(0.051)	(0.050)	(0.050)	(0.050)	(0.051)	(0.054)
Score				0.010***		
				(0.001)		
Constant	0.241***	0.275***	0.276***	0.241***	0.281***	0.257***
	(0.028)	(0.017)	(0.017)	(0.017)	(0.017)	(0.017)
Adjusted R^2	0.068	0.074	0.073	0.101	0.064	0.076
Observations	1,918	1,914	1,914	1,914	1,876	1,756
B. Subsample with male supervisor						
Female	-0.003	-0.036***	-0.036***	-0.033***	-0.039***	-0.038***
	(0.011)	(0.010)	(0.010)	(0.010)	(0.012)	(0.011)
In top-3 ex-post	0.355***	0.287***	0.286***	0.171***	0.281***	0.306***
	(0.015)	(0.008)	(0.008)	(0.009)	(0.009)	(0.008)
Female × in top-3 ex-post	-0.066***	-0.061***	-0.061***	-0.025	-0.057***	-0.055***
	(0.020)	(0.018)	(0.018)	(0.018)	(0.019)	(0.019)
Score				0.008***		
				(0.000)		
Constant	0.178***	0.210***	0.210***	0.164***	0.233***	0.199***
	(0.013)	(0.005)	(0.005)	(0.005)	(0.006)	(0.005)
Adjusted R^2	0.131	0.154	0.155	0.185	0.120	0.168
Observations	16,071	16,071	16,071	16,071	14,238	15,227
College-years with >3 participants	✓	✓	✓	✓	✓	✓
College FEs		✓	✓	✓	✓	✓
Year FEs			✓	✓	✓	✓
College-years with <25 participants					✓	
Maximum score of college is >0						✓

Notes:. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table 3.5: Prediction that a student will be the top-scorer for first-time participants

	(1)	(2)
female	-0.006	-0.002
	(0.017)	(0.018)
team_expost	0.259***	0.232***
	(0.014)	(0.015)
femteam_expost	-0.076**	-0.068**
	(0.030)	(0.031)
yearrunningsup	-0.013***	-0.016***
	(0.004)	(0.004)
femyear	-0.006	-0.007
	(0.006)	(0.006)
expostyear	-0.012**	-0.007
	(0.005)	(0.006)
femexpostyear	0.007	0.005
	(0.012)	(0.012)
Adjusted R^2	0.144	0.139
Observations	13,680	13,524
College-years with >3 participants	✓	✓
College FEs	✓	
College-supervisor FE		✓

Notes:. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

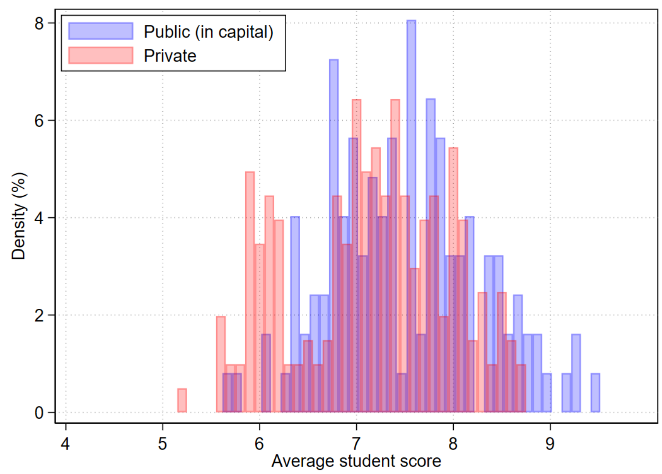
APPENDIX A

Appendix to Inequity in Centralized College Admissions with Public and Private Universities: Evidence from Albania

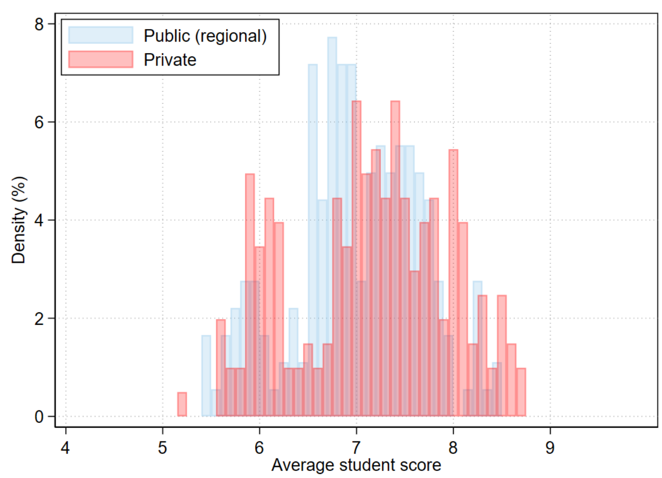
A.1 Additional Tables and Figures

Figure A.1: Quality of students enrolled in programs by public status and geography

(a) Public programs in capital and all private programs

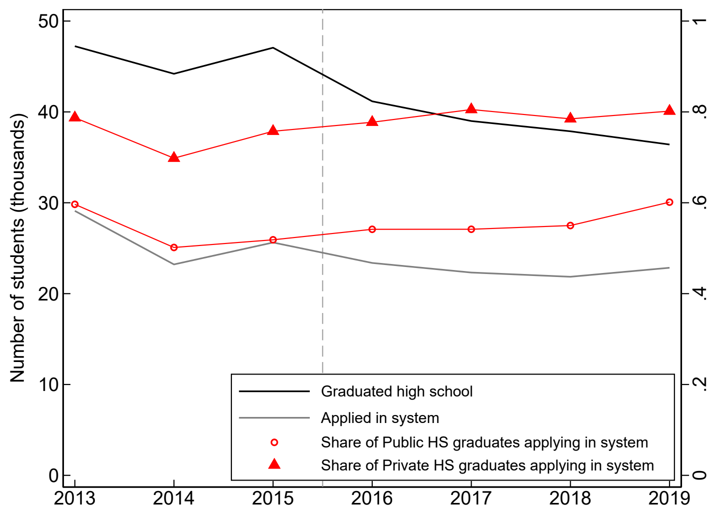


(b) Regional public programs and all private programs



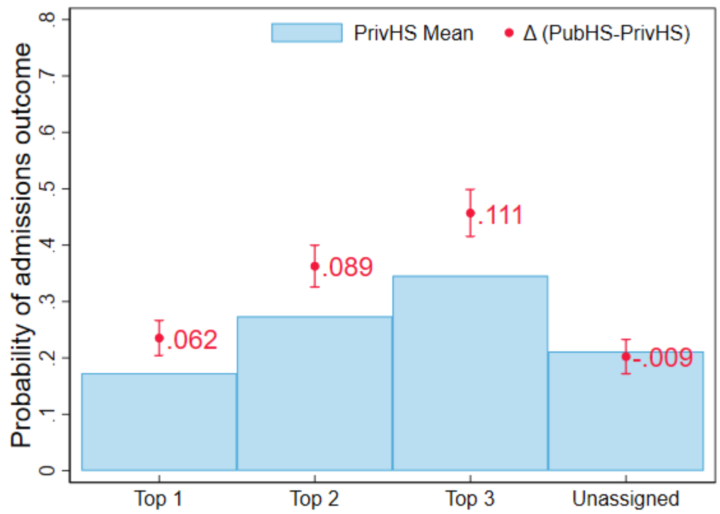
Note: This chart displays the distribution of the simple average of the weighted average score for enrollees in public university programs (blue) and private university programs (pink) in 2019. The top panel compares public programs in the capital to all private programs. 25 of 26 private universities are located in the capital and 7 out of 12 public universities are located in the capital. The bottom panel compares the average score of enrollees in private programs with enrollees in regional public universities.

Figure A.2: High School Graduates and College Applicants over Time



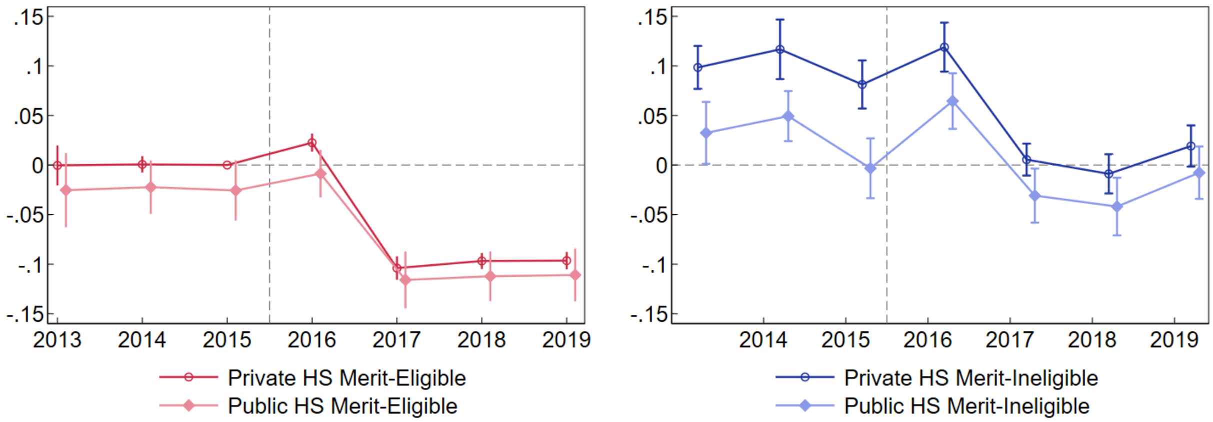
Note: Chart shows trends in number of students graduating high school and those applying to college through the centralized system. In shades of gray are the total number of students who graduated high school and the number that applied through the platform; in red, the share of graduating public HS students and private HS students separately that applied through the centralized platform.

Figure A.3: Final assignment outcomes in capital



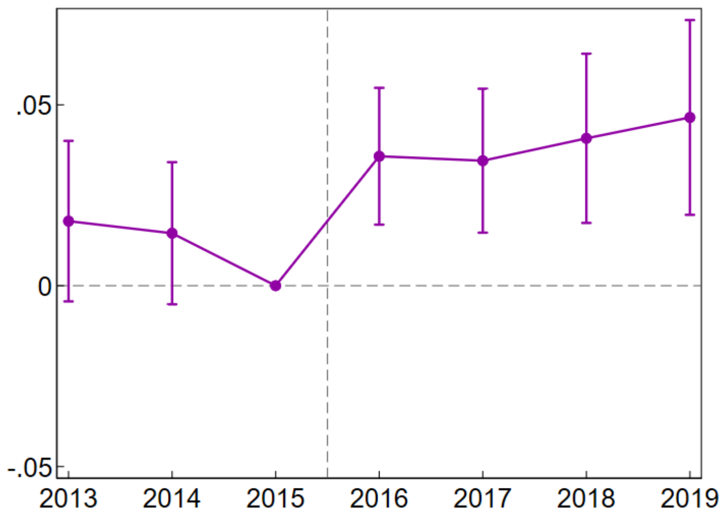
Note: Differences are conditional on score and district FE. Sample includes years 2013-2015. 95% confidence intervals are shown with standard errors clustered at the high school level. Score is the weighted average of test scores in Math, Language, choice subjects, and HS GPA.

Figure A.4: Selectivity of Most Selective Public Program



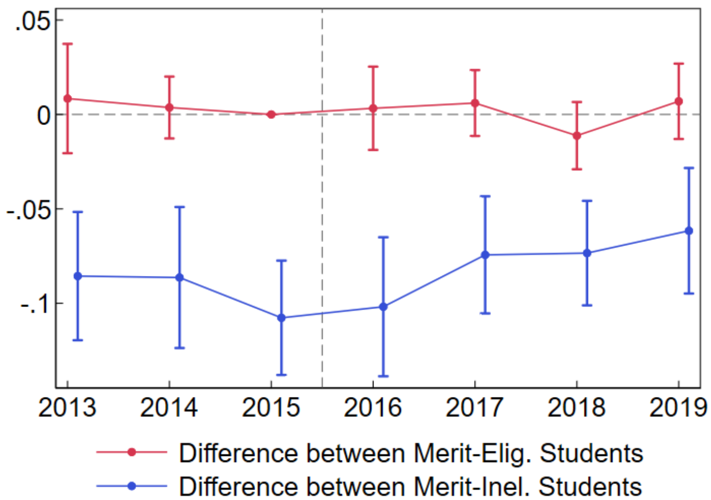
Note: Charts show selectivity of most selective public program for each of private high school top students, public high school top students, private high school non-top students, public high school non-top students. Top students are the set of students that would qualify for merit scholarships at private institutions based on their exam scores.

Figure A.5: Event Study of Triple Differences in Selectivity of Most Selective Public Programs
Top and Non-Top Students



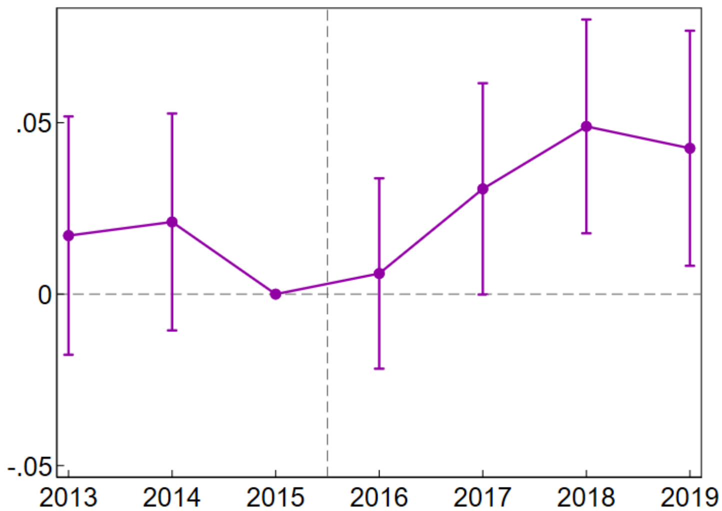
Note: This chart plots the differences in selectivity of most selective public programs on the centralized application between private high school and public high school students relative to the difference in the year before the reform (2015). Higher values reflect a reduction of the gap between the top choices of high- vs. lower-SES students. Regressions control for average exam score and include district FE. Standard errors clustered at district level.

Figure A.6: **Outcome**: average historical selectivity of programs in application
Outcome measure: program's standardized cutoff score in 2013



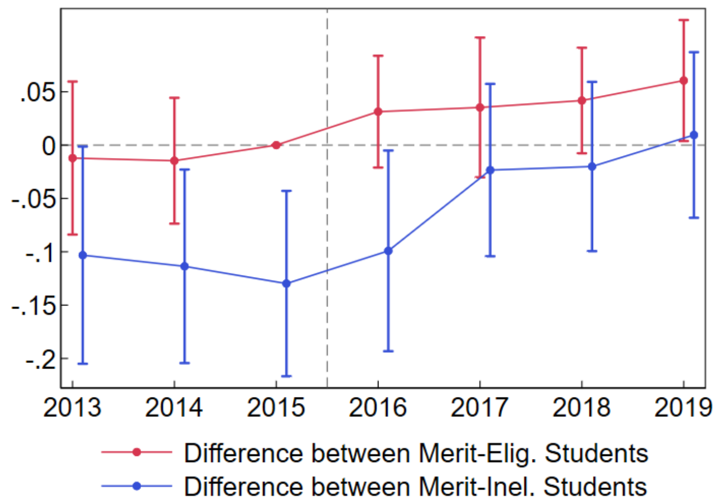
Note: This chart plots the differences in average selectivity of public programs included in the portfolio on the centralized application between private high school and public high school students. Negative differences reflect less selective portfolio choices for public high school students. Regressions control for average exam score and include district FE.

