Undermatching in New York City School Choice: Application Behavior and Feedback

Kenny Peng*[1,2], Emily Ryu*[1], Jon Kleinberg[1], Eva Tardos[1], and Nikhil Garg[1,2]

[1]Cornell University
[2]Cornell Tech

August 13, 2025

Abstract

In centralized public school matching, students submit ranked lists over options. In New York City, for example, eighth graders select from over 800 programs. Prior work documents disparities in how students apply and where they match, and has designed informational interventions in response. However, the relationship between application decisions (the full list of schools a student submits) and match outcomes is not fully understood, limiting the design of interventions. In this work, we contribute to this understanding by quantifying undermatching using counterfactuals of the 2022 New York City high school match, and by analyzing full student lists according to a theoretical model of portfolio decision-making. First, we quantify *undermatching*, the difference in quality between the program a student matched with and the best program a student *could have* matched to had they applied. Black or Hispanic students undermatched 1.5× as much than their Asian or white peers, and 2× as much when comparing the most competitive students. Disparities also exist between income groups, neighborhoods, and at-home languages, but we do not find significant differences by sex. These disparities are larger than those induced by admissions policies and geography alone. Second, we show that interpretable deviations from a theoretically-optimal reach-match-safety strategy explain heterogeneity in undermatching, unlike demographics alone (increasing explained variance from 4% to 41%). Finally, we study simple personalized feedback derived from our model. Changing a single program in a student's list reduces undermatching by 24% on average, while changing the full list reduces undermatching by 57%. In equilibrium, these recommendations reduce disparities in match outcomes, but also increase the number of students who do not match, suggesting a need for further incorporating congestion effects in designing recommendations.

1 Introduction

Each year, more than 70,000 eighth graders in New York City (NYC) choose which of over 800 high school programs to apply to as part of a centralized matching process. Similar systems are used in many other large school districts in the United States and globally. A student's school placement is important for downstream outcomes [2, 13, 15, 21, 32]. In NYC schools, there is substantial socioeconomic and demographic segregation [66, 74]. In the 2022-2023 general education match, our data shows that compared to the median Asian or white student, the median Black or Hispanic student matched to a program with 1.95× as many students

{klp98,er442,kleinberg,eva.tardos,ngarg}@cornell.edu. We thank Nageeb Ali, Sarah Cohodes, Viola Corradini, Clemence Idoux, Amélie Marian, Parag Pathak, Ran Shorrer, and Sabina Tomkins for helpful feedback and discussions. We also thank the New York City Department of Education's Enrollment Research and Policy team for invaluable discussions and for data access, especially Ben Cosman, Brielle McDaniel, and Lianna Wright. Finally, we thank collaborators and participants of internal seminars at Cornell for helpful questions and feedback. This research was supported by a Cornell Center for Social Sciences Grant and the center's Regulated Research Environment. NG is supported by NSF CAREER IIS-2339427 and a Cornell Tech Urban Tech Hub research award. ET was supported in part by AFOSR grant FA9550-23-1-0410, AFOSR grant FA9550-231-0068, and ONR MURI grant N000142412742. ER is supported by NSF Graduate Research Fellowship Program grant DGE-2139899.

who do not graduate, 1.43× as many students who do not enroll in college, and 0.3 standard deviations lower school impact.[1]

The application process is complex [56]. In theory, the deferred acceptance algorithm used by NYC and other large school districts [3] makes the process straightforward: with unlimited ranked lists, students can list programs in order of their true preference. In practice, this task is daunting. Forming preferences over 800 programs is hard. And when selecting only a subset of programs, students must be strategic: applying to only highly-selective programs, for example, leaves a student at risk of going unmatched. The NYC match has spawned dozens of expensive consulting services and Facebook groups where (some) parents discuss school options and admissions probabilities.

An important and natural concern is that this informational complexity exacerbates disparities in match outcomes in addition to those caused by geography and admissions criteria. In particular, there is a concern of *undermatching*: students, especially those already disadvantaged, may not apply to high quality schools that they could have matched to and are nearby. Indeed, substantial empirical work documents demographic disparities in how students apply. In NYC, research has found that students who are Asian, white, higher-income, or female apply to and match with higher quality schools compared to their peers, even when controlling for neighborhood and grades [26–28, 30]. Through a survey of NYC parents, Corradini and Idoux [30] show that this "choice gap" is explained substantially by differential awareness of high-performing schools.[2] Corradini [28] shows through administrative data that differences in application behavior cannot be explained by summary statistics of the set of schools to which a student has positive admissions probability. There are similar findings outside of the NYC context and in college admissions [16, 77]; for example Larroucau et al. [53], through surveys in Chile for college admissions, document and taxonomize types of application mistakes, including mistakes of awareness, valuations, and admission probabilities (causing not matching or undermatching). These results show that disparities in application behavior are not driven solely by heterogeneous preferences, differences in geography, or differences in admissions probability.

Motivated by such disparities, prior work has also designed and evaluated informational interventions, both experimentally [26, 27] and by analyzing policy changes [28] in NYC. Corcoran et al. [27], for example, provided students with lists of high-quality nearby schools [27]. In Chile, Larroucau et al. [53] implement and evaluate personalized application feedback, warning students if their probability matching to none of their choices is above 1% or if their probability of matching to their first choice is over 99%. The first warning encourages students to list more safety programs; the latter is to encourage a student to *reach* for higher quality but lower probability options as their first choice. Across these interventions, students change their behavior in response to information. Together, this past work has clearly established that disparities in application behavior are substantially driven by disparities in information, rather than only by differences in preferences, admissions probabilities, and geography. As a consequence, there has been continued interest in designing informational interventions to reduce such disparities.

The present work. However, despite this substantial work, an open question remains: *How do each student's ranking decisions empirically translate to undermatching?*. Answering this question informs *the design of interventions, both in broad policy and in how we can best build fine-grained personalized interventions.*

For an individual student, we would like to understand how much better they could have applied, how exactly they applied suboptimally, and how much behavioral changes would have changed outcomes. For example, are there students who do not apply to any "match" programs, only applying to schools that are either too competitive or not high-quality enough? Do these individual behaviors explain match outcomes and undermatching in particular? Could students have made alternative decisions (using only knowledge at the time of the application) to improve their outcomes? Empirically, would reordering programs, adding new programs to the top of one's list, or adding safety programs most improve one's match? Then, in the aggregate, we would like to understand the most common ways students undermatch, and how this varies across and within groups.

Understanding how precisely student application lists translates to undermatching can help guide the design of interventions. Consider the current landscape of interventions. New York City provides general advice to students apply to many programs and also warns students who are at risk of not matching. There

[1] a measure provided by the city that aims to be causal

[2] The gap is also partially explained by preferences over the school's racial composition, but not by family preferences over other school characteristics, beliefs of admissions chances, or knowledge of school quality conditional on awareness.

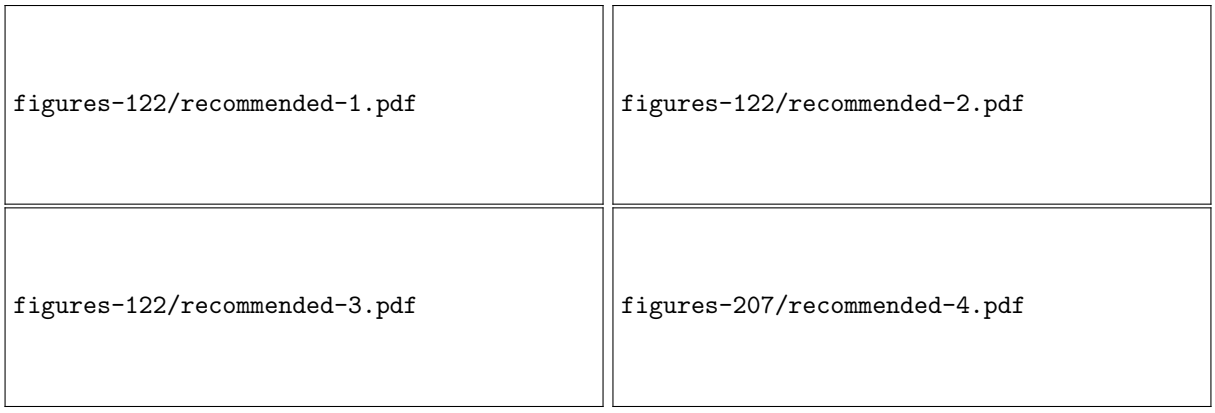


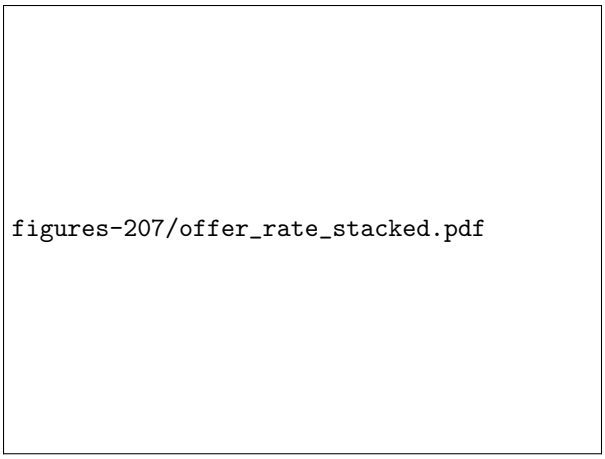
Figure 1: *Example portfolios and recommended changes, as an illustration of personalized feedback derived from our model. In portfolios 1 and 4, the model recommends the student apply to a less selective last choice, as a "safety program." In portfolio 2, the model recommends a more selective second choice, as a "reach" or "match" program. In portfolio 3, the model recommends a less selective second choice. The optimal change is a function of the applicant's entire ranked list. These changes are calculated using an algorithm that finds the optimal perturbation to a portfolio using the model of Ali and Shorrer [11], which optimizes expected program quality (selectivity) when admissions probabilities are fully correlated and thus favors portfolios that implement a "reach-match-safety" approach. Furthermore, in practice, individual student portfolios are highly non-monotonic in quality (though, as shown in Figure 25, average quality measures by list position decreases monotonically).*

is research highlighting the role of suboptimal first choice applications: in Chile, Larroucau et al. [53] warn students with high first choice admissions chances, and Cohodes et al. [26] in NYC illustrate that "an important channel [for successful interventions] is their effect on reducing nonoptimal first choice application strategies." However, it is unclear whether these interventions are optimal for all students, or what the gain from personalized feedback that considers a student's full current application set could be. For example, we find that in NYC, a student's *list length* does not predict substantially undermatching, unless the *ordering* of programs on the list is also considered. Especially as modern application systems allow for personalized feedback, it is important to consider their design and the gain from personalization.