Figure A.7: Double differences: portfolio selectivity



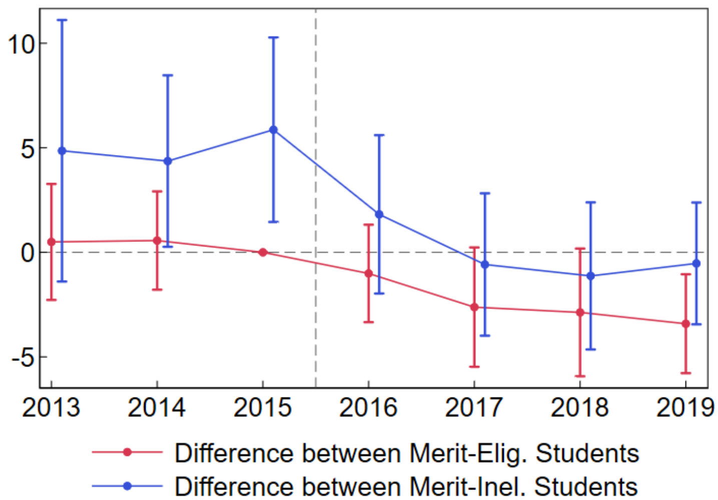
Note: This chart plots the differences in average selectivity of the public portion of the programs included in the portfolio on the centralized application between private high school and public high school students relative to the difference in the year before the reform (2015). Higher values reflect a reduction of the gap between the choices of high- vs. lower-SES students. Regressions control for average exam score and include district FE. Standard errors clustered at district level.

Figure A.8: First differences: most selective program
Outcome measure: standardized previous year's cutoff



Note: This chart plots the differences in most selected public programs included in the portfolio on the centralized application between private high school and public high school students. Negative differences reflect less selective portfolio choices for public high school students. The alternative measure of selectivity used in this graph is the standardized previous year's cutoff for each program. Regressions control for average exam score and include district FE.

Figure A.9: First differences: most selective program
Outcome measure: previous year's rank



Note: This chart plots the differences in most selected public programs included in the portfolio on the centralized application between private high school and public high school students. Negative differences reflect less selective portfolio choices for public high school students. The alternative measure of selectivity used in this graph is the previous year's rank for each program. Regressions control for average exam score and include district FE.

Table A.1: Relationship Between Attending a Public HS and Assignment Outcomes
Robustness Check for Initial Assignment Outcome

	National Outcomes			Outcomes in Capital		
	Assigned to First Choice	Assigned to One of Top Three Choices	Unassigned	Assigned to First Choice	Assigned to One of Top Three Choices	Unassigned
Public HS	0.032***	0.038***	-0.017	0.022	0.029*	-0.006
	(0.011)	(0.012)	(0.010)	(0.017)	(0.016)	(0.020)
Private HS Mean	0.298	0.540	0.188	0.192	0.404	0.288
Adjusted R^2	0.125	0.185	0.263	0.128	0.248	0.314
Observations	84,931	84,931	84,931	17,336	17,336	17,336
Condition on Score	✓	✓	✓	✓	✓	✓
District FEs	✓	✓	✓			
Year FEs	✓	✓	✓	✓	✓	✓

Notes: Sample includes years 2013-2015. Private HS Mean is the unconditional mean of the outcome variable for students attending private high schools. Standard errors are clustered at the district level for the national sample and are robust for the capital-only sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.2: Relationship Between Attending a Public HS and Assignment Outcomes
Robustness Check for Final Assignment Outcome of Those Who Did not Reject Centralized Offer

	National Outcomes			Outcomes in Capital		
	Assigned to First Choice	Assigned to One of Top Three Choices	Unassigned	Assigned to First Choice	Assigned to One of Top Three Choices	Unassigned
Public HS	0.028**	0.042***	-0.030***	0.026	0.043**	-0.029
	(0.012)	(0.012)	(0.010)	(0.021)	(0.021)	(0.019)
Private HS Mean	0.340	0.588	0.159	0.218	0.436	0.268
Adjusted R^2	0.124	0.181	0.249	0.142	0.255	0.307
Observations	76,556	76,556	76,556	15,498	15,498	15,498
Condition on Score	✓	✓	✓	✓	✓	✓
District FEs	✓	✓	✓			
Year FEs	✓	✓	✓	✓	✓	✓

Notes: Sample includes years 2013-2015. Private HS Mean is the unconditional mean of the outcome variable for students attending private high schools. Standard errors are clustered at the district level for the national sample and are robust for the capital-only sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.3: Relationship Between Attending a Public HS Application Selectivity Measures

	National Sample					Applicants from Capital				
	Listed Med.	Top Ranked	Second Ranked	Third Ranked	App. Sel.	Listed Med.	Top Ranked	Second Ranked	Third Ranked	App. Sel.
Public HS	-0.013***	-0.118***	-0.121***	-0.106***	-0.092***	-0.026***	-0.128***	-0.117***	-0.065***	-0.061***
	(0.005)	(0.023)	(0.021)	(0.024)	(0.018)	(0.006)	(0.019)	(0.020)	(0.019)	(0.009)
Score	0.094***	0.555***	0.571***	0.525***	0.500***	0.099***	0.460***	0.499***	0.454***	0.430***
	(0.002)	(0.024)	(0.020)	(0.020)	(0.019)	(0.002)	(0.006)	(0.006)	(0.006)	(0.003)
Priv. HS Mean	0.141	6.782	6.742	6.661	6.587	0.139	6.985	6.967	6.875	6.792
Adjusted R^2	0.170	0.371	0.377	0.362	0.631	0.170	0.301	0.314	0.299	0.614
Obs.	84,931	84,563	84,073	83,244	84,909	17,336	17,307	17,145	16,985	17,336
District FEs	✓	✓	✓	✓	✓					
Year FEs	✓	✓	✓	✓	✓	✓	✓	✓	✓	✓

Notes: Sample includes years 2013-2015. Private HS Mean is the unconditional mean of the outcome variable for students attending private high schools. Standard errors are clustered at the district level for the national sample and are heteroskedasticity robust for the capital-only sample. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.4: Relationship Between High School GPA and Results in the Matura Exams by Type of High School

	Exam mean	Math score	Literature score	First elective score	Second elective score
hs_gpa	0.707***	0.569***	0.911***	0.652***	0.694***
	(0.021)	(0.018)	(0.030)	(0.024)	(0.022)
Public HS	0.932***	0.325*	1.416***	0.815***	1.175***
	(0.178)	(0.185)	(0.248)	(0.212)	(0.194)
gpa_pub	-0.095***	-0.009	-0.161***	-0.086***	-0.123***
	(0.023)	(0.019)	(0.034)	(0.026)	(0.025)
Adjusted R^2	0.621	0.524	0.518	0.359	0.363
Obs.	84929.000	84901.000	84896.000	84875.000	84842.000
Year FEs	✓	✓	✓	✓	✓

Notes: Sample includes years 2013-2015. * $p < 0.10$, ** $p < 0.05$, *** $p < 0.01$.

Table A.5: Tests of the Parallel Trends Assumption on Selectivity of "Reach" Schools

	(1)	(2)	(3)
Public HS × Non-Top × Year	-0.030	-0.009	0.012
	(0.021)	(0.022)	(0.029)
Public HS × Non-top	-0.060	-0.106	-0.145*
	(0.084)	(0.079)	(0.079)
Public HS × Year	0.014	-0.004	-0.023
	(0.022)	(0.017)	(0.015)
Non-Top × Year	-0.075***	-0.071***	-0.086***
	(0.022)	(0.021)	(0.026)
Public HS	-0.030	0.015	-0.052
	(0.047)	(0.034)	(0.071)
Non-top	0.324***	0.346***	0.382***
	(0.092)	(0.079)	(0.069)
Year	0.009	0.006	0.015
	(0.024)	(0.020)	(0.012)
Adjusted R^2	0.352	0.405	0.474
Observations	50,029	50,029	50,029
Score Controls	Yes	Yes	Yes
Score and HS Path Controls	No	Yes	Yes
District FEs	No	No	Yes

Notes: Regression are run on data for the application cycles in years 2013-2015, immediately before the 2016 reform. The three years of data are coded Year 1 through 4. The three specifications test for parallel trends in the double difference across performance groups and types of high school. Standard errors are clustered at the locality level and are shown in parentheses.

Table A.6: Triple Difference Estimate of Exposure to Contraction of Outside Options on Selectivity of "Reach" Programs Chosen

	Including All Years			Excluding 2016		
	(1)	(2)	(3)	(4)	(5)	(6)
Public HS × Non-top × Post-reform	0.115***	0.132***	0.109**	0.102**	0.119***	0.095**
	(0.039)	(0.036)	(0.043)	(0.041)	(0.038)	(0.045)
Public HS × Non-top	-0.143***	-0.153***	-0.155***	-0.137***	-0.150***	-0.151***
	(0.049)	(0.047)	(0.042)	(0.050)	(0.048)	(0.043)
Public HS × Post-reform	0.007	-0.014	0.025	0.008	-0.015	0.032
	(0.030)	(0.029)	(0.027)	(0.030)	(0.030)	(0.026)
Non-top × Post-reform	0.230***	0.196***	0.212***	0.190***	0.144***	0.164***
	(0.037)	(0.034)	(0.039)	(0.037)	(0.036)	(0.043)
Public HS	0.019	0.035	-0.055	0.020	0.040	-0.054
	(0.029)	(0.026)	(0.042)	(0.029)	(0.025)	(0.043)
Non-top	-0.093*	-0.006	0.004	-0.028	0.084**	0.092**
	(0.049)	(0.040)	(0.039)	(0.046)	(0.038)	(0.038)
Post-reform	-0.458***	-0.380***	-0.410***	-0.367***	-0.264***	-0.303***
	(0.029)	(0.029)	(0.019)	(0.032)	(0.032)	(0.019)
Adjusted R^2	0.279	0.332	0.379	0.290	0.354	0.400
Observations	131,926	131,926	131,926	111,900	111,900	111,900
Score Controls	Yes	Yes	Yes	Yes	Yes	Yes
Score and HS Path Controls	No	Yes	Yes	No	Yes	Yes
District FEs	No	No	Yes	No	No	Yes

Notes: Scores represent the weighted average of end-of-high-school exam scores. HS Path is a binary variable that represents whether the path chosen in the second year of high school is "scientific" or "social", which affects the weights programs give to the elective exams. Standard errors are clustered at the locality level and are shown in parentheses.

Table A.7: Parameter estimates

		Estimate	SE
a. Preference parameters			
γ	cutoff	0.320	0.021
γ	capital	0.706	0.067
γ	public	-2.270	0.060
γ	private	-1.790	0.003
γ	price	-0.001	0.000
γ	distance	-0.011	0.000
λ	hises x public	-0.168	0.064
λ	hises x private	-3.421	0.063
λ	hises x capital	0.157	0.111
λ	hises x price	0.001	0.000
λ	hises x cutoff	-0.082	0.007
λ	hises x dist	0.005	0.001
γ	applied science	-0.217	0.047
γ	health	1.009	0.048
γ	social science and humanities	-0.891	0.063
b. Belief parameters			
σ_0	Standard deviation intercept	6.040	0.074
σ_1	Standard deviation slope on cutoff	-0.497	0.012
$\mu_{0,lowses}$	Mean shift intercept (lower-SES)	-12.004	0.104
$\mu_{0,highses}$	Mean shift intercept (high-SES)	-3.368	0.096
$\mu_{1,lowses}$	Mean shift slope on cutoff (lower-SES)	1.276	0.033
$\mu_{1,highses}$	Mean shift slope on cutoff (high-SES)	0.314	0.082

Notes: The table shows estimated parameters from the model.

A.2 Description of pre and post-reform admissions procedures

A.2.1 Pre-reform mechanism: deferred acceptance

In the pre-reform period, the Center for Educational Services conducted admissions to all public programs through a DA algorithm. The stages of the application process are as follows:

1. In June of each year, students take national exams, two obligatory exams in math and literature, and two elective exams in subjects chosen by each student.

2. In July, grades from each of the exams become public through a set of lists in which students are ranked in each of the exams from best to worst performing.

3. In August, the admissions process for the public programs begins, with each student applying to up to ten programs through the centralized platform, and ranking programs in the order of most to least preferred.

4. Round 1, phase 1: A DA algorithm runs and assigns each student to a seat.

   (a) Step 1: Each student proposes to their first choice. Then each program tentatively assigns its seats to its proposers in descending order of program-specific weighted scores. All of the other students are tentatively rejected.

   (b) Step 2: All students that were rejected from their first choice propose to their second choice. Any of the programs that have seats left and are proposed to in step 2 assign their remaining seats to proposers in descending order of priority.

5. After learning the initial assignment, each student chooses one of three options: (1) to enroll in the given assignment, foregoing a reassignment round where they can be assigned to a program ranked at least as high in their list as the one they were

initially assigned to, (2) exit the centralized assignment process and enroll in a private university, or (3) participate in a reassignment round where they are guaranteed to be assigned in a program they ranked at least as high as the program they were initially assigned to

6. Round 1, phase 2 (the reassignment round): students who decide to participate in the reassignment round are assigned a seat among the seats remaining in the programs that were not filled in the initial round of assignment again through a DA procedure. The allocation in round 2 is the final assignment for each student who participated in the centralized assignment. At this stage, students can choose to enroll in their assigned program, enroll in a private program, or reject their assignment and wait for the second round.

7. Round 2: this round mainly serves for students who failed to qualify for university admission in the main round, those who were unassigned in the main round, and students who rejected their assignment in phase 2 of the main round. I do not describe this round because it is not relevant for the paper as the set of students who participate in this round would not be eligible to enroll in university at all in the post-reform period.

A.2.2 Post-reform mechanism: dynamic multi-offer

1. In June of each year starting in 2016, students take national exams, three obligatory exams in math, Albanian language and literature, and a foreign language, and one elective exam in a subject chosen by each student.

2. In July, grades from each of the exams become public through a set of lists.

3. In August, the admissions process for all programs begins, with each student applying to up to ten programs through the centralized platform, and submitting *unordered* portfolios.

4. The admissions procedure with 7 phases lasting 48 hours each unfolds:

   (a) Phase 1: Ranked lists of applicants in decreasing order of weighted average score are published by the mechanism for each program and students observe their position on each list and whether they have cleared the cutoff for each program in this phase. The student also observes the last person to be admitted by each of the programs. At this stage a few possible scenarios may happen:

      - Students who have cleared the cutoff of at least one of the programs but not all, face three choices. The first is to accept any of the offers received in phase 1, and forgo all other options in their original portfolio. The second option is to forgo all options received in phase 1 and wait for results of the next phase for the remainder of the programs in their portfolio. Third, they may choose to exit the mechanism unmatched.
      - Students who have cleared the cutoff of all their programs may either choose to enroll in one of the programs, or exit the mechanism and forgo all offers received.

      At the end of phase 1, all seats but those taken by students who decided to accept an offer free up for students in the next phase. All rejected offers are removed from students' lists.

   (b) Phases 2-6: At the beginning of each phase n between 2 and 6, students are ranked by each of the remaining programs in their portfolio that have empty seats left. They observe their new ranking relative to the remaining applicants in each of the programs, observe their phase-n offers in that program and observe the phase n cutoffs. They make their enrollment or waiting decisions as in phase 1.

   (c) Phase 7: The final offers of the main round realize and students make their last-chance enrollment decisions for the round.

5. Round 2: this round mainly serves for students who failed were unassigned in the main

round. I do not describe this round.

A.3 Data appendix

A.3.1 Merit scholarship policies in private universities

Figure A.10: An example of a university posting its scholarship policy on the website

Bachelor/Integrated study program Scholarship

Epoka University grants merit-based scholarships for the best ones for the entire normal duration of their studies as follows:

Scholarships based on the results of high school studies and State Matura Exam of the Republic of Albania:

Point Range	Scholarship
6600 or more	100%
6400 – 6599.99	75%
6200 – 6399.99	50%

Scholarships based on SAT Exam results:

Point Range	Scholarship
1350 or more	100%
1290 - 1349	75%
1220 - 1289	50%

The number of scholarships is limited (10 - 15 scholarships for each study program). Applications after the aforementioned position will benefit from a scholarship with minus 25% of the respective scholarship amount. Candidates interested to receive a scholarship should apply at the earliest in order to benefit from the above-mentioned scholarships since their allocation will be carried out according to the date of application.

Note: Are eligible to apply for scholarship the candidates that have high school average at least 9.0

Click here to calculate your Matura points.

Social Scholarship

- Two scholarships for candidates who belong to the " children of Albanian policemen killed in line of duty"
- Two scholarships for candidates who are sportsmen with very high results in the respective Olympic sports category in Albania
- A scholarship for candidates who belong to the Roma/Egyptian community from Albania.
- Two scholarships for candidates who belong to the "orphan coming from a low-income family.

Note: This figure shows scholarship policies posted by Epoka University. https://admissions.epoka.edu.al/home-bachelor-integrated-study-program-scholarship-2745-497.html

A.4 Simplifying the likelihood function

A.4.1 One-shot swaps

Proposition 1 of Larroucau and Rios (2020) shows that for a portfolio selection problem with ordered lists where probabilities of admission to each program are independent, it suffices to show that the chosen portfolio is preferred to all its one-shot swaps for the portfolio to be optimal. The following reformulation of the proposition is applicable to portfolios in settings where the ordering of the list does not matter for payoffs.

[adapted from Larroucou and Rios] Let $C = \{j_1, ..., j_k\}$ be an unordered application list of length at most K, i.e. k $\leq K$. Without loss of generality, let $u_{j_1} \geq u_{j_2} \geq ... \geq u_{j_k}$ so that the utility from submitting application portfolio C is

$$V(C) = p_{j_1}u_{j_1} + (1-p_{j_1})p_{j_2}u_{j_2} + ... + \left(\prod_{l=1}^{l=k-1}(1-p_{j_l})\right)p_{j_k}u_{j_1}$$

If $\mathcal{S}(C)$ is the set of one-shot swaps of portfolio C and

$$V(C) \geq V(C') \ \forall \ C' \in \mathcal{S}(C), \tag{A.1}$$

then

$$V(C) \geq V(C') \ \forall \ \ C' \ \ s.t. \ \ |C'| = K. \tag{A.2}$$

Discussion: In the context of rank-ordered lists, unprofitability of one-shot swaps implies more restrictions than in the case of unordered lists. Take for example a case with lists of size 2. The set of one-shot swaps of a portfolio $C_o = (j_1, j_2)$ are $\mathcal{S}(C_o) = \{(j_1, x), (x, j_1), (x, j_2), (j_2, x)\}$ whereas the set of one-shot swaps for an unordered portfolio $C_u = \{j_1, j_2\}$ are $\mathcal{S}(C_u) = \{\{j_1, x\}, \{j_2, x\}\}$. I show below that even starting with fewer inequalities on OSS as in the case of unordered portfolios, optimality of a portfolio given the unprofitability of its OSS is satisfied.

Proof. Proof for the DA rank ordered lists can be found in Larroucau and Rios (2020). See below for a fast sketch of the proof for the case of unordered portfolios that follows the original proof. The proof is done by induction. The case for $K = 1$ is obvious. For $K = 2$, suppose that $C = \{j_1, j_2\}$ and let $\mathcal{S}(C) = \Big\{\{x, j_1\}, \{x, j_2\}\Big\} \forall x \in \mathcal{J} \backslash C$. Let $C \succeq \{x, j_1\}$ and $C \succeq \{x, j_2\}$ $\forall x \in \mathcal{J}$. Then for all $\{x, y\}$, suppose WLOG that $u_x \geq u_y$ and $u_{j_1} \geq u_{j_2}$. We have a few cases:

Case 1: $u_{j_1} \geq u_y \geq u_{j_2}$: Using the fact that $\{j_1, j_2\} \succeq \{j_1, y\}$ implies that $p_{j_2}u_{j_2} \geq p_y u_y$,

$$V(\{x, y\}) = p_x u_x + (1 - p_x)p_y u_y \leq p_x u_x + (1 - p_x)p_{j_2}u_{j_2} = V(\{x, j_2\}) \leq V(\{j_1, j_2\}) = V(C).$$

Case 2a: $u_{j_1} \geq u_{j_2} \geq u_y$ and $u_x \geq u_{j_2}$:

$$V(\{x, y\}) = p_x u_x + (1 - p_x)p_y u_y \leq p_x u_x + (1 - p_x)p_{j_2}u_{j_2} = V(\{j_2, x\}) \leq V(\{j_1, j_2\}) = V(C)$$

Case 2b: $u_{j_1} \geq u_{j_2} \geq u_y$ and $u_x \leq u_{j_2}$:

$$\begin{aligned} V(\{x, y\}) = p_x u_x + (1 - p_x)p_y u_y &\leq p_x u_x + (1 - p_x)p_{j_2}u_{j_2} \\ &\leq p_{j_2}u_{j_2} + (1 - p_{j_2})p_x u_x = V(\{j_2, x\}) \leq V(\{j_1, j_2\}) = V(C) \end{aligned}$$

The first inequality holds because $\{j_1, j_2\} \succeq \{j_1, y\} \implies p_{j_2}u_{j_2} \geq p_y u_y$. The second inequality holds because $j_2 \succeq x$ so in the case of admission to both, the applicant will choose j_2.

Case 3: $u_y \geq u_{j_1} \geq u_{j_2}$:

$$V(\{x, y\}) = p_x u_x + (1 - p_x) p_y u_y \leq p_x u_x + (1 - p_x) p_{j_2} u_{j_2} = V(\{x, j_2\}) \leq V(\{j_1, j_2\}) = V(C)$$

For the inductive step, assume that the theorem holds for portfolios of length k. It remains to show that the theorem holds for portfolios of length $k + 1$. The rest of the proof goes as follows: suppose portfolio C_{k+1} satisfies $C_{k+1} \succeq C'_{k+1}\ \forall\ C'_{k+1} \in \mathcal{S}(C_{k+1})$. First show that, C_k, the portfolio that has highest utility among the k-sized subsets of C_{k+1} satisfies $C_k \succeq C'_k\ \forall\ C'_k \in \mathcal{S}(C_k)$. This implies that C_k is the optimal portfolio among all k-sized portfolios by the inductive assumption. Then show that the remaining element added to form C_{k+1} is added by Marginal Improvement Algorithm (which **?** show to be optimal), which implies that the final C_{k+1} is the optimal portfolio of size $k + 1$. $\square$

APPENDIX B

How Costs Limit Contraceptive Use among Low-Income Women in the U.S.: A Randomized Control Trial

B.1 Transitions in Contraceptive Method Use

Table B.1: Voucher Amounts by Income Group and Study Phase

Income as Share of	Sliding Scale:	50% Phase		100% Phase	
Federal Poverty Line (FPL)	% of Fee Charged	Control	Treatment	Control	Treatment
$\leq$ 100%	0%	\$0	\$0	\$0	\$0
101-150%	25%	\$0	\$123	\$0	\$223
151-200%	50%	\$0	\$246	\$0	\$446
201-250%	75%	\$0	\$369	\$0	\$669
$\geq$ 251%	100%	\$0	\$492	\$0	\$892

Notes: Participants in the 50% phase received vouchers between August 20, 2018, and March 3, 2019. Participants in the 100% phase received vouchers between March 4, 2019, and November 3, 2019. The tablet customized voucher amounts to each patient's out-of-pocket costs for contraceptives based on PPMI's assessment of their income. Patients who were below the FPL (fee scale 1/A) are not charged for contraceptive services and are, therefore, excluded from the study. Uninsured patients with incomes at 101-150% of the federal poverty line (FPL, fee scale 2) pay 25% of PPMI prices; 151-200% (fee scale 3) pay 50%; 201-250% (fee scale 4) pay 75%; and above 250% (fee scale 5) pay 100%.

Table B.2: Method Transitions

A. Contraceptive Switching Matrix, 50% Treatment Group

	Most Effective Method Billed Post-Visit and within 100 Days of Enrollment						
Most Effective Birth Control Method Pre-Visit	LARC	Shot	Pill	Ring/Patch	Non-Hormonal[2]	Did not purchase BC at PPMI	Total
LARC	5	0	6	2	27	0	40
Shot	2	11	0	1	2	0	16
Pill	10	0	80	2	20	0	112
Ring/Patch	1	0	1	5	2	0	9
Non-Hormonal[2]	12	6	27	5	22	1	73
No Method[3]	11	13	25	0	22	0	71
Total	41	30	139	15	95	1	321
Switched to more effective							113 (0.352)
Stayed on same method							102 (0.318)
Switched to less effective							11 (0.034)
No purchase of BC at PPMI							95 (0.296)

B. Contraceptive Switching Matrix, 50% Control Group

	Most Effective Method Billed Post-Visit and within 100 Days of Enrollment						
Most Effective Birth Control Method Pre-Visit	LARC	Shot	Pill	Ring/Patch	Non-Hormonal[2]	Did not purchase BC at PPMI	Total
LARC	2	0	0	3	1	38	45
Vasectomy/Sterilization	0	0	0	0	0	1	1
Shot	0	16	2	0	0	1	19
Pill	6	2	48	2	2	45	103
Ring/Patch	0	0	1	5	0	0	6
Non-Hormonal[2]	9	7	18	7	0	34	75
No Method[3]	5	9	19	3	0	34	69
Total	22	34	88	20	1	153	318
Switched to more effective							85 (0.267)
Stayed on same method							71 (0.223)
Switched to less effective							9 (0.028)
No purchase of BC at PPMI							153 (0.481)

C. Contraceptive Switching Matrix, 100% Treatment Group

Most Effective Birth Control Method Pre-Visit	Most Effective Method Billed Post-Visit and within 100 Days of Enrollment						
	LARC	Shot	Pill	Ring/Patch	Non-Hormonal[2]	Did not purchase BC at PPMI	Total
LARC	19	0	10	2	0	39	70
Shot	4	26	1	0	0	3	34
Pill	19	5	94	3	0	25	146
Ring/Patch	2	0	2	5	0	2	11
Non-Hormonal[2]	32	8	30	9	1	39	119
No Method[3]	15	13	36	3	0	49	116
Total	91	52	173	22	1	157	496
Switched to more effective							178 (0.359)
Stayed on same method							145 (0.292)
Switched to less effective							16 (0.032)
No purchase of BC at PPMI							157 (0.317)

D. Contraceptive Switching Matrix, 100% Control Group

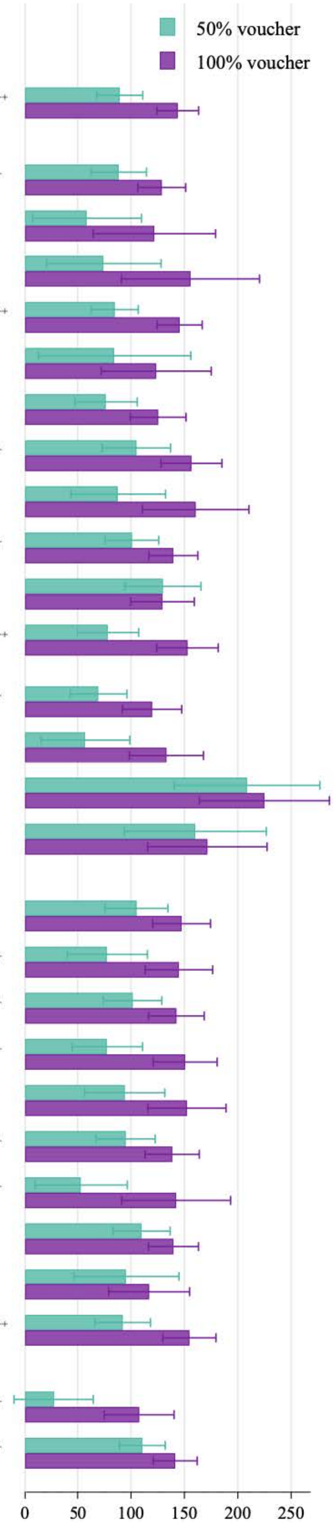
Most Effective Birth Control Method Pre-Visit	Most Effective Method Billed Post-Visit and within 100 Days of Enrollment						
	LARC	Shot	Pill	Ring/Patch	Non-Hormonal[2]	Did not purchase BC at PPMI	Total
LARC	4	0	10	2	0	47	63
Shot	0	21	1	0	0	6	28
Pill	5	4	60	1	0	62	132
Ring/Patch	1	0	0	8	0	3	12
Non-Hormonal[2]	6	12	28	2	1	60	109
No Method[3]	4	14	35	4	0	61	118
Total	20	51	134	17	1	239	462
Switched to more effective							115 (0.249)
Stayed on same method							94 (0.203)
Switched to less effective							14 (0.030)
No purchase of BC at PPMI							239 (0.517)

Notes: [1]Post-enrollment birth control methods come from the PPMI billing records. [2]Non-Hormonal includes: diaphragm, condom, withdrawal, rhythm, spermicide. [3]Baseline no method includes: abstinence, Plan B, abortion, miscarriage, and no method reported.