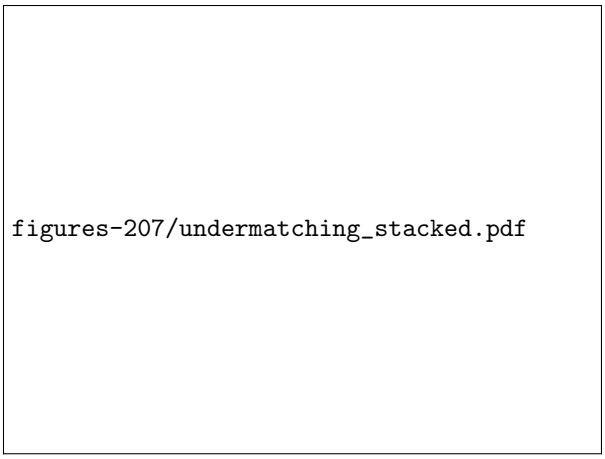
We make two conceptual and methodological contributions in understanding the relationship between students' application decisions and undermatching.

(a) Quantifying undermatching using counterfactuals. To understand how application behavior leads to undermatching, we first need to quantify *individual-level* undermatching. Notice that a student cannot be said to have truly undermatched if they would not have been admitted to other programs had they applied. Therefore, measuring individual-level undermatching requires access to precise counterfactuals, which have historically been unavailable in prior work on undermatching in high school or college admissions.[3] Leveraging now-available data in New York City on individual student lottery numbers and remaining application

[3]In the analysis of college application undermatching in the United States, one can study application behaviors but not admissions counterfactuals, and often even data on where a student was actually accepted is missing [77], a known limitation [19]. For this reason, literature on college undermatching primarily considers individual colleges or categories of schools (e.g., "most competitive" or any "selective college") [47, 77], in contrast to our continuous measures of undermatching based on quality measures such as program offer rate, impact, performance, and graduation and college enrollment rates. Such work also cannot tie an *individual's* behavior to their level of undermatching. Also in the NYC High School context up through the 2016-2017 admissions cycle, Corradini [28] considers school-level (not program-level) counterfactual admissions, creating sets of schools to which students have non-zero probability of admissions; she then shows that these choice sets cannot fully explain the "choice gap," that Black and Hispanic students apply to lower quality schools. We build on this work by calculating precise individual-level counterfactual admissions for each program (a school may have multiple programs, with different admissions criteria), using lottery numbers and other admissions-relevant information. We can do so because individual lottery numbers and detailed admissions criteria for each program are now available to researchers.



(a) *Offer rate for match and counterfactuals*



(b) *Undermatching across groups*

Figure 2: *(a) Average offer rate (for the applicant's match and counterfactuals) and (b) undermatching (difference between match and counterfactuals) by race/ethnicity. Figure (a) shows that on average, Asian and white students match to more selective programs than do Black and Hispanic students. Furthermore, the gaps in where students could be admitted (in (a), differences in dark blue bars) are much smaller than where they are matched, due to application behavior (in (a), differences in light blue bars). Counterfactual matches are restricted to those as geographically close as the student's actual match. Analogously, figure (b) shows that Black and Hispanic students undermatch more relative to Asian and white students; that is, there is a larger gap between the offer rate of the program they matched to and the lowest offer rate program they could have matched to. Both figures also show these effects when restricting counterfactuals to programs in a student's portfolio. These latter effects are thus due to list inversions in which students list higher offer rate programs earlier than lower offer rate programs, and account for about one third of undermatching. Error bars show 95% nonparametric bootstrap CI.*

data (such as academic performance and diversity characteristic eligiblity) in combination with detailed publicly-available information about how programs rank students, we can accurately compute precise *counterfactual* match outcomes at the individual level for most programs—where students would have matched, if either one or many students changed their application behavior.

With these counterfactuals, we can define and measure **undermatching** as the difference in program quality between where a student *actually matched* and the highest quality program they *would have been accepted to* had they applied and that is also nearby geographically (i.e., no longer commute distance than their actual match). For example, if an applicant matched to a program with measured quality as 0.4 but could have matched to one with quality 0.6 had they applied, their individual undermatching level would be 0.2. Our definition of undermatching isolates the effects of application behavior from those of admission policies and the geographic availability of high-quality programs, since it compares applicant outcomes to counterfactuals holding fixed everything but application behavior. This measure—by taking into account where individuals would have been admitted—allows us to measure how undermatching correlates with characteristics that are used by programs to make admissions decisions, such as lottery number, Free/Reduced Lunch eligibility, and neighborhood. Without precise counterfactuals, it would be unclear whether correlations between behavior and these characteristics are because students know if they will be admitted. By considering only geographically nearby programs, the measure further accounts for differences explainable by geography and commute preferences.

We quantify undermatching with respect to several measures of quality: program offer rate, graduation rate, college enrollment rate, NYC-provided measures of school performance and impact, and an average of NYC-provided school quality scores on different features (performance and impact scores combined with survey ratings from each school's community of the degree of rigorous instruction, collaborative teachers, supportive environment, effective school leadership, strong family-community ties, and trust). An advantage of studying New York City public school matching is that a wide variety of such measures exist, are provided

by the city to parents, and have been validated as being important to parents (e.g., the features used in our combined measure include those identified in the survey by Corradini and Idoux [30]). Thus, while each applicant's preferences may not be fully captured by such scores, undermatching according to these metrics is an important measure for policy, and our measurements further quantify how big the differences between individual preferences and the metrics would need to be to provide a preference-driven explanation of applicant behavior.[4]

In Section 2.1, we quantify undermatching in the 2022-23 NYC high school match, showing disparities in undermatching across different groups of students. Quantifying undermatching allows us to then study the sources of undermatching and the potential effects of interventions, as discussed next.

(b) Analyzing full application lists. To then understand how interventions can reduce disparities in undermatching, we would like to quantify the *preventable* sources of undermatching. Of course, in hindsight, a student could have applied to exactly the single highest quality program they would have been admitted to, but in practice, they must make decisions *ex ante*, under uncertainty. To fully understand the space of possible interventions, we would also like to understand how a student's *full list* of programs contributes to undermatching (does undermatching arise primarily from the beginning, middle, or end of lists? from not reaching or not listing safety programs?), and how sources of undermatching vary across students and groups (do different demographic groups tend to undermatch for the same or different reasons?). To do this, we build on a recent theoretical model of Ali and Shorrer [11], who characterize optimal "portfolios" under correlated uncertainty in admissions and show that they contain a mix of reach, match, and safety choices.

We use this model empirically in two ways. In Section 2.2, we quantify the extent to which different groups of students deviate from this strategy by examining three interpretable behaviors indicative of "no reach," "no match," and "no safety." While the theory tells us that each of these are mistakes a student can make, it does not tell us which of the mistakes are most common in practice. Different mistakes require different interventions. The theory also shows that longer lists allow students to implement a more effective reach-match-safety strategy. We examine the extent to which longer lists help students in practice, by considering a measure of *effective* list length. Then, in Section 2.3, we study an empirical instantiation of Ali and Shorrer's model in our setting, where we assume students have a uniform prior over a common score used by programs to make admissions; students are admitted to a program if and only if their score is below the program's threshold, which we set to be the program's offer rate. In this empirical model, we study optimal local perturbations of existing portfolios, mimicking explainability methods in machine learning to further understand common sources of undermatching. We evaluate the efficacy of such recommendations in reducing undermatching at the individual level, and in reducing disparities in match outcomes in equilibrium.

Findings. Using data from the 2022-2023 NYC High School Match, we connect ranking behavior to undermatching and design optimal feedback mechanisms. In the process, we quantify undermatching disparities, using precise program counterfactuals.

(1) There are substantial undermatching disparities by race/ethnicity, income, neighborhood, at-home language, and application competitiveness. We find substantial undermatching. On average, a student matched to a program that accepted 64% of applicants, but counterfactually could have

[4]In the main text, our primary operationalization uses the program offer rate—out of those who applied (in the stable matching sense), the fraction accepted. As shown in Appendix A.4, offer rate correlates with the other measures (performance, impact, graduation rate, college enrollment rate, and the aggregate score), and our results replicate for these measures: similar undermatching patterns hold, and recommendations to reduce undermatching by offer rate also reduce undermatching according to other measures. Selectivity and other quality measures are commonly used in the study of matching disparities, e.g., [4, 26, 27, 77]. Thus, regardless of student preferences, undermatching nevertheless is an important measure, to the extent that district-published school quality measures are meaningful—our undermatching measures document how much of the disparities in match quality (as measured by the district) are due to application behaviors as opposed to admission policies. For this reason, undermatching is well-studied for higher education [77], though less so for pre-collegiate education.