B.2 Heterogeneity in main paper results for each index component

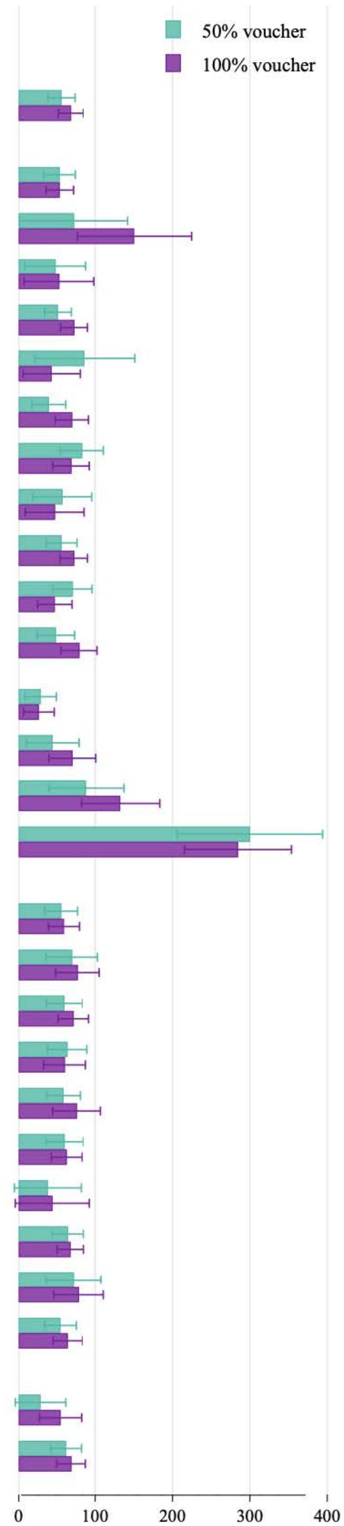
Figure B.1: Heterogeneity in the Treatment Effects of Receiving a Voucher on the Five Primary Outcomes

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on PPMI charges	300	0.72 (0.03)	181 (25)	268 (33)	287	0.66 (0.02)	270 +++ (22)	413 +++ (29)
A. Pre-specified demographic groups								
Non-Hispanic White	309	0.73 (0.03)	200 (32)	274 (41)	308	0.69 (0.03)	281 + (28)	397 ++ (35)
Non-Hispanic Black	279	0.60 (0.09)	90 (54)	163 (73)	238	0.49 (0.07)	137 (51)	291 (70)
Hispanic any race	310	0.72 (0.08)	158 (72)	230 (86)	270	0.63 (0.07)	316 (70)	420 (89)
Women without children	311	0.74 (0.03)	184 (27)	263 (35)	286	0.66 (0.02)	274 ++ (24)	416 +++ (31)
Mothers	271	0.62 (0.07)	131 (70)	229 (99)	301	0.70 (0.06)	268 (59)	371 (79)
Age<26	309	0.71 (0.04)	159 (34)	235 (47)	281	0.63 (0.03)	229 (28)	353 + (37)
Age≥26	290	0.72 (0.04)	206 (37)	304 (47)	296	0.69 (0.03)	308 ++ (34)	463 ++ (43)
Below associate's degree	265	0.77 (0.06)	176 (52)	232 (61)	254	0.63 (0.05)	243 (47)	408 + (65)
Associate's degree or higher	300	0.74 (0.03)	211 (30)	302 (38)	299	0.73 (0.03)	305 ++ (29)	417 ++ (35)
Married or cohabitating	268	0.80 (0.04)	276 (43)	348 (49)	296	0.75 (0.03)	285 (38)	384 (45)
Single	311	0.70 (0.04)	153 (34)	243 (46)	287	0.69 (0.03)	297 +++ (33)	439 +++ (42)
Fee scale								
101-150% FPL	312	0.75 (0.04)	157 (34)	216 (43)	307	0.66 (0.03)	249 ++ (32)	368 ++ (44)
151-200% FPL	358	0.71 (0.05)	134 (59)	204 (77)	295	0.69 (0.04)	273 + (44)	393 + (52)
201-250% FPL	241	0.76 (0.07)	358 (70)	503 (84)	244	0.68 (0.06)	361 (65)	550 (76)
250+% FPL	203	0.59 (0.08)	186 (49)	325 (69)	257	0.62 (0.06)	254 (54)	441 (73)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	283	0.75 (0.03)	219 (34)	297 (42)	285	0.74 (0.03)	309 + (33)	420 + (39)
Do not have a usual place for birth control	331	0.71 (0.05)	174 (46)	256 (64)	295	0.67 (0.04)	272 + (37)	428 ++ (48)
Using Tier 1 or 2 method	276	0.74 (0.03)	201 (31)	281 (38)	283	0.70 (0.03)	291 ++ (31)	404 ++ (38)
Not using Tier 1 or 2 method	336	0.70 (0.04)	159 (42)	260 (57)	288	0.62 (0.03)	259 + (31)	434 ++ (44)
Delayed getting birth control	346	0.90 (0.04)	277 (60)	325 (66)	297	0.77 (0.04)	353 (48)	453 (55)
Did not delay getting birth control	280	0.68 (0.03)	181 (30)	266 (39)	287	0.68 (0.03)	270 ++ (29)	397 ++ (37)
Positive desire to have a baby	305	0.68 (0.06)	89 (50)	161 (68)	277	0.56 (0.05)	215 + (45)	394 ++ (72)
Negative desire to have a baby	293	0.75 (0.03)	229 (33)	323 (40)	295	0.75 (0.02)	302 + (30)	413 (35)
More likely to meet career aspirations	293	0.80 (0.05)	216 (59)	279 (74)	312	0.66 (0.04)	237 (47)	365 (60)
Less likely to meet career aspirations	301	0.73 (0.03)	195 (31)	277 (40)	282	0.71 (0.03)	314 +++ (29)	436 +++ (36)
C. Exploratory								
Planning to get a LARC before appointment	649	0.77 (0.06)	98 (110)	176 (124)	516	0.83 (0.05)	473 +++ (83)	555 ++ (86)
Not planing to get a LARC before appointment	247	0.70 (0.03)	186 (21)	273 (27)	256	0.63 (0.02)	225 (20)	363 ++ (27)



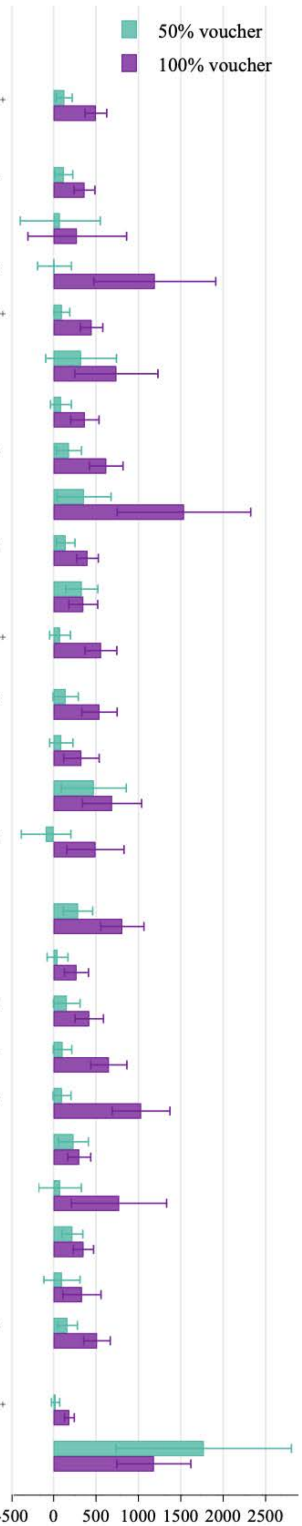
(a) PPMI charges in dollars

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on any birth control purchase	0.52	0.72 (0.03)	0.18 (0.04)	0.29 (0.05)	0.48	0.66 (0.02)	0.19 (0.03)	0.33 (0.04)
A. Pre-specified demographic groups								
Non-Hispanic White	0.53	0.73 (0.03)	0.19 (0.04)	0.28 (0.05)	0.52	0.69 (0.03)	0.17 (0.04)	0.28 (0.05)
Non-Hispanic Black	0.45	0.60 (0.09)	0.14 (0.12)	0.32 (0.16)	0.35	0.49 (0.07)	0.21 (0.09)	0.53 (0.13)
Hispanic any race	0.60	0.72 (0.08)	0.10 (0.11)	0.28 (0.12)	0.55	0.63 (0.07)	0.18 (0.10)	0.29 (0.13)
Women without children	0.55	0.74 (0.03)	0.18 (0.04)	0.28 (0.05)	0.47	0.66 (0.02)	0.20 (0.03)	0.34 (0.04)
Mothers	0.40	0.62 (0.07)	0.16 (0.10)	0.35 (0.13)	0.55	0.70 (0.06)	0.13 (0.09)	0.23 (0.10)
Age<26	0.58	0.71 (0.04)	0.13 (0.05)	0.23 (0.06)	0.49	0.63 (0.03)	0.18 (0.04)	0.34 (0.06)
Age≥26	0.45	0.72 (0.04)	0.24 (0.06)	0.37 (0.07)	0.48	0.69 (0.03)	0.21 (0.05)	0.33 (0.06)
Below associate's degree	0.53	0.77 (0.06)	0.19 (0.09)	0.30 (0.10)	0.53	0.63 (0.05)	0.12 (0.08)	0.25 (0.10)
Associate's degree or higher	0.52	0.74 (0.03)	0.20 (0.04)	0.29 (0.05)	0.49	0.73 (0.03)	0.25 (0.04)	0.35 (0.05)
Married or cohabitating	0.50	0.80 (0.04)	0.28 (0.06)	0.35 (0.07)	0.57	0.75 (0.03)	0.20 (0.05)	0.26 (0.06)
Single	0.54	0.70 (0.04)	0.15 (0.05)	0.26 (0.07)	0.45	0.69 (0.03)	0.23 (0.04)	0.36 (0.06)
Fee scale								
101-150% FPL	0.62	0.75 (0.04)	0.11 (0.05)	0.18 (0.06)	0.62	0.66 (0.03)	0.08 (0.05)	0.16 (0.06)
151-200% FPL	0.54	0.71 (0.05)	0.15 (0.08)	0.24 (0.09)	0.47	0.69 (0.04)	0.21 (0.06)	0.33 (0.07)
201-250% FPL	0.45	0.76 (0.07)	0.30 (0.11)	0.40 (0.11)	0.36	0.68 (0.06)	0.32 (0.08)	0.48 (0.09)
250+% FPL	0.22	0.59 (0.08)	0.35 (0.10)	0.67 (0.11)	0.24	0.62 (0.06)	0.41 (0.08)	0.68 (0.08)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	0.54	0.75 (0.03)	0.19 (0.05)	0.30 (0.06)	0.53	0.74 (0.03)	0.22 (0.05)	0.31 (0.05)
Do not have a usual place for birth control	0.48	0.71 (0.05)	0.23 (0.07)	0.33 (0.08)	0.45	0.67 (0.04)	0.23 (0.05)	0.35 (0.07)
Using Tier 1 or 2 method	0.51	0.74 (0.03)	0.20 (0.05)	0.30 (0.06)	0.50	0.70 (0.03)	0.23 (0.04)	0.35 (0.05)
Not using Tier 1 or 2 method	0.51	0.70 (0.04)	0.17 (0.06)	0.32 (0.07)	0.47	0.62 (0.03)	0.15 (0.05)	0.28 (0.07)
Delayed getting birth control	0.59	0.90 (0.04)	0.32 (0.07)	0.35 (0.07)	0.47	0.77 (0.04)	0.27 (0.06)	0.36 (0.08)
Did not delay getting birth control	0.49	0.68 (0.03)	0.18 (0.05)	0.29 (0.06)	0.51	0.68 (0.03)	0.20 (0.04)	0.31 (0.05)
Positive desire to have a baby	0.53	0.68 (0.06)	0.12 (0.09)	0.20 (0.12)	0.46	0.56 (0.05)	0.06 (0.08)	0.20 (0.11)
Negative desire to have a baby	0.51	0.75 (0.03)	0.22 (0.05)	0.33 (0.05)	0.52	0.75 (0.02)	0.24 (0.04)	0.35 (0.05)
More likely to meet career aspirations	0.50	0.80 (0.05)	0.26 (0.08)	0.36 (0.09)	0.47	0.66 (0.04)	0.20 (0.06)	0.37 (0.08)
Less likely to meet career aspirations	0.53	0.73 (0.03)	0.19 (0.05)	0.28 (0.06)	0.51	0.71 (0.03)	0.22 (0.04)	0.32 (0.05)
C. Exploratory								
Planning to get a LARC before appointment	0.64	0.77 (0.06)	0.10 (0.10)	0.18 (0.11)	0.60	0.83 (0.05)	0.25 (0.08)	0.32 (0.08)
Not planing to get a LARC before appointment	0.50	0.70 (0.03)	0.19 (0.04)	0.31 (0.05)	0.47	0.63 (0.02)	0.18 (0.03)	0.32 (0.05)



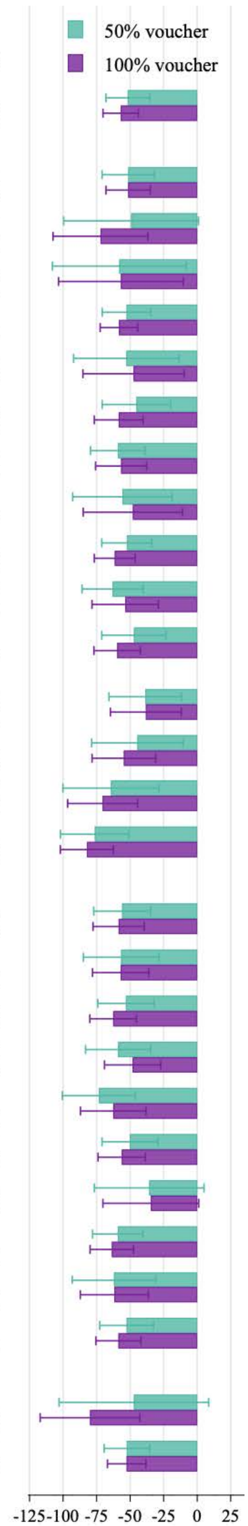
(b) Any birth control purchase

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on larc insertion	0.07	0.72 (0.03)	0.05 (0.02)	0.09 (0.03)	0.04	0.66 (0.02)	0.14^{+++} (0.02)	0.22^{+++} (0.03)
A. Pre-specified demographic groups								
Non-Hispanic White	0.08	0.73 (0.03)	0.06 (0.03)	0.09 (0.04)	0.06	0.69 (0.03)	0.15^{++} (0.03)	0.20^{++} (0.03)
Non-Hispanic Black	0.03	0.60 (0.09)	0.01 (0.04)	0.02 (0.06)	0.02	0.49 (0.07)	0.01 (0.04)	0.05 (0.05)
Hispanic any race	0.10	0.72 (0.08)	0.03 (0.07)	0.01 (0.10)	0.02	0.63 (0.07)	0.21^{+} (0.06)	0.27^{++} (0.08)
Women without children	0.08	0.74 (0.03)	0.05 (0.03)	0.07 (0.04)	0.05	0.66 (0.02)	0.13^{++} (0.02)	0.20^{+++} (0.03)
Mothers	0.04	0.62 (0.07)	0.07 (0.06)	0.14 (0.09)	0.03	0.70 (0.06)	0.19 (0.05)	0.23 (0.08)
Age<26	0.08	0.71 (0.04)	0.05 (0.03)	0.07 (0.05)	0.04	0.63 (0.03)	0.12^{+} (0.03)	0.17 (0.04)
Age≥26	0.06	0.72 (0.04)	0.06 (0.03)	0.11 (0.05)	0.04	0.69 (0.03)	0.16^{++} (0.03)	0.25^{++} (0.04)
Below associate's degree	0.04	0.77 (0.06)	0.11 (0.05)	0.14 (0.06)	0.01	0.63 (0.05)	0.13 (0.04)	0.23 (0.06)
Associate's degree or higher	0.07	0.74 (0.03)	0.06 (0.03)	0.10 (0.04)	0.05	0.73 (0.03)	0.15^{++} (0.03)	0.22^{++} (0.04)
Married or cohabitating	0.05	0.80 (0.04)	0.14 (0.04)	0.17 (0.05)	0.05	0.75 (0.03)	0.13 (0.04)	0.18 (0.05)
Single	0.07	0.70 (0.04)	0.03 (0.03)	0.05 (0.04)	0.04	0.69 (0.03)	0.16^{+++} (0.03)	0.24^{+++} (0.04)
Fee scale								
101-150% FPL	0.05	0.75 (0.04)	0.05 (0.03)	0.08 (0.04)	0.04	0.66 (0.03)	0.14^{++} (0.03)	0.22^{++} (0.04)
151-200% FPL	0.11	0.71 (0.05)	0.05 (0.05)	0.10 (0.08)	0.05	0.69 (0.04)	0.11 (0.04)	0.15 (0.05)
201-250% FPL	0.05	0.76 (0.07)	0.14 (0.07)	0.24 (0.10)	0.05	0.68 (0.06)	0.22 (0.06)	0.31 (0.08)
250+% FPL	0.07	0.59 (0.08)	-0.02 (0.05)	-0.06 (0.10)	0.04	0.62 (0.06)	0.12^{++} (0.05)	0.22^{++} (0.08)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	0.05	0.75 (0.03)	0.10 (0.03)	0.13 (0.04)	0.03	0.74 (0.03)	0.17^{+} (0.03)	0.24^{+} (0.04)
Do not have a usual place for birth control	0.11	0.71 (0.05)	0.04 (0.05)	0.05 (0.07)	0.07	0.67 (0.04)	0.12 (0.03)	0.18 (0.05)
Using Tier 1 or 2 method	0.05	0.74 (0.03)	0.05 (0.03)	0.07 (0.04)	0.04	0.70 (0.03)	0.13^{++} (0.03)	0.18^{++} (0.04)
Not using Tier 1 or 2 method	0.10	0.70 (0.04)	0.06 (0.04)	0.11 (0.06)	0.04	0.62 (0.03)	0.16^{+} (0.03)	0.27^{++} (0.05)
Delayed getting birth control	0.13	0.90 (0.04)	0.11 (0.06)	0.13 (0.07)	0.03	0.77 (0.04)	0.25^{+} (0.04)	0.33^{++} (0.05)
Did not delay getting birth control	0.04	0.68 (0.03)	0.07 (0.03)	0.10 (0.04)	0.05	0.68 (0.03)	0.11 (0.03)	0.16 (0.04)
Positive desire to have a baby	0.05	0.68 (0.06)	0.00 (0.04)	0.04 (0.07)	0.02	0.56 (0.05)	0.09 (0.04)	0.18 (0.07)
Negative desire to have a baby	0.07	0.75 (0.03)	0.10 (0.03)	0.14 (0.04)	0.06	0.75 (0.02)	0.15 (0.03)	0.21 (0.04)
More likely to meet career aspirations	0.06	0.80 (0.05)	0.05 (0.05)	0.06 (0.07)	0.05	0.66 (0.04)	0.11 (0.04)	0.16 (0.05)
Less likely to meet career aspirations	0.07	0.73 (0.03)	0.07 (0.03)	0.11 (0.04)	0.05	0.71 (0.03)	0.17^{++} (0.03)	0.24^{++} (0.04)
C. Exploratory								
Planning to get a LARC before appointment	0.50	0.77 (0.06)	0.06 (0.11)	0.11 (0.12)	0.29	0.83 (0.05)	0.46^{+++} (0.08)	0.54^{+++} (0.09)
Not planing to get a LARC before appointment	0.00	0.70 (0.03)	0.04 (0.01)	0.06 (0.02)	0.01	0.63 (0.02)	0.07 (0.01)	0.12^{+} (0.02)



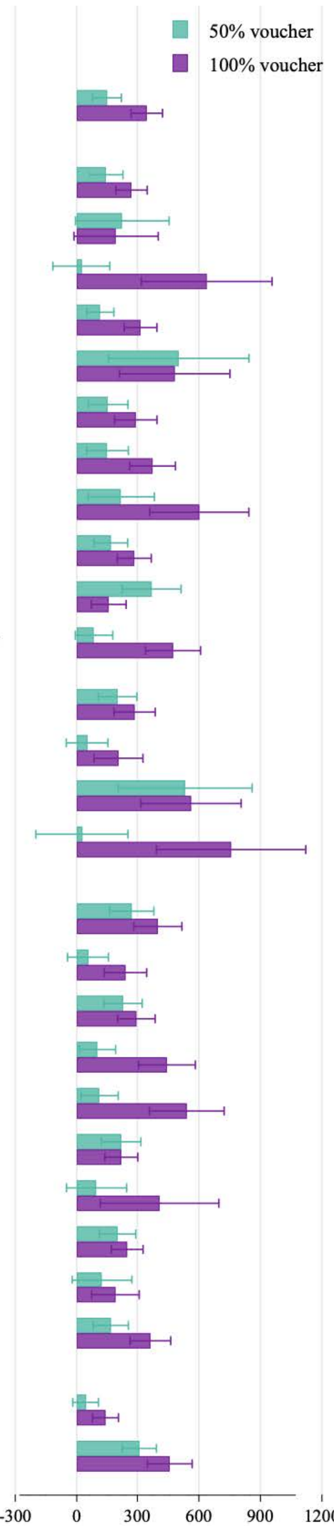
(c) LARC insertion

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on expected annual pregnancies	0.52	0.72 (0.03)	-0.17 (0.03)	-0.27 (0.04)	0.55	0.66 (0.02)	-0.19 (0.03)	-0.32 (0.04)
A. Pre-specified demographic groups								
Non-Hispanic White	0.51	0.73 (0.03)	-0.18 (0.04)	-0.26 (0.05)	0.52	0.69 (0.03)	-0.17 (0.03)	-0.27 (0.04)
Non-Hispanic Black	0.58	0.60 (0.09)	-0.13 (0.11)	-0.29 (0.15)	0.68	0.49 (0.07)	-0.20 (0.09)	-0.49 (0.12)
Hispanic any race	0.45	0.72 (0.08)	-0.09 (0.10)	-0.26 (0.11)	0.50	0.63 (0.07)	-0.18 (0.09)	-0.28 (0.12)
Women without children	0.49	0.74 (0.03)	-0.17 (0.04)	-0.26 (0.05)	0.56	0.66 (0.02)	-0.20 (0.03)	-0.33 (0.04)
Mothers	0.63	0.62 (0.07)	-0.15 (0.10)	-0.33 (0.13)	0.50	0.70 (0.06)	-0.14 (0.08)	-0.24 (0.10)
Age<26	0.46	0.71 (0.04)	-0.12 (0.05)	-0.21 (0.06)	0.55	0.63 (0.03)	-0.17 (0.04)	-0.32 (0.05)
Age≥26	0.58	0.72 (0.04)	-0.23 (0.05)	-0.35 (0.06)	0.56	0.69 (0.03)	-0.20 (0.04)	-0.32 (0.05)
Below associate's degree	0.51	0.77 (0.06)	-0.18 (0.08)	-0.28 (0.10)	0.51	0.63 (0.05)	-0.12 (0.07)	-0.25 (0.10)
Associate's degree or higher	0.52	0.74 (0.03)	-0.18 (0.04)	-0.27 (0.05)	0.55	0.73 (0.03)	-0.24 (0.04)	-0.34 (0.04)
Married or cohabitating	0.53	0.80 (0.04)	-0.26 (0.06)	-0.34 (0.06)	0.48	0.75 (0.03)	-0.20 (0.05)	-0.26 (0.06)
Single	0.50	0.70 (0.04)	-0.14 (0.05)	-0.24 (0.06)	0.58	0.69 (0.03)	-0.22 (0.04)	-0.35 (0.05)
Fee scale								
101-150% FPL	0.43	0.75 (0.04)	-0.11 (0.05)	-0.17 (0.06)	0.43	0.66 (0.03)	-0.09 (0.04)	-0.16 (0.06)
151-200% FPL	0.50	0.71 (0.05)	-0.15 (0.07)	-0.22 (0.09)	0.57	0.69 (0.04)	-0.20 (0.06)	-0.31 (0.07)
201-250% FPL	0.58	0.76 (0.07)	-0.28 (0.10)	-0.37 (0.11)	0.66	0.68 (0.06)	-0.31 (0.08)	-0.47 (0.09)
250+% FPL	0.79	0.59 (0.08)	-0.32 (0.09)	-0.60 (0.10)	0.78	0.62 (0.06)	-0.38 (0.07)	-0.64 (0.08)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	0.50	0.75 (0.03)	-0.18 (0.05)	-0.28 (0.05)	0.51	0.74 (0.03)	-0.21 (0.04)	-0.30 (0.05)
Do not have a usual place for birth control	0.55	0.71 (0.05)	-0.21 (0.06)	-0.31 (0.08)	0.58	0.67 (0.04)	-0.22 (0.05)	-0.33 (0.06)
Using Tier 1 or 2 method	0.53	0.74 (0.03)	-0.19 (0.05)	-0.28 (0.06)	0.54	0.70 (0.03)	-0.22 (0.04)	-0.34 (0.05)
Not using Tier 1 or 2 method	0.52	0.70 (0.04)	-0.16 (0.05)	-0.31 (0.06)	0.57	0.62 (0.03)	-0.15 (0.04)	-0.27 (0.06)
Delayed getting birth control	0.45	0.90 (0.04)	-0.31 (0.06)	-0.33 (0.06)	0.56	0.77 (0.04)	-0.26 (0.06)	-0.35 (0.07)
Did not delay getting birth control	0.54	0.68 (0.03)	-0.17 (0.05)	-0.27 (0.06)	0.53	0.68 (0.03)	-0.19 (0.04)	-0.30 (0.05)
Positive desire to have a baby	0.52	0.68 (0.06)	-0.11 (0.08)	-0.18 (0.11)	0.58	0.56 (0.05)	-0.07 (0.07)	-0.20 (0.11)
Negative desire to have a baby	0.52	0.75 (0.03)	-0.21 (0.04)	-0.31 (0.05)	0.52	0.75 (0.02)	-0.23 (0.04)	-0.33 (0.04)
More likely to meet career aspirations	0.54	0.80 (0.05)	-0.25 (0.07)	-0.33 (0.09)	0.56	0.66 (0.04)	-0.19 (0.06)	-0.35 (0.07)
Less likely to meet career aspirations	0.51	0.73 (0.03)	-0.18 (0.04)	-0.27 (0.05)	0.53	0.71 (0.03)	-0.22 (0.04)	-0.31 (0.05)
C. Exploratory								
Planning to get a LARC before appointment	0.37	0.77 (0.06)	-0.09 (0.10)	-0.18 (0.11)	0.42	0.83 (0.05)	-0.27 (0.08)	-0.34 (0.08)
Not planing to get a LARC before appointment	0.54	0.70 (0.03)	-0.18 (0.04)	-0.28 (0.05)	0.57	0.63 (0.02)	-0.17 (0.03)	-0.30 (0.04)



(d) Expected annual pregnancies

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on temporal coverage	184	0.72	186	280	145	0.66	328 ++	503 ++
		(0.03)	(48)	(66)		(0.02)	(41)	(57)
A. Pre-specified demographic groups								
Non-Hispanic White	202	0.73	210	296	174	0.69	338	473
		(0.03)	(61)	(85)		(0.03)	(53)	(68)
Non-Hispanic Black	85	0.60	100	191	95	0.49	56	184
		(0.09)	(72)	(100)		(0.07)	(88)	(101)
Hispanic any race	250	0.72	56	61	92	0.63	444 ++	585 ++
		(0.08)	(134)	(180)		(0.07)	(121)	(149)
Women without children	206	0.74	168	241	150	0.66	315 ++	475 ++
		(0.03)	(53)	(71)		(0.02)	(44)	(61)
Mothers	83	0.62	254	417	116	0.70	432	561
		(0.07)	(125)	(145)		(0.06)	(107)	(160)
Age<26	176	0.71	187	272	141	0.63	291	411
		(0.04)	(58)	(89)		(0.03)	(54)	(74)
Age≥26	194	0.72	182	295	152	0.69	366 +	568 +
		(0.04)	(78)	(103)		(0.03)	(63)	(86)
Below associate's degree	145	0.77	227	318	75	0.63	264	450
		(0.06)	(106)	(122)		(0.05)	(64)	(92)
Associate's degree or higher	180	0.74	212	308	175	0.73	355 +	498 +
		(0.03)	(55)	(75)		(0.03)	(55)	(74)
Married or cohabitating	127	0.80	377	470	210	0.75	249	337
		(0.04)	(77)	(93)		(0.03)	(73)	(91)
Single	206	0.70	114	178	125	0.69	397 +++	591 +++
		(0.04)	(65)	(98)		(0.03)	(61)	(85)
Fee scale								
101-150% FPL	140	0.75	215	284	152	0.66	295	437
		(0.04)	(57)	(67)		(0.03)	(57)	(78)
151-200% FPL	288	0.71	71	151	170	0.69	243	353
		(0.05)	(114)	(153)		(0.04)	(76)	(103)
201-250% FPL	130	0.76	413	692	120	0.68	474	674
		(0.07)	(137)	(216)		(0.06)	(129)	(150)
250+% FPL	187	0.59	38	51	107	0.62	442 ++	807 ++
		(0.08)	(119)	(217)		(0.06)	(127)	(198)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	136	0.75	271	370	128	0.74	376	511
		(0.03)	(60)	(75)		(0.03)	(59)	(76)
Do not have a usual place for birth control	268	0.71	113	152	190	0.67	281	460 +
		(0.05)	(95)	(140)		(0.04)	(73)	(100)
Using Tier 1 or 2 method	129	0.74	222	297	150	0.70	317	442
		(0.03)	(50)	(61)		(0.03)	(54)	(70)
Not using Tier 1 or 2 method	264	0.70	145	275	136	0.62	356 +	605 ++
		(0.04)	(88)	(121)		(0.03)	(65)	(96)
Delayed getting birth control	314	0.90	322	357	127	0.77	527	688 +
		(0.04)	(133)	(148)		(0.04)	(97)	(118)
Did not delay getting birth control	126	0.68	194	279	167	0.68	248	369
		(0.03)	(46)	(62)		(0.03)	(50)	(68)
Positive desire to have a baby	175	0.68	68	172	116	0.56	250	476
		(0.06)	(98)	(133)		(0.05)	(89)	(172)
Negative desire to have a baby	180	0.75	256	366	189	0.75	345	472
		(0.03)	(63)	(82)		(0.02)	(58)	(75)
More likely to meet career aspirations	158	0.80	142	199	179	0.66	244	345
		(0.05)	(92)	(119)		(0.04)	(85)	(105)
Less likely to meet career aspirations	188	0.73	221	321	147	0.71	376 +	534 +
		(0.03)	(61)	(82)		(0.03)	(55)	(74)
C. Exploratory								
Planning to get a LARC before appointment	911	0.77	289	411	644	0.83	839 +	930
		(0.06)	(253)	(301)		(0.05)	(193)	(208)
Not planing to get a LARC before appointment	74	0.70	156	228	78	0.63	212	357 ++
		(0.03)	(25)	(31)		(0.02)	(29)	(43)



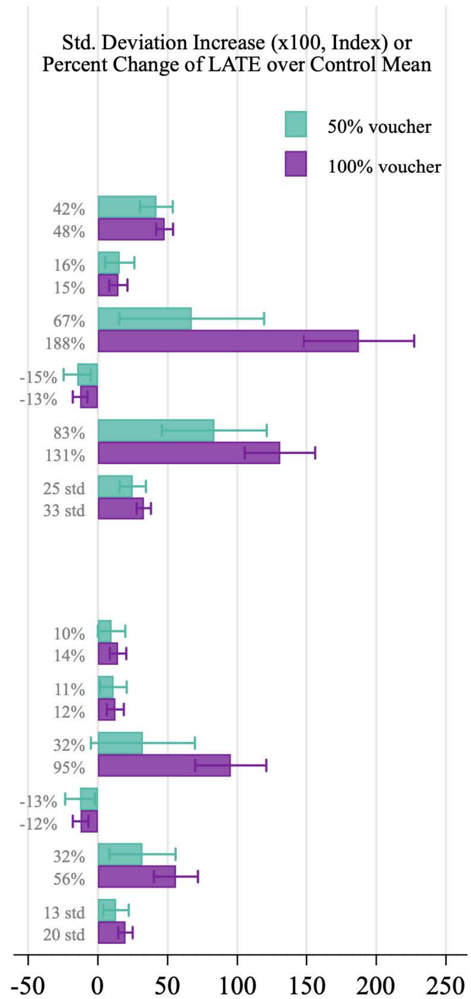
(e) Days covered by purchased contraception