Other dimensions are also important for student preferences [20]. There may also be true heterogeneity in whether families prefer, e.g., their children to attend diverse schools or schools with other children like them [5, 39, 41, 45, 62]; school effectiveness also differs by student [49]. Programs also differ substantially by sports and specialty programs (such as arts, STEM, and language) offerings, and applicants have heterogeneous preferences over these characteristics. To the extent that our measures do not capture individual true applicant preferences, our measure of undermatching may not reflect student and family disutility. However, Corradini and Idoux [30] find through a survey in NYC that, while there are substantial behavioral differences, guardians across racial groups generally similarly value school characteristics, especially safety, though there is also racial homophily. We use NYC-provided scores on various dimensions to approximate the features identified by the survey.

matched to a program that only accepted 37% (assuming only that single student changed their application behavior); this difference is 1.5× larger for Black and Hispanic students relative to Asian and white students. Interestingly, we also find that undermatching disparities are *highest* for the most competitive students: Black and Hispanic students with the best lottery numbers and most selective counterfactual admissions have the highest undermatching (2× as much undermatching) compared to their Asian and white counterparts. Similar results hold for undermatching with respect to all other program quality measures. We also find substantial inter-group disparities by Free and Reduced Lunch eligibility, at-home language, and neighborhood, but not by sex. In some zip codes, undermatching is over twice as high as a neighboring zip code. Furthermore, these gaps are much larger than those induced by geography or admission policies: disparities in where students *can* match (that are as close geographically) are much smaller than where they *do* apply and match. However, there is substantial *intra-group* heterogeneity: applicant demographic and academic performance characteristics explain just 3.8% of the variance in undermatching.

(2) Simple and interpretable deviations from reach-match-safety strategies explain undermatching, unlike student characteristics alone. Calculating individual-level counterfactuals further enables connecting individual-level application behavior to undermatching, explaining both individual undermatching and group-level disparities. Asian and white students disproportionately undermatch as a result of not listing enough safety programs and subsequently failing to match.[5] In comparison, Black and Hispanic students disproportionately undermatch as a result of ranking non-selective programs above selective ones.[6] In particular, Asian and white students are more than 2× more likely than Black and Hispanic students to not match to any program (i.e., not list any safety programs), while more than 3× less likely to have listed a 100% offer rate program first in their list (i.e., not list a reach program). Furthermore, while Black and Hispanic students on average submit lists of similar length to Asian and white students, their "effective" list lengths are much shorter, since they disproportionately list safety programs higher. These simple behaviors are predictive of undermatching—in a regression, including such portfolio behaviors increases explained variance of undermatching to 40.9%, from the just 3.8% without behavioral features.

(3) Simple personalized feedback can reduce individual-level undermatching. Finally, we show how individual rankings can be (minimally) modified to decrease undermatching, using only program-level offer rate information from the previous year (even without student-level information like academic performance or lottery number). According to our model, Black and Hispanic students benefit relatively more from applying to more selective programs early in their lists, while Asian and white students benefit relatively more from applying to less selective programs later in their lists, especially in the middle. These individualized perturbations—which modify portfolios by only one program and do not consider individual-level admissions probabilities—decrease undermatching by 24% on average (in the setting where only the single individual changes their ranking behavior according to the recommendation). Adopting a full reach-match-safety portfolio reduces undermatching by 57% on average. We then evaluate the feedback when *every* student adopts them simultaneously: the feedback reduces disparities (for example, reducing the offer rate gap between Black and white students by 15%), further quantifying how disparate application behavior affects matching disparities.

Putting things together, our work lays a foundation for personalized application feedback and broader interventions: we identify how individual students can change their ranking behavior to decrease undermatching. This offers a useful lens to understand the potential effects of interventions: do they help students list more safety programs, or do they help students list more reach programs? In New York City alone, recent (proposed and deployed) interventions include providing students with more information about high quality nearby programs, providing information about admissions probabilities to help students avoid not matching [75], and removing caps on list length (as of the 2024 cycle). Our results suggest that the latter two deployed changes would especially benefit students who do not match (who are disproportionately high-income, Asian, or white), while the first type of change would especially benefit students who do not list reach programs (who are disproportionately low-income, Black, or Hispanic).

[5] We do not yet use enrollment data, and so cannot rule out the possibility that some students are (rationally) applying more aggressively due to having higher-quality outside options in the form of private schools, cf. Calsamiglia and Güell [23]. However, especially if Asian, white, and higher-income applicants disproportionately have access to such outside options, then such an effect would mean that our disparity measures are underestimates.

[6] As explained in the discussion and supplement, these inversions cannot be explained by behavioral models in which students sequentially search for and select schools, in which each choice is locally optimal under an information set.

In Section 2, we detail our results. In Section 3, we discuss the implications for particular interventions. Methods are detailed in Section 4, and Section 5 contains a discussion of related work.

2 Results

We analyze data from the 2022-2023 admissions cycle for the New York City High School match, with 58,901 General Education students who applied to at least one of 867 unique programs across 434 schools in the main high school match. For 579 programs accounting for 78.9% of matches, we have access to all information needed to compute precise counterfactuals for each student. That is, would a given student have been matched to a program had they applied to that program above their actual match (and no other student changed behavior)? In 2022, programs used student lottery numbers, grade tiers to rank applicants, diversity criteria (based on grades or Free and Reduced Lunch status), and priorities for students from a certain borough, who attended the associated middle school, or who live in the zoned area. These priorities are documented for each program in the program directory [61]. Such priorities and rankings can get complex, as detailed in the Methods section below; however, we can accurately simulate these decisions, given the data. The remaining programs used information that was not available to us, or which does not exist for students who did not apply (such as musical auditions).

We quantify undermatching as the difference in program quality of a student's actual match and the best quality among programs a student could have matched with that is also closer than their actual match. We consider several measures of program quality: offer rate, graduation rate, college attendance rate, and city-calculated measures of performance and impact.[7] In the main text, to simplify exposition, we focus on offer rate as a measure of quality, since it also serves as a measure of overall "demand" for the program and thus as a proxy for admissions likelihoods. Offer rate is defined as the ratio between students who—in the deferred acceptance algorithm—were admitted by a program and students who were considered by the program. In the supplemental material, we reproduce results for alternative measures of undermatching, finding qualitatively similar results.[8] When studying students' application lists, we also use offer rate as a way to understand the relatively likelihood that a student will be matched to different programs. Methods are detailed in Section 4.

2.1 There are substantial disparities in undermatching

We first analyze patterns in the quantity of undermatching, before turning to patterns in *how* students undermatch.

Demographic and spatial disparities. Figure 2 illustrates undermatching and disparities in undermatching. Students on average matched to a program with offer rate 64%, but counterfactually could have matched to a program no farther away with offer rate 37% (meaning that on average, students undermatched by 27 p.p.). Undermatching is about 1.5× higher on average for Black and Hispanic students compared to Asian and white students. Notice, also, that undermatching disparities generally are much larger than disparities in where students could be admitted if they had applied, as illustrated by the differences in the dark blue bars in Figure 2.

The figure also shows that undermatching exists even when counterfactual programs are restricted to those within a student's existing list, meaning that students listed less selective programs above more selective programs that they could have been admitted to. Across all groups, rank order *within* the portfolio accounts for about one-third of undermatching: the most selective program in a student's application to which they would have been admitted has an average offer rate of 56%, but was listed below the less selective program to which they were actually admitted. This illustrates how our measure of undermatching considers not only the *set* of programs a student applies to, but also the *order* in which they are listed.

Figure 3 illustrates geographic patterns in match quality and undermatching. For example, students in the upper Bronx tend to match to relatively high offer rate programs, consistent with their relative geographic

[7] For offer rate, a lower number is viewed as higher quality. For all other measures, higher numbers are interpreted as higher quality. In all cases, we fix the sign of the undermatching measure so that it is positive, and so that higher numbers correspond to more undermatching.

[8] Alternative measures tend to be highly correlated with offer rate, as shown in Figure 22.

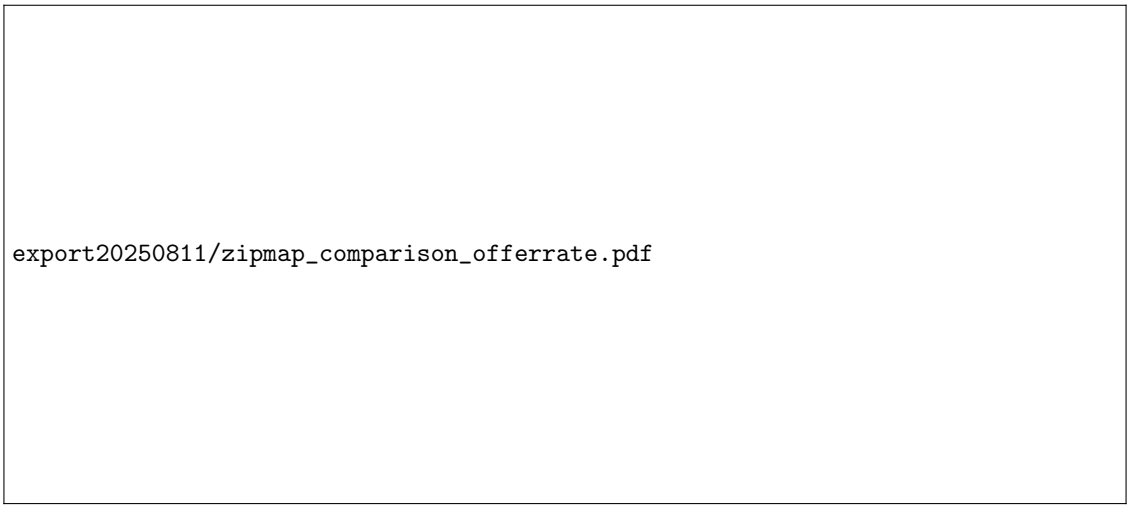


Figure 3: *Map of New York City by zip code, showing the geographic distribution of match offer rate and undermatching. On the left is the average match offer rate for students who live in that zip code; on the right is the average amount of undermatching. Undermatching varies substantially by geography. Geographic patterns of match offer rates and undermatching distinguish between geographic access to selective programs and application behavior; some zip codes have 2x higher undermatching than a neighboring one. For example, students in Riverdale (northwest corner of the map) match to relatively high offer rate (less selective) programs (around .65), but have little undermatching (around .1 difference from counterfactual). This may reflect its relatively remote but affluent status. In the rest of the Bronx (north end of map), students match to higher offer rate programs on average, but also experience much more undermatching. Meanwhile, in downtown Manhattan (bottom tip of the island on middle left), students on average match to selective programs (left plot is largely blue), but they could have matched to even more selective programs, given the high density of such programs. For privacy reasons, zip codes with 20 or fewer applicants are omitted.*

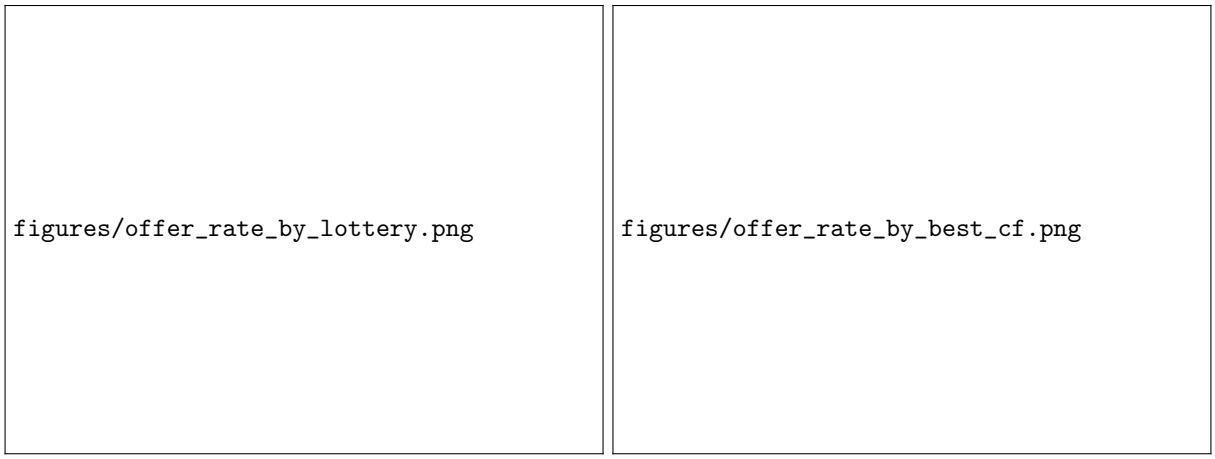


Figure 4: *Undermatching by lottery number (lower is better) and offer rate of the most selective program they counterfactually would have been admitted to (that is as close as their actual match), by race and ethnicity. The most competitive (by grades, lottery numbers, or other admission criteria) Black and Hispanic students are especially undermatching compared to their Asian and white counterparts, indicating that they are not applying as aggressively as they could have. Similar trends hold when measuring competitiveness by applicant grade tier (Fig. 15). Error bars show 95% nonparametric bootstrap CI.*

isolation; meanwhile, within the upper Bronx, more affluent areas (like Riverdale) have less undermatching on average (around 10 p.p.) compared to neighboring, less affluent areas (30-40 p.p.). This example also illustrates how our measure of undermatching isolates the effect of behavior from geography.

Disparities conditional on applicant competitiveness and grades. Figure 4 shows that disparities in undermatching are most prominent among the most competitive applicants, as measured by lottery number or by the selectivity of the program they counterfactually could have been admitted to. This suggests that undermatching disparities are not explained by group differences in applicant competitiveness. Regressions in Appendix A.5 show more generally that undermatching disparities by race and ethnicity persist after conditioning on FRL status and grades, across measures of performance—i.e., group differences between where a student actually matched and could have matched persist even when comparing students with similar middle school grades. Patterns with respect to competitiveness also suggest subtlety in how undermatching can occur; for example, the non-monotonicity in undermatching for Asian and white applicants suggests that those at the middle-competitiveness range are not applying to sufficient "match" programs, and are instead opting for either "reach" or "safety" programs. We explore behaviors more directly in Section 2.2: What behaviors explain undermatching, and how do these behaviors differ in prevalence across groups?

A natural related question is if applicant competitiveness causally influences applicant behavior. Given knowledge of their own competitiveness, do students change their behavior? In the application year under study, students had access to their lottery numbers before they submitted their applications. Since lottery number determine competitiveness and are assigned at random, correlations in behavior by lottery number provide a mechanism by which to study this question. Interestingly, we did not find evidence that behavior is affected by lottery numbers, by any demographic group or overall. Specifically, we regressed multiple application behaviors on lottery number, and found no statistically significant effect, even when considering group-specific interactions (after correcting for multiple hypotheses). These regressions are given in Appendix A.8. In other words, in the year we considered, students do not appear to meaningfully use lottery numbers when choosing which programs to apply to. As a consequence, the disparities in undermatching we find are not driven by strategic use of lottery information. It should be noted, however, that lottery numbers were only provided as a hexadecimal code, which may be hard to interpret [57]. Clearer explanations of how to use lottery numbers may influence student behavior.



Figure 5: *We analyze the prevalence of three interpretable deviations from optimal reach-match-safety behavior: (i)* ***no reach*** *(listing an unselective program first), (ii)* ***no match*** *(being rejected by program(s) and then matching to an unselective program), (iii)* ***no safety*** *(being rejected by all programs). Notice that the "no reach" deviation is much more common among Black and Hispanic applicants, while the "no safety" deviation is much more common among Asian and white applicants. Also observe the importance of ordering: even though the first portfolio has lower offer rate programs, since the first-ranked program is unselective, these other choices are inconsequential. Lastly, within each group, students with Free/Reduced Lunch eligibility tend to undermatch due to not listing reach and match programs, while FRL ineligible students tend to undermatch due to not listing safety programs. Error bars show 95% nonparametric bootstrap CI.*

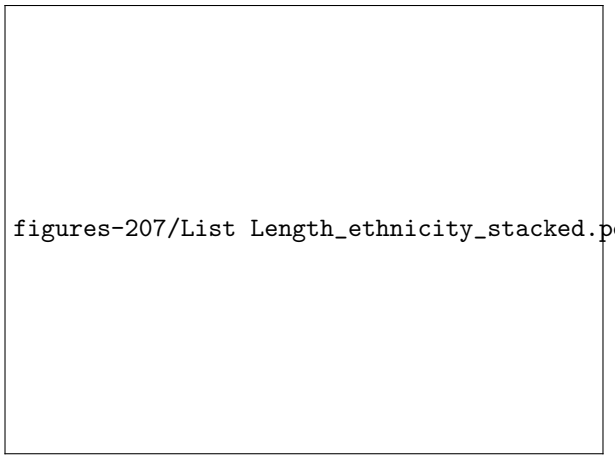


Figure 6: *Students across groups list a similar number of programs overall (out of a maximum of 12). However, Black and Hispanic students have lower effective list length: on average, Black and Hispanic students list an unselective, 100% offer rate program earlier in their list, rendering subsequent programs inconsequential. Error bars show 95% nonparametric bootstrap CI.*

Further results: Additional demographic characteristics. Appendix A.3 shows undermatching distributions by race and ethnicity, sex, FRL eligibility, and at-home language. There is substantial inter-group heterogeneity, except by sex. We find, for example, those whose at home language is *Mandarin* or *Chinese (any)* have the lowest undermatching rates, regardless of Free/Reduced Lunch status.

Further results: Additional measures of undermatching. Appendix A.4 replicates Figures 2 to 4 for undermatching in terms of program graduation rate, program college enrollment rate, school performance, and school impact, and our aggregate quality metric that combines survey scores; the documented disparities continue to hold, including spatially. One difference is worth noting: optimal counterfactual matches tend to be more similar across students when measured by these other metrics, since there exist relatively uncompetitive programs that perform well by these metrics (meanwhile, low offer rate programs are by definition competitive). As a result, the exact patterns in Figure 4 vary by metric.

2.2 Interpretable deviations from reach-match-safety explain undermatching

Prevalence of interpretable deviations from reach-match-safety. Figure 5 shows the prevalence of three interpretable categories of behaviors that indicate deviation from a reach-match-safety strategy. *No reach* is defined as having listed a 100% offer rate program first. This is an extreme form of not reaching, since the student is guaranteed to match to this unselective program. *No match* is defined as matching to a 100% offer rate program later in a student's list, indicating that the student only listed programs that were too competitive before matching to a "safety" program. *No safety* is defined as not matching to any program at all, indicating that the student did not list a sufficiently safe choice. 38.3% of applicants exhibit one of these three (mutually exclusive) behaviors. Black, Hispanic, and FRL-eligible students disproportionately fall under *no reach* and *no match*, while the opposite is true for *no safety*. Race and ethnicity correlate with these behaviors more than does family income: for each behavior, patterns within a racial and ethnic group but differing on FRL eligibility are more similar than those across racial and ethnic groups.

List length and effective list length. Figure 6 compares list lengths across race and ethnicity. Longer lists can benefit students since they can apply to a more complete set of reach, match, and safety programs. Notably, list length does not differ substantially across groups. However, there are disparities in *effective* list length, defined as the number of programs up to and including the first 100% offer rate program. This is the "effective" list length, since the student will never match to programs listed below a 100% offer rate program under the deferred acceptance algorithm: adding additional programs beyond such a program has

no effect. Figure 6 indicates that Black and Hispanic students, despite having similar or higher list lengths, have significantly lower effective list lengths. Thus, stronger advice to students simply to rank more programs is likely to be ineffective without advice that also recommends stronger consideration of order.