B.3 Alternative first stage dependent variable: Voucher dollars spent

B.3.1 Treatment Effects of Receiving a Voucher on Contraceptive Efficacy

Figure B.2: Treatment Effects of Receiving a Voucher on Contraceptive Efficacy

	50% Voucher Group			100% Voucher Group		
	Control mean	ITT effect	LATE	Control mean	ITT effect	LATE
A. Contraceptive efficacy within 100 days of enrollment						
First stage: Voucher $ Spent		109.62			150.30+++	
		(6.42)			(8.65)	
PPMI charges in dollars	300	181	126	287	270 +++	137
		(25)	(18)		(22)	(9)
Any birth control purchase	0.52	0.18	0.08	0.48	0.19	0.07
		(0.04)	(0.03)		(0.03)	(0.02)
LARC insertion	0.07	0.05	0.05	0.04	0.14 +++	0.08
		(0.02)	(0.02)		(0.02)	(0.01)
Expected annual pregnancies	0.52	-0.17	-0.08	0.55	-0.19	-0.07
		(0.03)	(0.03)		(0.03)	(0.01)
Temporal coverage	184	186	154	145	328 ++	190
		(48)	(35)		(41)	(19)
Index of contraceptive efficacy	0.00	0.39	0.25	0.00	0.65 +++	0.33
		(0.07)	(0.05)		(0.06)	(0.03)
B. Contraceptive efficacy within two years of enrollment						
First stage: Voucher $ Spent		109.62			150.30+++	
		(6.42)			(8.65)	
PPMI charges in dollars	596	120	58	561	191	81
		(41)	(30)		(33)	(17)
Any birth control purchase	0.57	0.15	0.06	0.52	0.18	0.06
		(0.04)	(0.03)		(0.03)	(0.02)
LARC insertion	0.11	0.03	0.04	0.08	0.12 ++	0.07
		(0.03)	(0.02)		(0.02)	(0.01)
Expected annual pregnancies	0.47	-0.14	-0.06	0.52	-0.17	-0.06
		(0.03)	(0.03)		(0.03)	(0.01)
Temporal coverage	350	135	112	283	265 +	158
		(56)	(42)		(46)	(23)
Index of contraceptive efficacy	0.00	0.23	0.13	0.00	0.41 ++	0.20
		(0.06)	(0.05)		(0.06)	(0.03)

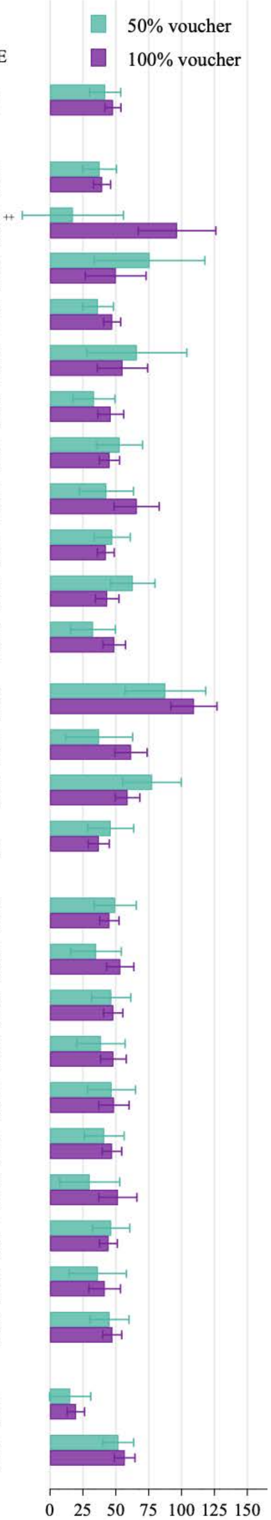


Note: Panel A presents the estimated treatment effects using equation 1 for participants up to 100 days after enrollment when the voucher expired; panel B presents the estimated treatment effects for participants at two years after enrollment. +++, ++ and + indicate that the 100% effect is statistically different from the 50% effect at the 1, 5, and 10% levels, respectively.

B.3.2 Heterogeneity in the Treatment Effects of Receiving a Voucher on the Five Primary Outcomes

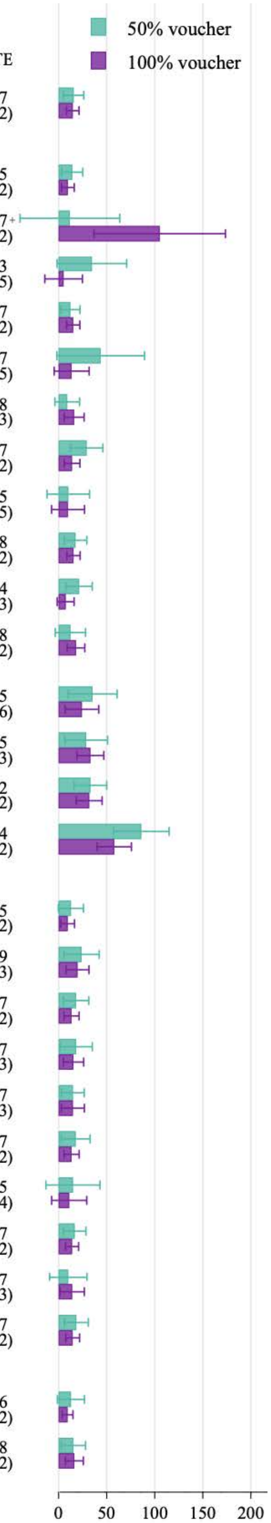
Figure B.3: Heterogeneity in the Treatment Effects of Receiving a Voucher on the Five Primary Outcomes

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on PPMI charges	300	110 (6)	181 (25)	126 (18)	287	150 (9)	270 +++ (22)	137 (9)
A. Pre-specified demographic groups								
Non-Hispanic White	309	120 (8)	200 (32)	116 (20)	308	168 (11)	281 + (28)	122 (10)
Non-Hispanic Black	279	80 (21)	90 (54)	48 (55)	238	64 (15)	137 (51)	230 ++ (36)
Hispanic any race	310	81 (13)	158 (72)	235 (67)	270	132 (27)	316 (70)	135 (32)
Women without children	311	118 (7)	184 (27)	113 (19)	286	156 (10)	274 ++ (24)	135 (9)
Mothers	271	60 (11)	131 (70)	179 (53)	301	125 (22)	268 (59)	166 (29)
Age<26	309	107 (9)	159 (34)	103 (25)	281	129 (11)	229 (28)	130 (14)
Age≥26	290	110 (10)	206 (37)	153 (26)	296	169 (13)	308 ++ (34)	134 (12)
Below associate's degree	265	107 (14)	176 (52)	114 (28)	254	112 (16)	243 (47)	167 (22)
Associate's degree or higher	300	117 (8)	211 (30)	142 (21)	299	176 (11)	305 ++ (29)	127 (10)
Married or cohabitating	268	129 (11)	276 (43)	169 (23)	296	162 (14)	285 (38)	129 (13)
Single	311	104 (9)	153 (34)	101 (27)	287	165 (14)	297 +++ (33)	140 (12)
Fee scale								
101-150% FPL	312	61 (4)	157 (34)	274 (49)	307	75 (5)	249 ++ (32)	336 (27)
151-200% FPL	358	109 (11)	134 (59)	134 (47)	295	152 (14)	273 + (44)	182 (18)
201-250% FPL	241	190 (24)	358 (70)	187 (27)	244	259 (32)	361 (65)	144 (12)
250+% FPL	203	203 (30)	186 (49)	94 (18)	257	271 (39)	254 (54)	95 (11)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	283	115 (9)	219 (34)	140 (23)	285	180 (15)	309 + (33)	128 (11)
Do not have a usual place for birth control	331	117 (12)	174 (46)	116 (32)	295	146 (13)	272 + (37)	158 (16)
Using Tier 1 or 2 method	276	121 (9)	201 (31)	129 (21)	283	164 (13)	291 ++ (31)	136 (11)
Not using Tier 1 or 2 method	336	99 (10)	159 (42)	130 (32)	288	141 (13)	259 + (31)	139 (14)
Delayed getting birth control	346	138 (15)	277 (60)	162 (32)	297	183 (18)	353 (48)	144 (18)
Did not delay getting birth control	280	104 (8)	181 (30)	115 (22)	287	155 (12)	270 ++ (29)	135 (11)
Positive desire to have a baby	305	92 (13)	89 (50)	92 (36)	277	126 (20)	215 + (45)	143 (20)
Negative desire to have a baby	293	121 (9)	229 (33)	136 (21)	295	175 (12)	302 + (30)	131 (10)
More likely to meet career aspirations	293	122 (15)	216 (59)	106 (33)	312	145 (17)	237 (47)	129 (19)
Less likely to meet career aspirations	301	112 (8)	195 (31)	136 (23)	282	165 (11)	314 +++ (29)	133 (10)
C. Exploratory								
Planning to get a LARC before appointment	649	132 (18)	98 (110)	99 (52)	516	341 (30)	473 +++ (83)	101 (17)
Not planing to get a LARC before appointment	247	104 (7)	186 (21)	128 (15)	256	119 (8)	225 (20)	145 (10)



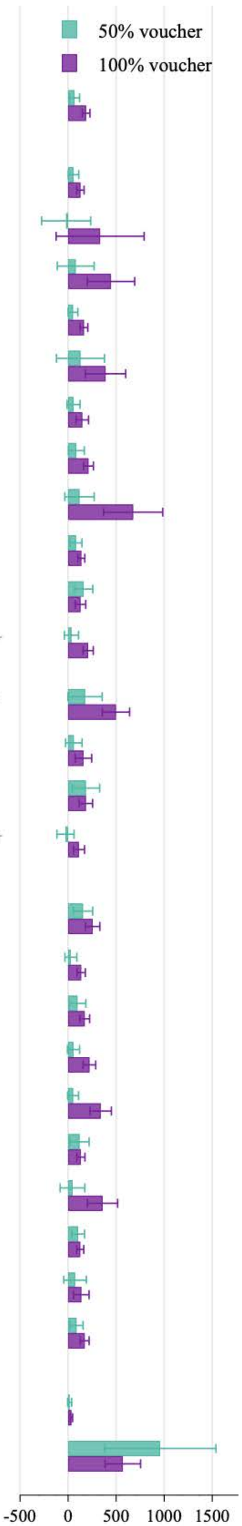
(a) PPMI charges in dollars

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on any birth control purchase	0.52	110	0.18	0.08	0.48	150	0.19	0.07
		(6)	(0.04)	(0.03)		(9)	(0.03)	(0.02)
A. Pre-specified demographic groups								
Non-Hispanic White	0.53	120	0.19	0.07	0.52	168	0.17	0.05
		(8)	(0.04)	(0.03)		(11)	(0.04)	(0.02)
Non-Hispanic Black	0.45	80	0.14	0.05	0.35	64	0.21	0.37
		(21)	(0.12)	(0.12)		(15)	(0.09)	(0.12)
Hispanic any race	0.60	81	0.10	0.21	0.55	132	0.18	0.03
		(13)	(0.11)	(0.11)		(27)	(0.10)	(0.05)
Women without children	0.55	118	0.18	0.07	0.47	156	0.20	0.07
		(7)	(0.04)	(0.03)		(10)	(0.03)	(0.02)
Mothers	0.40	60	0.16	0.18	0.55	125	0.13	0.07
		(11)	(0.10)	(0.09)		(22)	(0.09)	(0.05)
Age<26	0.58	107	0.13	0.05	0.49	129	0.18	0.08
		(9)	(0.05)	(0.04)		(11)	(0.04)	(0.03)
Age≥26	0.45	110	0.24	0.13	0.48	169	0.21	0.07
		(10)	(0.06)	(0.04)		(13)	(0.05)	(0.02)
Below associate's degree	0.53	107	0.19	0.05	0.53	112	0.12	0.05
		(14)	(0.09)	(0.06)		(16)	(0.08)	(0.05)
Associate's degree or higher	0.52	117	0.20	0.09	0.49	176	0.25	0.08
		(8)	(0.04)	(0.03)		(11)	(0.04)	(0.02)
Married or cohabitating	0.50	129	0.28	0.11	0.57	162	0.20	0.04
		(11)	(0.06)	(0.03)		(14)	(0.05)	(0.03)
Single	0.54	104	0.15	0.07	0.45	165	0.23	0.08
		(9)	(0.05)	(0.04)		(14)	(0.04)	(0.02)
Fee scale								
101-150% FPL	0.62	61	0.11	0.22	0.62	75	0.08	0.15
		(4)	(0.05)	(0.08)		(5)	(0.05)	(0.06)
151-200% FPL	0.54	109	0.15	0.15	0.47	152	0.21	0.15
		(11)	(0.08)	(0.06)		(14)	(0.06)	(0.03)
201-250% FPL	0.45	190	0.30	0.15	0.36	259	0.32	0.12
		(24)	(0.11)	(0.04)		(32)	(0.08)	(0.02)
250+% FPL	0.22	203	0.35	0.19	0.24	271	0.41	0.14
		(30)	(0.10)	(0.03)		(39)	(0.08)	(0.02)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	0.54	115	0.19	0.07	0.53	180	0.22	0.05
		(9)	(0.05)	(0.04)		(15)	(0.05)	(0.02)
Do not have a usual place for birth control	0.48	117	0.23	0.12	0.45	146	0.23	0.09
		(12)	(0.07)	(0.05)		(13)	(0.05)	(0.03)
Using Tier 1 or 2 method	0.51	121	0.20	0.09	0.50	164	0.23	0.07
		(9)	(0.05)	(0.03)		(13)	(0.04)	(0.02)
Not using Tier 1 or 2 method	0.51	99	0.17	0.09	0.47	141	0.15	0.07
		(10)	(0.06)	(0.04)		(13)	(0.05)	(0.03)
Delayed getting birth control	0.59	138	0.32	0.09	0.47	183	0.27	0.07
		(15)	(0.07)	(0.04)		(18)	(0.06)	(0.03)
Did not delay getting birth control	0.49	104	0.18	0.09	0.51	155	0.20	0.07
		(8)	(0.05)	(0.04)		(12)	(0.04)	(0.02)
Positive desire to have a baby	0.53	92	0.12	0.08	0.46	126	0.06	0.05
		(13)	(0.09)	(0.08)		(20)	(0.08)	(0.04)
Negative desire to have a baby	0.51	121	0.22	0.09	0.52	175	0.24	0.07
		(9)	(0.05)	(0.03)		(12)	(0.04)	(0.02)
More likely to meet career aspirations	0.50	122	0.26	0.05	0.47	145	0.20	0.07
		(15)	(0.08)	(0.05)		(17)	(0.06)	(0.03)
Less likely to meet career aspirations	0.53	112	0.19	0.10	0.51	165	0.22	0.07
		(8)	(0.05)	(0.03)		(11)	(0.04)	(0.02)
C. Exploratory								
Planning to get a LARC before appointment	0.64	132	0.10	0.08	0.60	341	0.25	0.06
		(18)	(0.10)	(0.05)		(30)	(0.08)	(0.02)
Not planing to get a LARC before appointment	0.50	104	0.19	0.08	0.47	119	0.18	0.08
		(7)	(0.04)	(0.03)		(8)	(0.03)	(0.02)