Portfolio behaviors explain much more of undermatching than student characteristics alone. We regress each applicant's undermatching difference against demographic and other individual characteristics, both with (Table 4) and without (Table 3) the measures of portfolio behaviors.

As suggested by Figures 2 to 4 and quantified in Table 3, undermatching correlates significantly with several performance, demographic, and socioeconomic characteristics; coefficients for most racial/ethnic minority groups, FRL status, and average middle school grades are significant ($p < .01$); the coefficient for sex is not ($p > .1$). However, these characteristics only explain 3.8% of the variation in undermatching: there is substantial individual-level variation not captured by student characteristics. As shown in Table 4, portfolio behaviors are substantially more predictive of undermatching than are student characteristics alone: adding measures of portfolio behaviors (such as effective list length, minimum and average offer rate information, a clustering of portfolios, and the presence of the deviations described above) increases explained variance to 40.9%. Appendix A.5 shows qualitatively identical results when undermatching is calculated with respect to the other program quality measures, and when the measures of portfolio behavior are only defined with respect to application-time characteristics (the measures discussed above are *ex post*, after outcomes are realized, while we also seek to understand behavioral measures using information at the time of the application).

Together, these analyses suggest heterogeneity in how students can undermatch, and therefore, potential benefits of providing personalized application feedback to applicants to reduce undermatching. Next, motivated by this analysis, we calculate improvements for each student's portfolio using only information available at the time of the application.

2.3 Recommendations reduce individual undermatching and close disparities

We now recommend changes to each student's individual portfolio to reduce undermatching. Using a model of portfolio decision-making under uncertainty [11], we identify the optimal single change to each applicant's portfolio, and then evaluate how outcomes change when one or many applicants adopt the recommendations. Figure 1 illustrates example portfolios and changes recommended by the model.

Model and evaluation summary. We consider an empirical version of the theoretical model studied by Ali and Shorrer [11]. The objective is to minimize the offer rate of the program the student is matched to in expectation. The model uses a common score assumption, where each program is assumed to rank students according to a single score (e.g., all programs use the same lottery number or test score to rank students). In our empirical setup, when making recommendations we assume that a student has a uniform prior over their score (between 0 and 1), and they are admitted to a program if and only if their score is below a threshold set equal to the program's offer rate. This means that a student is less likely to be admitted to a program with a lower offer rate. A student matches to their highest listed program that they are admitted to. Therefore, portfolios in the model are represented as a length-12 vector of offer rates. To only use information available *ex ante*, we compute offer rates in the *previous* year's match (outcomes are evaluated according to the current year's offer rates).[9]

At a high level, the model rewards portfolios that exhibit more of a reach-match-safety pattern [11] where programs are listed in ascending order of offer rate. If the model recommends a student apply to a program with a given offer rate, we select the program with the most similar offer rate that is also geographically closer than the student's true match. See Section 4.5 for details about the model.

We then evaluate recommendations according to the empirical counterfactual admissions calculations, using our matching simulator—i.e., while our model to *make* recommendations is naive in several respects, recommendations are *evaluated* using the full complexity of how programs accept applicants. We first evaluate the quality of these recommendations if each student adopted them but no other students did, and examine the distribution of recommendations made. Then, we evaluate outcomes *in equilibrium*, where every student adopts the model's recommendations.

[9] In Figure 24, we show that offer rate is largely stable across years, except for small programs.

Adopting the model's recommendation decreases individual undermatching. We first analyze the impact when individual students adopt the model's recommendations, holding the behavior of all other students fixed. For Asian students, the recommendation reduces undermatching by 4.1 percentage points (from 20.6% to 16.5%); for Black students: 8.9 p.p (from 32.1%); for Hispanic students: 7.4 p.p (from 29.2%); and for white students: 3.7 p.p (from 20.4%). This represents a 24.2% decrease in undermatching overall. As shown in Figure 16, these recommendations would also generally reduce undermatching as measured by the other quality metrics, especially for Black and Hispanic students.

We also find that instead recommending an entire portfolio of programs—employing a full reach-match-safety strategy, using ex ante information and without using personalized probabilities—would reduce individual undermatching as measured by offer rates by 57.1% on average. Figure 18 shows the effect on other measures. Interestingly, for Asian and white students, while the full portfolio recommendations improve outcomes by offer rate (selectivity), they sometimes worsen outcomes slightly with respect to other metrics—especially when the full portfolios are modified with respect to the model. One reason could be that while the single best change finds more "obvious" deviations, completely changing a list may override information that the model does not consider: for example, students may especially choose programs that are relatively unselective but also perform well by other metrics, or leverage personal knowledge of admissions likelihoods. More generally, to the extent that recommendations may have varying effects depending on metric, it is important to provide explainable and diverse recommendations to applicants.

Recommendations vary both within and between groups. Figure 7 visualizes the distribution of recommendations made by the model for different groups. Consistent with findings from Section 2.2, the model is relatively more likely to recommend Black and Hispanic students to apply to a more selective program earlier in their lists, and is relatively more likely to recommend Asian and white students to apply to less selective programs later in their lists. However, there is substantial intra-group heterogeneity, consistent with the fact that student characteristics alone do not substantially explain undermatching.

Top-choice recommendations have a large impact. As a whole, we find that recommendations made by the model are concentrated towards the beginning of student lists. In particular, 27.0% of recommendations are to change a student's first choice program. We also find that recommendations earlier in the list tend to be more effective—perhaps because students for whom such recommendations are optimal had higher initial undermatching. For example, while on average student undermatching decreases by 24.2% overall from the recommendations, undermatching decreases by 37.9% for students who receive a recommendation to their first choice (these students also undermatch the most in the status quo). Recommendations at all other positions, however, also decrease undermatching by at least 10%.

We can also study optimal recommendations that are constrained to recommending a more selective program in a student's first choice. This is an especially natural choice since it is both relatively easy to explain and deploy, and has been studied in past interventions [27]. We find that such recommendations reduce undermatching by 13.7% (as compared to 24.2% when a single recommendation can be made at any position). So, in our model, 56.6% of the fixable undermatching from changing one program can be achieved by only changing the first program to be more selective. As shown in Figure 17, these recommendations benefit Black and Hispanic students relatively more compared to their Asian and white peers, consistent with our general findings that Black and Hispanic students are relatively more likely to undermatch due to listing unselective programs early in their lists.

Adopting the model's recommendation decreases disparities in equilibrium, but increases congestion. Finally, we analyze the effectiveness of the recommendations *in equilibrium*, when every student follows the feedback simultaneously. This induces *congestion*, as students may be recommended the same programs. Holding the number of students admitted to each program constant, feedback cannot improve average match outcomes (since the total sum of students matched to each program is fixed), but can reduce disparities between groups. Simulating such changes, we find that each student making the single change suggested by our model reduces disparities, (for example, reducing the offer rate gap between Black and white students by 15.3%). However, the resulting congestion leads to 16.3% of students going unmatched (in comparison to 7.9% in the original match); this occurs because more students are now applying to more



Figure 7: *Recommended changes to portfolios differ substantially across demographic groups. For example, the dark blue square in the recommendations for Hispanic students indicates that about 3000 such students would be most helped (concerning match offer rate in expectation) by changing their first-ranked program to one that has about 50 percentage points lower offer rate. Asian and white students are relatively more likely to benefit from applying to less selective programs later in their list, while Black and Hispanic students would primarily benefit from applying to more selective programs early in their lists.*

competitive programs, which then may reject some students who previously matched to that program—these students may not have, in turn, applied to other less selective programs later down in their list. When all students fully employ a reach-match-safety approach, the offer rate gap between Black and white students decreases by 66.0%—this further suggests about $\frac{1}{3}$ of the gap in the status quo is driven by factors other than application behavior, such as admissions criteria and geography. However, this latter change also substantially increases the number of students who do not match (and must be administratively matched), to 35.0%. Thus, while feedback can reduce disparities, effective recommendations when all students use them must be "decongesting," and recommendations must take into account recommendations also given to other applicants. For example, it may be important to also recommend applicants add safety programs to their portfolios. The design of such recommendations is important future work.

3 Discussion

We quantify the behavioral sources of undermatching in the New York City High School Match—i.e., the extent to which students match to less selective programs than they could have matched to had they applied differently. Differences in match outcomes are driven substantially by complexity in the application process (as opposed to program admissions criteria or geography). Moreover, we find that the application behaviors

that cause undermatching differ by racial/ethnic and socioeconomic groups, as do the interventions that would mitigate undermatching.

Implications for policy and interventions. Our findings help inform the design of interventions, both broad and personalized. We quantify heterogeneity in sources of undermatching, both within and across demographic groups. This informs the potential efficacy of different interventions (for different groups). In both Sections 2.2 and 2.3, we find that undermatching can arise at the beginning, middle, or end of a student's list, and that patterns differ across and within demographic groups. This suggests that various proposed and deployed interventions will have different distributional effects, depending on whether they focus on helping students apply to higher quality programs earlier in their list or on helping students apply to more safety programs. The first category includes interventions that provide information about high quality programs, emphasize the strategyproof-ness of deferred acceptance (encouraging students to list more competitive programs early in the list), and flag when a student applies to an extremely high probability program first in their list. The second category includes interventions that flag high risk portfolios or flag programs that a student is extremely unlikely to match to.

Notice also, that providing information at different levels of coarseness and explanations can have different effects. For example, *reaching* for selective programs necessarily equates to applying to programs with low offer likelihoods. Consider the difference between flagging programs that a student is almost certainly not going to be admitted, to flagging reach programs that a student has some likelihood of being admitted to. Furthermore, more direct guidance on how students should use this information will also have an effect: does the intervention encourage or discourage students to apply to lower likelihood programs? For example, at the college application level, Tomkins et al. [77] find that showing students average score cutoffs increased undermatching, by discouraging *reach* applications.