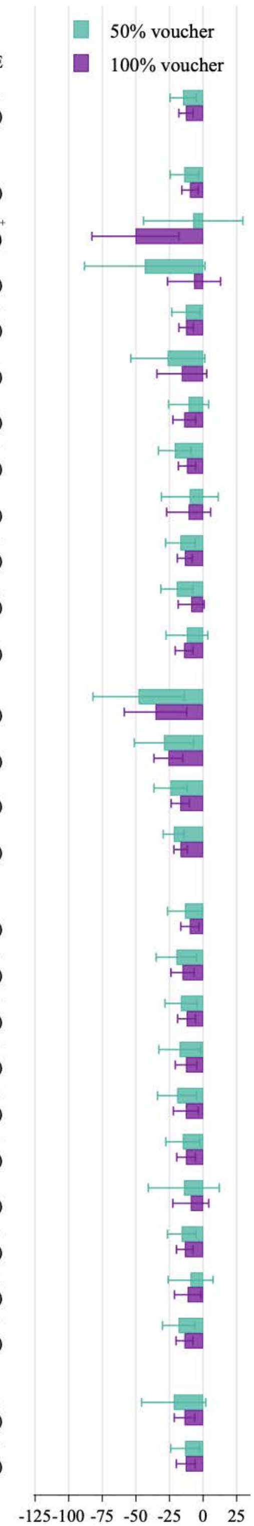
(b) Any birth control purchase

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on larc insertion	0.07	110 (6)	0.05 (0.02)	0.05 (0.02)	0.04	150 (9)	0.14 +++ (0.02)	0.08 (0.01)
A. Pre-specified demographic groups								
Non-Hispanic White	0.08	120 (8)	0.06 (0.03)	0.04 (0.02)	0.06	168 (11)	0.15 ++ (0.03)	0.07 (0.01)
Non-Hispanic Black	0.03	80 (21)	0.01 (0.04)	-0.01 (0.03)	0.02	64 (15)	0.01 (0.04)	0.06 (0.04)
Hispanic any race	0.10	81 (13)	0.03 (0.07)	0.08 (0.09)	0.02	132 (27)	0.21 + (0.06)	0.10 (0.03)
Women without children	0.08	118 (7)	0.05 (0.03)	0.04 (0.02)	0.05	156 (10)	0.13 ++ (0.02)	0.07 (0.01)
Mothers	0.04	60 (11)	0.07 (0.06)	0.05 (0.05)	0.03	125 (22)	0.19 (0.05)	0.12 (0.03)
Age<26	0.08	107 (9)	0.05 (0.03)	0.04 (0.03)	0.04	129 (11)	0.12 + (0.03)	0.07 (0.01)
Age≥26	0.06	110 (10)	0.06 (0.03)	0.05 (0.03)	0.04	169 (13)	0.16 ++ (0.03)	0.09 (0.01)
Below associate's degree	0.04	107 (14)	0.11 (0.05)	0.05 (0.03)	0.01	112 (16)	0.13 (0.04)	0.10 (0.02)
Associate's degree or higher	0.07	117 (8)	0.06 (0.03)	0.06 (0.02)	0.05	176 (11)	0.15 ++ (0.03)	0.07 (0.01)
Married or cohabitating	0.05	129 (11)	0.14 (0.04)	0.08 (0.03)	0.05	162 (14)	0.13 (0.04)	0.07 (0.01)
Single	0.07	104 (9)	0.03 (0.03)	0.02 (0.03)	0.04	165 (14)	0.16 +++ (0.03)	0.09 ++ (0.01)
Fee scale								
101-150% FPL	0.05	61 (4)	0.05 (0.03)	0.09 (0.05)	0.04	75 (5)	0.14 ++ (0.03)	0.20 + (0.03)
151-200% FPL	0.11	109 (11)	0.05 (0.05)	0.06 (0.05)	0.05	152 (14)	0.11 (0.04)	0.07 (0.02)
201-250% FPL	0.05	190 (24)	0.14 (0.07)	0.09 (0.04)	0.05	259 (32)	0.22 (0.06)	0.08 (0.02)
250+% FPL	0.07	203 (30)	-0.02 (0.05)	-0.02 (0.03)	0.04	271 (39)	0.12 ++ (0.05)	0.05 ++ (0.01)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	0.05	115 (9)	0.10 (0.03)	0.07 (0.02)	0.03	180 (15)	0.17 + (0.03)	0.08 (0.01)
Do not have a usual place for birth control	0.11	117 (12)	0.04 (0.05)	0.03 (0.03)	0.07	146 (13)	0.12 (0.03)	0.09 (0.02)
Using Tier 1 or 2 method	0.05	121 (9)	0.05 (0.03)	0.04 (0.02)	0.04	164 (13)	0.13 ++ (0.03)	0.07 (0.01)
Not using Tier 1 or 2 method	0.10	99 (10)	0.06 (0.04)	0.06 (0.03)	0.04	141 (13)	0.16 + (0.03)	0.09 (0.01)
Delayed getting birth control	0.13	138 (15)	0.11 (0.06)	0.07 (0.04)	0.03	183 (18)	0.25 + (0.04)	0.11 (0.02)
Did not delay getting birth control	0.04	104 (8)	0.07 (0.03)	0.05 (0.02)	0.05	155 (12)	0.11 (0.03)	0.07 (0.01)
Positive desire to have a baby	0.05	92 (13)	0.00 (0.04)	0.02 (0.03)	0.02	126 (20)	0.09 (0.04)	0.08 (0.02)
Negative desire to have a baby	0.07	121 (9)	0.10 (0.03)	0.07 (0.02)	0.06	175 (12)	0.15 (0.03)	0.07 (0.01)
More likely to meet career aspirations	0.06	122 (15)	0.05 (0.05)	0.05 (0.04)	0.05	145 (17)	0.11 (0.04)	0.06 (0.02)
Less likely to meet career aspirations	0.07	112 (8)	0.07 (0.03)	0.06 (0.02)	0.05	165 (11)	0.17 ++ (0.03)	0.08 (0.01)
C. Exploratory								
Planning to get a LARC before appointment	0.50	132 (18)	0.06 (0.11)	0.08 (0.05)	0.29	341 (30)	0.46 +++ (0.08)	0.10 (0.02)
Not planing to get a LARC before appointment	0.00	104 (7)	0.04 (0.01)	0.03 (0.01)	0.01	119 (8)	0.07 (0.01)	0.06 (0.01)



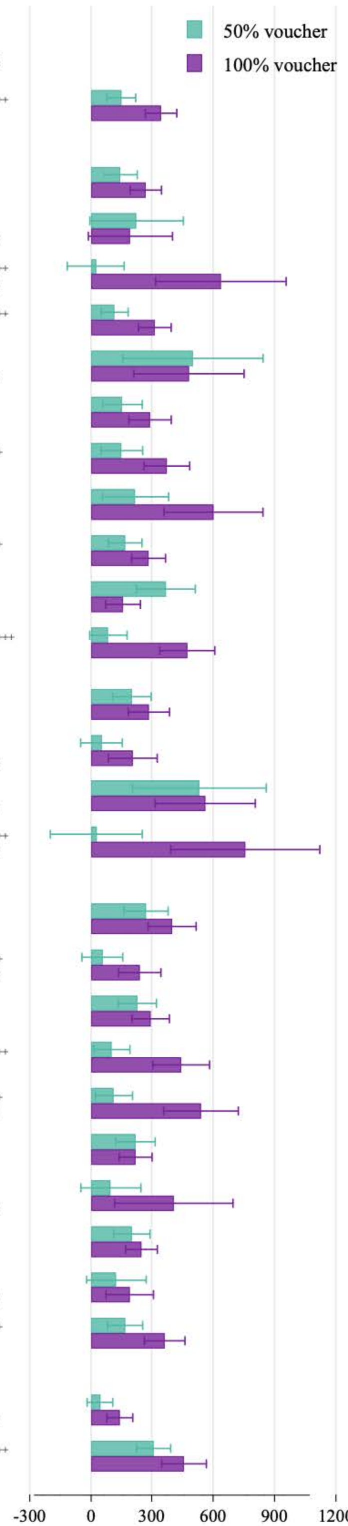
(c) LARC insertion

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on expected annual pregnancies	0.52	110 (6)	-0.17 (0.03)	-0.08 (0.03)	0.55	150 (9)	-0.19 (0.03)	-0.07 (0.01)
A. Pre-specified demographic groups								
Non-Hispanic White	0.51	120 (8)	-0.18 (0.04)	-0.07 (0.03)	0.52	168 (11)	-0.17 (0.03)	-0.05 (0.02)
Non-Hispanic Black	0.58	80 (21)	-0.13 (0.11)	-0.04 (0.11)	0.68	64 (15)	-0.20 (0.09)	-0.34+ (0.11)
Hispanic any race	0.45	81 (13)	-0.09 (0.10)	-0.19 (0.10)	0.50	132 (27)	-0.18 (0.09)	-0.03 (0.05)
Women without children	0.49	118 (7)	-0.17 (0.04)	-0.06 (0.03)	0.56	156 (10)	-0.20 (0.03)	-0.07 (0.02)
Mothers	0.63	60 (11)	-0.15 (0.10)	-0.16 (0.09)	0.50	125 (22)	-0.14 (0.08)	-0.08 (0.05)
Age<26	0.46	107 (9)	-0.12 (0.05)	-0.05 (0.04)	0.55	129 (11)	-0.17 (0.04)	-0.08 (0.02)
Age≥26	0.58	110 (10)	-0.23 (0.05)	-0.12 (0.04)	0.56	169 (13)	-0.20 (0.04)	-0.07 (0.02)
Below associate's degree	0.51	107 (14)	-0.18 (0.08)	-0.05 (0.05)	0.51	112 (16)	-0.12 (0.07)	-0.05 (0.04)
Associate's degree or higher	0.52	117 (8)	-0.18 (0.04)	-0.09 (0.03)	0.55	176 (11)	-0.24 (0.04)	-0.07 (0.02)
Married or cohabitating	0.53	129 (11)	-0.26 (0.06)	-0.10 (0.03)	0.48	162 (14)	-0.20 (0.05)	-0.04 (0.02)
Single	0.50	104 (9)	-0.14 (0.05)	-0.06 (0.04)	0.58	165 (14)	-0.22 (0.04)	-0.08 (0.02)
Fee scale								
101-150% FPL	0.43	61 (4)	-0.11 (0.05)	-0.21 (0.07)	0.43	75 (5)	-0.09 (0.04)	-0.15 (0.05)
151-200% FPL	0.50	109 (11)	-0.15 (0.07)	-0.15 (0.06)	0.57	152 (14)	-0.20 (0.06)	-0.15 (0.03)
201-250% FPL	0.58	190 (24)	-0.28 (0.10)	-0.14 (0.04)	0.66	259 (32)	-0.31 (0.08)	-0.11 (0.02)
250+% FPL	0.79	203 (30)	-0.32 (0.09)	-0.17 (0.03)	0.78	271 (39)	-0.38 (0.07)	-0.13 (0.02)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	0.50	115 (9)	-0.18 (0.05)	-0.07 (0.03)	0.51	180 (15)	-0.21 (0.04)	-0.05 (0.02)
Do not have a usual place for birth control	0.55	117 (12)	-0.21 (0.06)	-0.11 (0.04)	0.58	146 (13)	-0.22 (0.05)	-0.09 (0.03)
Using Tier 1 or 2 method	0.53	121 (9)	-0.19 (0.05)	-0.09 (0.03)	0.54	164 (13)	-0.22 (0.04)	-0.07 (0.02)
Not using Tier 1 or 2 method	0.52	99 (10)	-0.16 (0.05)	-0.09 (0.04)	0.57	141 (13)	-0.15 (0.04)	-0.07 (0.02)
Delayed getting birth control	0.45	138 (15)	-0.31 (0.06)	-0.09 (0.03)	0.56	183 (18)	-0.26 (0.06)	-0.07 (0.03)
Did not delay getting birth control	0.54	104 (8)	-0.17 (0.05)	-0.08 (0.03)	0.53	155 (12)	-0.19 (0.04)	-0.07 (0.02)
Positive desire to have a baby	0.52	92 (13)	-0.11 (0.08)	-0.07 (0.07)	0.58	126 (20)	-0.07 (0.07)	-0.05 (0.04)
Negative desire to have a baby	0.52	121 (9)	-0.21 (0.04)	-0.08 (0.03)	0.52	175 (12)	-0.23 (0.04)	-0.07 (0.02)
More likely to meet career aspirations	0.54	122 (15)	-0.25 (0.07)	-0.05 (0.05)	0.56	145 (17)	-0.19 (0.06)	-0.06 (0.03)
Less likely to meet career aspirations	0.51	112 (8)	-0.18 (0.04)	-0.09 (0.03)	0.53	165 (11)	-0.22 (0.04)	-0.07 (0.02)
C. Exploratory								
Planning to get a LARC before appointment	0.37	132 (18)	-0.09 (0.10)	-0.08 (0.05)	0.42	341 (30)	-0.27 (0.08)	-0.06 (0.02)
Not planing to get a LARC before appointment	0.54	104 (7)	-0.18 (0.04)	-0.07 (0.03)	0.57	119 (8)	-0.17 (0.03)	-0.07 (0.02)



(d) Expected annual pregnancies

	50% Voucher Group				100% Voucher Group			
	Control mean	First stage	ITT effect	LATE	Control mean	First stage	ITT effect	LATE
Overall effect on temporal coverage	184	0.72 (0.03)	186 (48)	280 (66)	145	0.66 (0.02)	328 ++ (41)	503 ++ (57)
A. Pre-specified demographic groups								
Non-Hispanic White	202	0.73 (0.03)	210 (61)	296 (85)	174	0.69 (0.03)	338 (53)	473 (68)
Non-Hispanic Black	85	0.60 (0.09)	100 (72)	191 (100)	95	0.49 (0.07)	56 (88)	184 (101)
Hispanic any race	250	0.72 (0.08)	56 (134)	61 (180)	92	0.63 (0.07)	444 ++ (121)	585 ++ (149)
Women without children	206	0.74 (0.03)	168 (53)	241 (71)	150	0.66 (0.02)	315 ++ (44)	475 ++ (61)
Mothers	83	0.62 (0.07)	254 (125)	417 (145)	116	0.70 (0.06)	432 (107)	561 (160)
Age<26	176	0.71 (0.04)	187 (58)	272 (89)	141	0.63 (0.03)	291 (54)	411 (74)
Age≥26	194	0.72 (0.04)	182 (78)	295 (103)	152	0.69 (0.03)	366 + (63)	568 + (86)
Below associate's degree	145	0.77 (0.06)	227 (106)	318 (122)	75	0.63 (0.05)	264 (64)	450 (92)
Associate's degree or higher	180	0.74 (0.03)	212 (55)	308 (75)	175	0.73 (0.03)	355 + (55)	498 + (74)
Married or cohabitating	127	0.80 (0.04)	377 (77)	470 (93)	210	0.75 (0.03)	249 (73)	337 (91)
Single	206	0.70 (0.04)	114 (65)	178 (98)	125	0.69 (0.03)	397 +++ (61)	591 +++ (85)
Fee scale								
101-150% FPL	140	0.75 (0.04)	215 (57)	284 (67)	152	0.66 (0.03)	295 (57)	437 (78)
151-200% FPL	288	0.71 (0.05)	71 (114)	151 (153)	170	0.69 (0.04)	243 (76)	353 (103)
201-250% FPL	130	0.76 (0.07)	413 (137)	692 (216)	120	0.68 (0.06)	474 (129)	674 (150)
250+% FPL	187	0.59 (0.08)	38 (119)	51 (217)	107	0.62 (0.06)	442 ++ (127)	807 ++ (198)
B. Pre-specified pre-randomization categories								
Have a usual place for birth control	136	0.75 (0.03)	271 (60)	370 (75)	128	0.74 (0.03)	376 (59)	511 (76)
Do not have a usual place for birth control	268	0.71 (0.05)	113 (95)	152 (140)	190	0.67 (0.04)	281 (73)	460 + (100)
Using Tier 1 or 2 method	129	0.74 (0.03)	222 (50)	297 (61)	150	0.70 (0.03)	317 (54)	442 (70)
Not using Tier 1 or 2 method	264	0.70 (0.04)	145 (88)	275 (121)	136	0.62 (0.03)	356 + (65)	605 ++ (96)
Delayed getting birth control	314	0.90 (0.04)	322 (133)	357 (148)	127	0.77 (0.04)	527 (97)	688 + (118)
Did not delay getting birth control	126	0.68 (0.03)	194 (46)	279 (62)	167	0.68 (0.03)	248 (50)	369 (68)
Positive desire to have a baby	175	0.68 (0.06)	68 (98)	172 (133)	116	0.56 (0.05)	250 (89)	476 (172)
Negative desire to have a baby	180	0.75 (0.03)	256 (63)	366 (82)	189	0.75 (0.02)	345 (58)	472 (75)
More likely to meet career aspirations	158	0.80 (0.05)	142 (92)	199 (119)	179	0.66 (0.04)	244 (85)	345 (105)
Less likely to meet career aspirations	188	0.73 (0.03)	221 (61)	321 (82)	147	0.71 (0.03)	376 + (55)	534 + (74)
C. Exploratory								
Planning to get a LARC before appointment	911	0.77 (0.06)	289 (253)	411 (301)	644	0.83 (0.05)	839 + (193)	930 (208)
Not planing to get a LARC before appointment	74	0.70 (0.03)	156 (25)	228 (31)	78	0.63 (0.02)	212 (29)	357 ++ (43)



(e) Days covered by purchased contraception

BIBLIOGRAPHY

Atila Abdulkadiroğlu, Nikhil Agarwal, and Parag A. Pathak. The welfare effects of coordinated assignment: Evidence from the New York city high school match. *American Economic Review*, 107(12):3635–3689, 2017. ISSN 00028282. doi: 10.1257/aer.20151425.