In addition, while longer application lists can theoretically help students form better portfolios (as emphasized by NYC advice to applicants), we find that many students—especially Black and Hispanic students—already submit long but *effectively* short lists, due to the early presence of a near-guaranteed admissions program. This suggests the limitations of interventions targeted at increasing list lengths (such as NYC's recent policy removing list length restrictions), both in decreasing undermatching overall, and in closing disparities. Our work thus suggests the potential benefits of a wide range of interventions currently being explored and deployed [16, 26, 27, 52], while suggesting the lack of a "one size fits all" approach.

Along these lines, our work provides a principled foundation on which to build personalized feedback that can handle heterogeneous sources of undermatching. For example, while existing interventions focus primarily on the beginning or end of a student's list, our model suggests intervening on the *middle* of many students' lists, often to avoid forms of correlation neglect [68] in which students list unselective programs above selective programs. We find that recommendations can substantially reduce individual undermatching, and can reduce disparities when adopted widely. However, our empirical findings suggest that the resulting congestion can increase the number of students who do not match, revealing the need for implementing decongesting measures and ensuring that students list sufficient safety programs.

Assumptions and limitations. We focus on undermatching, primarily operationalized in the main text as a student being matched to a less selective program than they could have been. Program selectivity correlates with other quality measures (performance, impact, graduation rate, college enrollment rate, and survey measures of school environment by the community), and our results hold when undermatching is defined with respect to those quality measures. However, as discussed in Footnote 4, it is important to note that—to the extent that these program quality measures do not capture true applicant preference—undermatching may not reflect student disutility, deviations from reach-match-safety approaches may not mean "mistakes," and recommendations may not cause behavioral changes. We note especially that our recommendations are designed at the individual student level; if applicants have strong preferences for attending high schools with specific peers or schools with specific peer compositions, then such recommendations may be ineffective. However, measuring undermatching—and isolating effects from behaviors versus admissions policies and geographic access—remains important in designing interventions and policy changes more broadly.

More generally, our work suggests "how large" such student utility differences would have to be from the quantified program quality measures to explain behavior. It is worth reflecting on how the measured student behavior could be rationalized by an optimal portfolio or search model (such as those of Chade and

Smith [25] or Ali and Shorrer [11]). (1) It is possible that individual applicants truly prefer the programs listed and in that order, and prefer their outside option over all non-listed programs (though findings of Corradini and Idoux [30] and others suggest that this is not the case on average); then, all rankings could be fully rationalized. (2) Now instead suppose that our quality measures well-capture student utility. (a) Then, regardless of whether offer rates accurately capture student admissions probabilities, our undermatching measures would correctly quantify *ex post* application mistakes and disparities between groups—since we calculate undermatching in comparison to a counterfactual program to which the student would have been admitted. Furthermore, ordering *inversions* (ranking a lower-quality program above a higher-quality one) could not be even *ex ante* rationalized because deferred acceptance is strategy-proof among programs that are listed; one would need students misreporting their preferences to explain such reversals (e.g., Hassidim et al. [43, 44], Shorrer and Sóvágó [76]). (b) Finally, suppose that utility is well-captured, but the program offer rate does not reflect a student's individualized admissions probability. Then, which programs students list could be *ex ante* rationalized, even if they were mistakes *ex post*: it is possible that the chosen programs correctly balance the utility of matching with each program with the probability of matching.[10]

Relatedly, matches are several steps removed from downstream outcomes of interest: (a) being *matched* to a school does not mean that a student will *attend*—especially if demographic changes leads to matches that students and families less prefer, they may abandon their match and attend, for example a private school; such abandonment may be especially acute among more privileged families. There is similarly an aftermarket of reassignments after cancellations and for unmatched students [36]. For example, in Ghana, Ajayi et al. [10] finds that an informational intervention changed ranked lists but had a more limited effect on enrollment. (b) *Attending* a given school or program does not necessarily imply educational benefits, and school diversity may not translate to *classroom* diversity [71]. Thus, our methods cannot directly measure changes in downstream educational outcomes.

Finally, while our work points to characteristics of optimal feedback, there are several important steps to turn our approach into actual recommendations that can be sent to students. For example, our feedback model is naive, assuming that the recommender does not know each student's idiosyncratic likelihood of being matched. In practice, personalized recommendations may leverage such information, improving recommendation quality. Furthermore, our equilibrium simulation assumes that program capacities are restricted to the number that were matched to that program historically, and so cannot by assumption improve match quality on average across the applicant pool. In practice, there may be high quality programs that can accept more students had they applied. An important direction for future work is to design recommendations that by design minimize congestion, while being individually optimal. Finally—while we quantify undermatching through a range of program quality measures and design recommendations to optimize the selectivity of the matched program in expectation—in practice, "high quality" programs should be identified in conversation with policymakers, educators, and community groups. Furthermore, applicants should receive explanations alongside the recommendations, so they can choose based on the dimensions most important to them.

4 Data and Methods

4.1 Data description

We analyze the 2022-2023 admissions cycle for the New York City High School match, with 58,901 General Education[11] students who applied to at least one of 867 unique programs[12] (across 434 schools[13]) in the main high school match.[14] We have each student's application package (including random lottery number),

[10] One quantification of the difference between *ex ante* and *ex post* suboptimality is that, when we simulate applicants as following an *ex ante* optimal reach/match/safety portfolio using population-level offer rate information, undermatching decreases by 57%, but is not eliminated.

[11] We limit the analysis to students participating in the General Education match, and so exclude the students as part of the Students With Disabilities match, for which there are separate seat groups.

[12] to which at least one General Education student applied.

[13] A school may house multiple programs; one may be Screened and focused on science, while another may be open/zoned. We consider each program separately, though performance and impact measures are only available at the school level.

[14] We do not consider the 9 *Specialized High Schools*, e.g., the Bronx High School of Science, which operate a separate, parallel admissions process through a specialized test (8 schools) or auditions (1 school, LaGuardia High School, with multiple programs). NYC high school applicants submit rankings to the main match and the specialized high schools separately. Crucially, students

method	# programs	# matches	offer rate	grad. rate	college rate	performance	impact
Ed. Opt.	393	25722	0.85	0.87	0.56	0.53	0.48
Screened	152	16245	0.50	0.94	0.75	0.67	0.56
Open	57	3797	0.86	0.86	0.56	0.52	0.48
Audition	71	2845	0.39	0.93	0.70	0.64	0.52
Other	194	5631	0.57	0.74	0.50	0.56	0.51

Table 1: *Overview of programs by admissions method.*

match outcomes, middle school grades and exam scores, and zip code (but not exact address). We calculate the transit distance between each student's zip code to each school's zip code, and use it as a proxy for the student's commute distance.[15]

For each program, we have characteristics such as how they rank students (using grades and exam scores, zoning priorities, Diversity in Admissions, auditions, etc) and historical student demographic composition; we further use performance and impact measures as calculated by the Department of Education [33].[16] There are four primary categories of programs: Ed. Opt., Screened, Open, and Audition. Open programs admit students by lottery number. Ed. Opt. programs also admit students by lottery number, but set aside seats for students from all academic levels (each of three grade tiers). Screened programs admit students first by academic performance measures and then by lottery number. Audition programs admit students based on their performance in an audition. Of students who match, 89.6% match to programs in these categories. Table 1 provides an overview of programs in each category. Screened programs (which evaluate students using measures of academic performance) on average have lower offer rates (are more selective), as well as higher graduation rate, college enrollment rate, performance, and impact.

4.2 Measuring undermatching

Given a measure of program quality (such as offer rate or college enrollment), we define **undermatching** as the difference between the quality of the student's matched program and that of the highest quality program the student (1) would have matched to had they applied, and (2) is *at least as close by commute distance* as the program the student matched to. Then, for example, if a student is matched to a program that has an offer rate of 0.6 and is a 40-minute commute, but there exists a program that the student could have been matched to that has an offer rate of 0.2 and is a 30-minute commute, then we can quantify the student's undermatching (with respect to offer rate) as 0.6-0.2=0.4. For a student who did not match to a program they ranked, we impute their matched offer rate to be 1 and their commute radius to 30 min (the median commute distance for a student is 33 min). Restricting to programs a student would have been admitted to removes effects from school admission policies (such as prioritizing those with higher grades), and restricting to programs closer than the matched program further removes effects from some students living closer to high-quality schools, cf. [54]—isolating the effect of application behavior on undermatching.

submit initial-round rankings before learning about any offer(s). At the initial results round, students may receive up to one offer from the main match, one offer from the test-based specialized high schools, and/or multiple offers from programs at the audition-based specialized school. After students accept up to one offer, there is a follow-up waitlist round and then further matches. In this paper, we consider a student's match as the offer they receive in the initial results round from the main match. When calculating undermatching, we compare this initial results round offer to the initial round counterfactual we calculate. Thus, our work may overestimate undermatching for students who ultimately enroll in a specialized high school (3,710 students received an offer from a test-based specialized high school; 868 students received an offer from a LaGuardia performing arts program). Similarly, for the admissions year under study, high school enrollment data is not yet available to researchers, and so we cannot measure students who choose not to attend their offered match and instead enroll in a private school.

[15]We use the commute distance from the MTA's trip planner tool (which measures transit time between zip codes), with a commute distance of 0 for a school in the same zip code as the student. Times are calculated for a commute on a weekday morning before school begins. Under this approach, all schools in the same zip code are presumed to have the same commute distance to the student. While imperfect, such an assumption is necessary given that more granular location information is unavailable for students, due to privacy considerations.

[16]These measures use a combination of student achievement quantitative metrics such as test scores, graduation rates, and attendance (including metrics split by racial/ethnic group). *Impact* is adjusted with a comparison group based on incoming student factors, and *Performance* is unadjusted. The DOE provides these scores at the *school* level, as opposed to *program* level. Using student-level data from previous years, we calculate program-level graduation and college enrollment rates.

Program quality measures To measure the quality of each program, we use the following measures: program-level graduation, college enrollment rates, and offer rates (selectivity); and school-level *performance* and *impact*. We calculate program-level graduation, college enrollment rates, and offer rates using student-level data from previous years. The *offer rate* of a program is the fraction admitted to the program j out of those who applied to the program (in the stable matching sense), where a student is only counted as *rejected* if they ranked the program j but then either did not match or were matched to a program they ranked lower than they ranked j. Thus, students who were matched to a program they ranked higher than program j are not included in the offer rate calculation, as it is possible that they would have been accepted. *Selectivity* is defined as $1 -$ offer rate. Performance and impact measures are calculated by the Department of Education [33], and are the district's summary measures of student achievement at each school. These measures use a combination of student achievement quantitative metrics such as test scores, graduation rates, and attendance (including metrics split by racial/ethnic group). *Impact* is adjusted with a comparison group based on incoming student factors, and *Performance* is unadjusted. The DOE provides these scores at the *school* level, as opposed to *program* level.

Note that for all program quality measures except offer rate, higher values are assumed to be higher-quality. For all undermatching measures, we normalize the sign so that undermatching is positive and higher values correspond to more undermatching.

As shown in Figures 22 and 23, these measures largely correlate with one another, with pairwise correlations ranging from 0.44 to 0.89. In the main text, we focus on offer rate (selectivity) as the quality measure. Selectivity is commonly used in studies of undermatching [77]; results hold for each measure, as shown in the Appendix.

4.3 Calculating counterfactual offers

Crucial to our approach is the ability to calculate *counterfactual offers* for individual students. For 579 programs accounting for 78.9% of matches, we are able to (almost) exactly reconstruct their ranking function over students, enabling us to compute counterfactual offers. That is, would a given student have been matched to a program had they applied to that program (first among programs to which they would have been accepted)? In 2022, programs primarily used student lottery numbers and student grade tiers to rank applicants; other types of priorities also used include priorities for students from a certain borough, who attended the associated middle school, or who live in the zoned area. These priorities are documented for each program in the program directory [61].

Such priorities and rankings can get complex: at a high level, each program may have several "seat groups," and students who belong to that seat group have priority over all students who do not; within a seat group, students are ranked, such as by grades or lottery numbers or those with other priorities. Each seat group has an associated capacity. At many programs, a certain number of seats are set aside for students with different grades (in 393 Ed. Opt. programs, which prioritize students with High, Medium, and Low grades, respectively, for different seats) and for students who were free or reduced lunch (FRL) eligible (in 62 Diversity in Admissions programs). All types of programs can be a Diversity in Admissions (DIA) program (for example, some Ed. Opt. programs are also DIA). In the match process, these programs can be viewed as individual seat groups—each of which has its own capacity and ranks students according to the priority associated with the seat group. Students who apply to such a program apply to all seat groups, and seat groups are filled in a certain order. For example, if fewer 'High' category students apply to an Ed. Opt. program than there are seats in the 'High' priority seat group, then the highest-score students from the 'Middle' and 'Low' groups will take those seats. Next, the 'Middle' seat group seats are filled, followed by the 'Low' group seats. For DIA programs, FRL priority seats are filled before general seats.[17]

Some programs used information that is not available for all students. For example, all Audition programs as well as 23 Screened programs use supplemental materials (e.g., essays, other assessments, or auditions) to rank students. Since students who did not apply to these programs did not submit these materials, we are not able to exactly simulate counterfactuals for these programs. For undermatching measures and potential application improvements, we only include the 579 Open, Ed. Opt., and Screened programs for which we can effectively calculate counterfactuals (we do not consider Other programs, many of which rank students

[17] The order in which seat groups are filled matters [35].

using information not available to us). Of these 579, 57 are Open programs, 129 are Screened programs, and 393 are Ed. Opt. These programs accounted for 78.9% of student matches.

For these programs, we used the following procedure to calculate counterfactuals.

- Informed by the program directory, we create ranking functions that take each student's features and generate each student's scores for each seat group within a program.
- The cutoff for each seat group is the lowest score of a student who matched to a seat group.
- A student receives a counterfactual offer from a program if and only if their score exceeds the seat group's cutoff for at least one seat group.

The counterfactuals are accurate. When applied to the students who applied to each program (for whom we have the ground truth admission decision), our counterfactuals have an average accuracy across programs of 99.4%. The minimum accuracy across programs is 78%, and only five programs have an accuracy of less than 95%.[18]

Finally, to evaluate the effectiveness of the recommendations in equilibrium, we simulate the entire match with all students' preferences changed. This simulation can have more error than when considering each program-student pair in isolation, as errors propagate (a student incorrectly rejected from one program may take the seat of a student at another program). Thus, to evaluate the entire simulator, we simulate the match (with all GE students and programs jointly) with historical preferences. The resulting match is 91% accurate. To carry out this simulation, we assume that we know the match outcomes of students who in the historical data applied to a program for which we cannot calculate counterfactuals; all other students who applied to such programs are assumed to be rejected.

When calculating match outcomes in terms of offer rate, we use the offer rate of the original match (rather than in the simulated match). We further assume that the overall number of seats at each program is the number of students historically accepted to the program in the main match plus the students administratively assigned to that program. Jointly, these two assumptions imply that changes in application behavior cannot change match outcomes by offer rate on average, since the mean offer rate over student matches is invariant. Thus, we consider how recommendations change disparities in match outcomes by group, while holding population-level average outcomes fixed.

4.4 Characterizing portfolio behavior

Finally, we need to characterize the *portfolio decisions* made by individual students. Crucially, the strategic challenge facing students is a *ranking* decision: the utility of ranking a program kth depends on both the programs ranked above it (due to correlated admissions outcomes) and those below it (relevant if the student is rejected from the kth ranked program). Motivated by the theoretical literature [11, 25], we characterize a student portfolio by the sequence of their admissions likelihoods, i.e., student portfolios are characterized by a vector $R \in [0,1]^{12}$. (In comparison to the theoretical literature, we also represent the order of programs, which is important in deferred acceptance.) For example, a portfolio without any safety programs would have all its entries close to 0. To do this analysis, we require a notion of students' admissions likelihoods.[19]

[18] We note that the set rejected or matched to a program may not be representative of all potential applicants for whom we calculate counterfactuals. However, the in-sample accuracy of the simulator, as well as the geographic limitation in using counterfactuals (which limits potential distribution shift), suggests that such errors do not substantially bias results.

[19] This likelihood depends on an assumption of "what does the student realistically know, and what is uncertain, at the time of their application." On the one extreme, since the match is almost deterministic given the lottery numbers and where students apply (the only uncertainty left is ratings on qualitative aspects like essays and auditions, as used by 151 screened and audition schools, or similar other unknown features), we as researchers can almost perfectly calculate in hindsight whether a student would have been admitted to a program. However, in practice, students misestimate their chances of admission, and there is real uncertainty about school cutoffs [9, 48, 50, 77]. Thus, close to the other extreme, we can use the true offer rate of a school (that year in hindsight, unconditioned on student characteristics) as a proxy for this probability. In practice, a student's uncertainty is likely between those implied by these assumptions—they know their features (grades, location-based priorities), but not true population-level statistics. Where relevant, we indicate where our results depend on such an assumption: for example, in our calculation of "optimal" application perturbations, we use the unconditional offer rate from the previous year to calculate optimal perturbations, but then measure the decrease in undermatching using true counterfactual outcomes in hindsight; our improvement measurement is thus an underestimate since it does not use applicant-level information available at the time of the recommendation. We further note that likelihood uncertainty changes over time, as a function of policy and

We use the *offer rate* of a program (unconditioned on student characteristics) as a proxy for this probability. For most of our results, we use the offer rate of the application year, a hindsight measure; when calculating optimal ex ante recommendations, we use the prior year's offer rates, to avoid information leakage.

Given offer likelihood and match outcomes, we define our three deviations from a reach-match-safety approach as follows.

1. (No Reach Program) When a student ranks a 100% offer rate program first. This is an extreme case of not reaching, as the student ranks first a program that all students would match to.
2. (No Match Program) When a student is matched with a 100% offer rate program, after getting rejected from earlier-ranked programs. This means that a student applied to programs that were too risky (that they were not matched to) or too safe (one with a 100% offer rate).
3. (No Safety Program) When a student gets rejected from all ranked programs. This pattern suggests that the student did not list a sufficiently unselective program.

We note that the three defined behaviors are mutually exclusive.

4.5 Deriving recommendations

Theoretical model. We first introduce a theoretical model of undermatching as a function of portfolio decisions. The model uses the cutoff characterization of stable matching [18], in which each program has a cutoff $\tau_i \in [0,1]$, and a student receives an offer from that program if and only if their score S_i at that program is *below* the cutoff. Then, a student is matched to the highest-ranked program in their list from which they receive an offer.

To make this model more empirically tractable, we adopt the "common score" assumption of Ali and Shorrer [11]. That is, each student has a score $S \in [0,1]$ that is used across all programs. A key property of this model is that student offers are correlated: being rejected from a selective program means that the student is unlikely to receive an offer from another selective program. This reflects our empirical setting, in which most programs use a combination of lottery number and grade tier to rank students, meaning that student outcomes are correlated.