Nikhil Agarwal and Paulo Somaini. Demand Analysis Using Strategic Reports: An Application to a School Choice Mechanism. *Econometrica*, 86(2):391–444, 2018. doi: 10.3982/ECTA13615. URL https://doi.org/10.3982/ECTA13615.

Nikhil Agarwal and Paulo Somaini. Revealed preference analysis of school choice models. *Annual Review of Economics*, 12:471–501, 2020. ISSN 19411391. doi: 10.1146/annurev-economics-082019-112339.

Ruchir Agarwal and Patrick Gaulé. Invisible Geniuses: Could the Knowledge Frontier Advance Faster? *SSRN Electronic Journal*, 2019. doi: 10.2139/ssrn.3325595.

Kehinde Ajayi and Modibo Sidibe. School Choice Under Imperfect Information. *SSRN Electronic Journal*, 2020. ISSN 1556-5068. doi: 10.2139/ssrn.3524535.

Mohammad Akbarpour, Adam Kapor, Christopher Neilson, Winnie Van Dijk, and Seth Zimmerman. Centralized School Choice with Unequal Outside. (December), 2021.

Elizabeth Oltmans Ananat and Daniel Hungerman. The power of the pill for the next generation: Oral contraception's effects on fertility, abortion, and maternal and child characteristics. *Review of Economics and Statistics*, 94(1):37–51, 2012.

Tommy Andersson, Umut Dur, Sinan Ertemel, and Onur Kesten. Sequential School Choice with Public and Private Schools. *Working Papers*, (2018:39):1–40, 2019.

Dan Ariely and Klaus Wertenbroch. Procrastination, deadlines, and performance: Self-control by precommitment. *Psychological Science*, 13(3):219–224, 2002. doi: 10.1111/1467-9280.00441.

Katherine Baicker, Sendhil Mullainathan, and Joshua Schwartzstein. Behavioral hazard in health insurance. *The Quarterly Journal of Economics*, 130(4):1623–1667, 2015. doi: 10.1093/qje/qjv029.

Martha J. Bailey. More power to the pill: The impact of contraceptive freedom on women's lifecycle labor supply. *Quarterly Journal of Economics*, 121(1):289–320, 2006.

Martha J. Bailey. Reexamining the impact of u.s. family planning programs on fertility: Evidence from the war on poverty and the early years of title x. *American Economic Journal: Applied Economics*, 4(2):62–97, 2012.

Martha J. Bailey. Fifty years of family planning: New evidence on the long-run effects of increasing access to contraception. NBER Working Paper 19493, 2013.

Martha J. Bailey and Jason M. Lindo. Access and use of contraception and its effects on women's outcomes in the u.s. 2018.

Martha J. Bailey, Brad J. Hershbein, and Amalia R. Miller. The opt-in revolution? contraception and the gender gap in wages. *American Economic Journal: Applied Economics*, 4(3):225–254, 2012.

Martha J. Bailey, Olga Malkova, and Zoë McLaren. Do family planning programs increase children's opportunities? evidence from the war on poverty and the early years of title x. *Journal of Human Resources*, 2018.

Martha J. Bailey, Lea Bart, and Vanessa Wanner Lang. The missing baby bust: The consequences of the covid-19 pandemic for contraceptive use, pregnancy, and childbirth among low-income women. *Population Research and Policy Review*, 41(4):1549–1569, 2022. doi: 10.1007/s11113-022-09703-9.

Gary S. Becker. *An Economic Analysis of Fertility*. Princeton University Press, Princeton, N.J., 1960.

Gary S. Becker. A theory of the allocation of time. *Economic Journal*, 75(299):493–517, 1965.

Gary S. Becker. *The Demand for Children*. Harvard University Press, Cambridge, MA, 1981.

Gary S. Becker and H. Gregg Lewis. On the interaction between the quantity and quality of children. *Journal of Political Economy*, 81(2):S279–S288, 1973.

Nora Becker. The impact of insurance coverage on utilization of prescription contraceptives: Evidence from the affordable care act. *Journal of Policy Analysis and Management*, 37 (3):571–601, 2018.

Natalia E. Birgisson, Qiuhong Zhao, Gina M. Secura, Tessa Madden, and Jeffrey F. Peipert. Preventing unintended pregnancy: The contraceptive choice project in review. *Journal of Women's Health*, 24(5):349–353, 2015.

Judith Blake. Population policy for americans: Is the government being misled. *Science*, 164(3879):522–529, May 1969.

Hilary O. Broughton, Christina M. Buckel, Karen J. Omvig, Jennifer L. Mullersman, Jeffrey F. Peipert, and Gina M. Secura. From research to practice: dissemination of the contraceptive choice project. *Translational Behavioral Medicine*, 1(9):1–9, 2016. doi: 10.1007/s13142-016-0404-x.

Thomas Buser and Huaiping Yuan. Do Women Give Up Competing More Easily? Evidence from the Lab and the Dutch Math Olympiad. *SSRN Electronic Journal*, 2016. doi: 10.2139/ssrn.2867346.

Caterina Calsamiglia, Chao Fu, and Maia Güell. Structural estimation of a model of school choices: The boston mechanism versus its alternatives. *Journal of Political Economy*, 128 (2):642–680, 2020. ISSN 1537534X. doi: 10.1086/704573.

CS Carlin, AR Fertig, and BE Dowd. Affordable care act's mandate eliminating contraceptive cost sharing influenced choices of women with employer coverage. *Health Affairs*, 35(9): 1608–1615, 2016.

Hector Chade and Lones Smith. Simultaneous search. *Econometrica*, 74(5):1293–1307, 2006. doi: 10.1111/j.1468-0262.2006.00705.x.

Yan Chen and Onur Kesten. Chinese college admissions and school choice reforms: A theoretical analysis. *Journal of Political Economy*, 125(1):99–139, 2017. ISSN 1537534X. doi: 10.1086/689773.

Vanessa K. Dalton, Michelle H. Moniz, Martha J. Bailey, Lindsay K. Admon, Giselle E. Kolenic, Anca Tilea, and A. Mark Fendrick. The impact of eliminating out-of-pocket costs for contraception on births following the affordable care act. *Journal of the American Medical Association Network Open*, 2020.

Jeff Diamant and Besheer Mohamed. What the data says about abortion in the u.s. *Pew Research Center*, 2023.

Alba DiCenso, Gordon Guyatt, Andrew Willan, and Lauren Griffith. Interventions to reduce unintended pregnancies among adolescents: Systematic review of randomized controlled trials. *British Medical Journal*, 324(7351):1426–1434, 2002.

Richard A. Easterlin. *Birth and Fortune: The Impact of Numbers on Personal Welfare.* Basic Books, New York, 1980.

Glenn Ellison and Ashley Swanson. Do Schools Matter for High Math Achievement? Evidence from the American Mathematics Competitions †. *American Economic Review*, 106 (6):1244–1277, 2016. doi: 10.1257/aer.20140308. URL http://dx.doi.org/10.1257/aer.20140308.

Gabrielle Fack, Julien Grenet, and Yinghua He. Beyond truth-telling: Preference estimation with centralized school choice and college admissions. *American Economic Review*, 109 (4):1486–1529, 2019. ISSN 19447981. doi: 10.1257/aer.20151422.

C.I. Fowler, J. Gable, J. Wang, B. Lasater, and E. Wilson. Family planning annual report: 2018 national summary, 2019.

J. Hainmueller. Entropy balancing for causal effects: A multivariate reweighting model to produce balanced samples in observational studies. *Political Analysis*, 20:25–46, 2012.

Oscar Harkavy, Frederick S. Jaffe, and Samuel M. Wishik. Family planning and public policy: Who is misleading whom? *Science*, 165(3891):367–373, 1969.

E Heisel, GE Kolenic, MM Moniz, et al. Intrauterine device insertion before and after mandated health care coverage: The importance of baseline costs. *Obstetric Gynecology*, 131(5):843–849, 2018.

Ryan Hill and Carolyn Stein. Race to the Bottom: Competition and Quality in Science. *Working paper*, 1122374(1122374), 2020.

P. J. Huber. The behavior of maximum likelihood estimates under nonstandard conditions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, pages 221–233, Berkeley, CA, 1967. University of California Press.

Clemence Idoux. Integrating New York City Schools: The Role of Admissions Criteria and Family Preferences. 2022.

Guido W. Imbens and Joshua D. Angrist. Identification and estimation of local average treatment effects. *Econometrica*, 62(2):467–475, 1994. doi: 10.2307/2951620.

Adam Kapor, Mohit Karnani, and Christopher Neilson. Aftermarket Frictions and the Cost of Off-Platform Options in Centralized Assignment Mechanisms. Technical report, 2022.

Adam J. Kapor, Christopher A. Neilson, and Seth D. Zimmerman. Heterogeneous beliefs and school choice mechanisms†. *American Economic Review*, 110(5):1274–1315, 2020. ISSN 19447981. doi: 10.1257/aer.20170129.

Melissa S. Kearney and Phillip Levine. Subsidized contraception, fertility, and sexual behavior. *Review of Economics and Statistics*, 91(1):137–151, 2009.

Douglas Kirby. *No Easy Answers: Research Findings on Programs to Reduce Teen Pregnancy.* The National Campaign to Prevent Teen Pregnancy, Washington, D.C., 1997.

Jeffrey R. Kling, Jeffrey B. Liebman, and Lawrence F. Katz. Experimental analysis of neighborhood effects. *Econometrica*, 75(1):83–119, 2007.

James C. Knowles, John S. Akin, and David K. Guilkey. The impact of population policies: Comment. *Population and Development Review*, 20(3):611–615, 1994.

Kathryn Kost and Laura D. Lindberg. Pregnancy intentions, maternal behaviors, and infant health: Investigating relationships with new measures and propensity score analysis. *Demography*, 52(1):83–111, 2015. doi: 10.1007/s13524-014-0359-9.

Kathryn Kost, Mia Zolna, and Rachel Murro. Pregnancies in the united states by desire for pregnancy: Estimates for 2009, 2011, 2013, and 2015, 2023.

Tomás Larroucau and Ignacio Rios. Do "Short-List" Students Report Truthfully? Strategic Behavior in the Chilean College Admissions Problem. (2018):1–65, 2020.

Jason M. Lindo and Analisa Packham. How much can expanding access to long-acting reversible contraceptives reduce teen birth rates. *American Economic Journal: Economic Policy*, 9(3):348–376, 2017.

Margaux Luflade. The Value of Information in Centralized School Choice Systems. *Working paper*, (August):1–72, 2017. ISSN 1932-6203. URL https://sites.duke.edu/margauxluflade/files/2017/11/LUFLADE_JobMarketPaper.pdf.

A. McCann, A. Schoenfeld Walker, A. Sasani, T. Johnston, L. Buchanan, and J. Huang. Tracking the states where abortion is now banned. *The New York Times*, 2023. Updated June 19, 2023, originally published October 13, 2022.

Colleen McNicholas, Tessa Madden, Gina Secura, and Jeffrey F. Peipert. The contraceptive choice project round up: What we did and what we learned. *Clinical Obstetrics and Gynecology*, 57(4):635–643, 2014.

Renee Mestad, Gina Secura, Jenifer Allsworth, Tessa Madden, Qiuhong Zhao, and Jeffrey Peipert. Acceptance of long-acting reversible contraceptive methods by adolescent participants in the contraceptive choice project. *Contraception*, 84(5):493–498, 2011.

Robert T. Michael and Robert J. Willis. *Contraception and Fertility: Household Production under Uncertainty*. National Bureau of Economic Research, Cambridge, MA, 1976.

Sarah Miller, Laura R. Wherry, and Diana Greene Foster. The economic consequences of being denied an abortion. *National Bureau of Economic Research Working Paper No. 26662*, 2020.

A. P. Mohllajee, K. M. Curtis, B. Morrow, and P. A. Marchbanks. Pregnancy intention and its relationship to birth and maternal outcomes. *Obstetrics Gynecology*, 109(3), 2007.

Diana Moreira. Success Spills Over. 2019.

Ted O'Donoghue and Matthew Rabin. Doing it now or later. *American Economic Review*, 89(1):103–124, 1999. doi: 10.1257/aer.89.1.103.

Sebastian Otero, Nano Barahona, and Caue Dobbin. Affirmative Action in Centralized College Admission Systems: Evidence from Brazil. 2021.

Analisa Packham. Family planning funding cuts and teen childbearing. *Journal of Health Economics*, 55:168–185, 2017.

Lant H. Pritchett. Desired fertility and the impact of population policies. *Population and Development Review*, 20(1):1–55, 1994a.

Lant H. Pritchett. The impact of population policies: Reply. *Population and Development Review*, 20(3):621–630, 1994b.

Norman B. Ryder and Charles F. Westoff. *The Contraceptive Revolution*. Princeton University Press, Princeton, 1971.

Gina M. Secura, Jenifer E. Allsworth, Tessa Madden, Jennifer Mullersman, and Jeffrey F. Peipert. The contraceptive choice project: reducing barriers to long-acting reversible contraception. *American Journal of Obstetrics and Gynecology*, 203(2):115.e1–115.e7, 2010.

Mindel C. Sheps. On the time required for conception. *Population Studies*, 19(3):85–97, 1964.

Mindel C. Sheps and Edward B. Perrin. Further results from a human fertility model with a variety of pregnancy outcomes. *Human Biology*, 38(3):189–193, 1966.

Adam Sonfield, Kathryn Kost, Rachel Benson Gold, and Lawrence B. Finer. The public costs of births resulting from unintended pregnancies: national and state-level estimates. *Perspectives on Sexual and Reproductive Health*, 43(2):94–102, 2011.

Daniel Waldinger. Targeting in-kind transfers through market design: A revealed preference analysis of public housing allocation. *American Economic Review*, 111(8):2660–2696, 2021. ISSN 19447981. doi: 10.1257/aer.20190516.

H. White. A heteroskedasticity-consistent covariance matrix estimator and a direct test for heteroskedasticity. *Econometrica*, 48:817–830, 1980.

Robert J. Willis. A new approach to the economic theory of fertility behavior. *Journal of Political Economy*, 81(2):S14–S64, 1973.