Under the common score assumption, if a student applies to a list of programs with *thresholds* $\tau_1, \tau_2, \cdots, \tau_k$, then they are matched to the program in position $\min\{i \in [k] : S < \tau_i\}$, the program in the earliest position i in their list for which their score is less than the threshold, $S < \tau_i$. Let u_i be the quality (i.e., as measured by offer rate or school performance) of the program ranked i in the student's portfolio. Then, assuming S is distributed uniformly at random on $[0,1]$, the expected quality of a student's match is

$$\int_0^1 u_{\min\{i \in [k] : S < \tau_i\}} \, dS. \tag{1}$$

Ali and Shorrer [11] demonstrate that optimal portfolios (according to the above objective) contain programs with a range of thresholds—roughly speaking, implementing a "reach-match-safety" strategy. Moreover, optimal portfolios can be computed efficiently. Under student proposing deferred acceptance, optimal portfolios must also be rank-ordered by offer rate.

Using the model to make recommendations. In our empirical setting, we study the model in which $u_i = R_i$, since we are interested in making recommendations that minimize expected undermatching (equivalent to minimizing offer rate). We also employ a heuristic, in which we set the threshold τ_i to be the offer rate R_i of the program. We use the offer rate of programs computed from the *previous* year. This captures the basic property of thresholds: that programs with lower thresholds are more selective. Then, the objective in our heuristic recommender is to choose a portfolio that minimizes

$$f(R) := \int_0^1 R_{\min\{i \in [k] : S < R_i\}} \, dS. \tag{2}$$

published information. In the admissions year analyzed (but not previous years), all students received their lottery number and the use of grades was simplified compared to previous years.

The model can then be used to estimate the expected offer rate of a student's match, given their portfolio. Therefore, we can recommend portfolios that improve the expected offer rate. (Minimizing expected offer rate also minimizes expected undermatching.) More specifically, we recommend a *single change* to a student's portfolio: to replace one program in their portfolio with a program of a different offer rate. Define $R(m, r)$ to be the portfolio R except with R_m replaced by an offer rate r. Then for a given portfolio R, we would like to make a recommendation (m, r) such that

$$f(R(m, r)) \tag{3}$$

is minimized. For example, for a student with portfolio $R = [0.6, 0.7, 0.5]$, the model's recommendation would be $(m, r) = (1, 0.35)$—that is, to change the offer rate of the first program to 0.35, so that $R(m, r) = [0.35, 0.7, 0.5]$. For each student, we first find such an (m, r). Note that this new portfolio is still suboptimal, as evidenced by the non-monotonicity; however, it is the best one can do to improve expected match quality while constrained to make only one change to the student's portfolio. (We show how we compute this optimal pair in Appendix A.1. Roughly, the idea is to fill the largest "gap" in the increasing subsequence of offer rates.) Then, to implement the recommendation implied by an optimal (m, r), we find a geographically nearby program to the student whose offer rate approximates the recommendation. For example, in the example above, we find the program that has an offer rate closest to 0.35 to substitute into the first position of the student's portfolio. We only consider programs at least as close as the student's matched program, as this approximates a student's distance constraint. We also only consider programs for which we can compute counterfactuals, and for which at least 25 students matched in the previous year (so that the offer rate of the previous year is a better estimate of the current year's offer rate). Finally, we evaluate the counterfactual outcome in hindsight, assuming the student adopted the recommendation: we find the earliest program in a student's list from which they would have received a counterfactual offer, and evaluate undermatching between this offer and their counterfactual best. We evaluate counterfactual undermatching using the current year's offer rates.

5 Related work

There is a large literature on centralized school matching and in NYC specifically [1, 3–5, 11, 14, 25, 37, 38, 55, 59, 63–65, 69, 70, 72, 73]. Our work sits at the intersection of two large literatures: (1) empirically understanding disparities in application behavior and guiding behavior through informational interventions, and (2) theoretically characterizing optimal application set behavior, under uncertainty of admissions.

Empirically understanding and intervening on application behavior. A substantial literature, both in NYC and other contexts, observes socioeconomic and demographic disparities in ranking behavior.

School matching in NYC. It is well known that the process favors privileged applicants, who have the resources to find and apply to "higher-quality" schools [22, 28]. Through surveys, Corradini and Idoux [29] find that "low-income and Black and Hispanic families appear to use fewer information sources and are less reliant on their social networks" when making application decisions, and Marian [56] highlights that many students apply to schools to which they have little admission chances due to their priority group. Correspondingly, a large literature has documented group-level disparities in application behavior [26–28, 30], e.g., that white and Asian, higher-income, and female students are more likely to apply to and match with schools with higher graduation rates, even conditional on test scores. A large literature implements and/or evaluates informational interventions to encourage students to apply to higher-performing schools, with mixed results on whether the interventions change behavior (see, e.g., Cohodes et al. [26] and references therein). These interventions are primarily impersonal (e.g., highlighting nearby high-graduation-rate schools), or personalized to students' stated preferences over school characteristics.

Most related to our work is that of Corradini [28] and Corradini and Idoux [30]. Corradini [28] studies racial/ethnic disparities in NYC high school admissions in terms of students applying to and matching with high value-add schools. Leveraging variation induced by NYC introducing in 2008 and then removing in 2014 school quality letter grades, she finds that Black and Hispanic (vs white and Asian) students choose lower-quality schools, and this cannot be fully explained by residential segregation or admissions policies—controlling for nearby schools to which a student has a nonzero match probability, Black and Hispanic

students' top three choices are lower quality on average than those of Asian and white students. Using counterfactual simulations, she finds that perfect information yields 23% of the gain of the first-best solution of students just applying to schools based on value-add, which "reduce[s] the mean cross-race achievement gap by 62% or .035σ." Corradini and Idoux [30] combine administrative and survey data from the NYC High School match to also study disparities in application behavior and match outcomes; like us, they identify racial/ethnic patterns in behavior (including in the "choice gap" as measured by offer rates and related measures) and match outcomes, even when controlling for neighborhood and middle school grades. Using survey data, they find that these differences are best explained by differential awareness of high-performing schools and preferences over the school's racial composition, as opposed to family preferences over other school characteristics, beliefs of admissions chances, or knowledge of school quality conditional on awareness. Leveraging randomness in middle school matches, they show that Black and Hispanic students who attend middle schools that are predominantly white or Asian in turn apply to similar schools as their middle school peers, illuminating the role of peers and social networks in school awareness and choice.

Finally, across multiple interventions, Cohodes et al. [26] find that a "a key mechanism by which interventions worked was through new information replacing students' default first-choice schools that had low graduation rates and guaranteed admission." We analyze the prevalence of such ranking effects within each applicant's entire portfolio, quantify its relationship to undermatching, and develop entire-portfolio-based feedback to help applicants implement a reach-match-safety approach. Other works also report such choice and matching disparities [60].

Undermatching and not matching in other contexts. In the context of college admissions in the United States, Tomkins et al. [77] find that the introduction of Naviance—which provides average test scores and grades of admitted students to each college—at a high school led to more conservative application behavior among the most competitive applicants, who may have been discouraged from applying to a college where their scores are below the average of admitted students; the authors suggest that an interface that emphasizes applying to a portfolio of reach-match-safety schools may prevent such behavior. Hoxby and Avery [47] find that a "vast majority of very high-achieving students who are low-income do not apply to any selective college or university," even though they would receive financial aid and are admitted and graduate at high rates when they do apply.

Outside of NYC, most related to our work methodologically is a line of work in school and college admissions in Chile [12, 31, 46]. Arteaga et al. [16], through a survey regarding high school matching, find that "costly search for schools is central to the way families experience choice" and "systematic overoptimism about admissions chances leads participants to submit applications with high nonplacement risk;" in both New Haven (in the United States) and Chile, they deploy a smart platform that gives personalized warnings to applicants at risk of *not matching* and find that it is effective. Larroucau et al. [53] conduct a multi-year collaboration for college admissions in Chile. Through surveys, they first document and taxonomize types of application mistakes, including mistakes of awareness, valuations, and admission probabilities (causing either unmatched or undermatching outcomes). They then implement and evaluate personalized feedback, including if the probabilities for being matched at all is below 99% or that of being matched to one's first choice is over 99%. Our work complements this analysis in two ways: (a) informed by prior survey work and school quality measures in NYC, we characterize the prevalence of similar application behaviors in actual portfolios and document how they explain disparities across groups, including undermatching caused by potential unawareness of high-quality options; (b) applying the theoretical model of Ali and Shorrer [11], we characterize "optimal" changes to individual lists, quantifying the relative effectiveness of heuristics that modify the applicant's first choice program or add safety programs to the end.

In New Orleans, Harris and Larsen [42] find via a mixed logit model that lower-income families have weaker preferences for academics and stronger preferences for extracurriculars. Hastings et al. [45] study the Charlotte Mecklenburg Public School District in North Carolina for elementary and middle school matching under the Boston (non-strategyproof) mechanism; they find that higher neighborhood income, those not receiving lunch subsidies, and (conditional on neighborhood income) non-white families, along with students with higher baseline test scores, placed more weight on school academic scores; non-white families had a larger preference for schools with more Black students. Agte et al. [8] study search costs, information access, and admission believes in Chile, and find that an intervention that highlighted the characteristics of nearby schools increased knowledge and enrollment in high-value add schools of high socioeconomic status (SES) households, but not of those with lower SES households.

At a higher level, the literature has established that applying to schools is a complex process, that incorrect beliefs and information access heterogeneously affect applicant behavior, and that informational interventions can steer such behavior—thus, even in "strategy-proof" settings, ranking behavior may not reflect "true" applicant preferences. To this literature, our analysis contributes (a) quantification of undermatching with respect to (almost fully known) counterfactual admission decisions, and so fully isolating the effects of application behavior from admission policies, and (b) connecting *individual*-level undermatching with empirical application ranking behavior (as motivated by the theoretical literature we discuss below), including the potential effects of application modifications. These behaviors—and the interventions we simulate—intermediate the relationship between information and outcomes. Thus, building on prior work that underscores the need to provide actionable information to applicants, helps design individual-level, personalized interventions feedback to steer application behavior, whether that behavior was due to informational awareness or other aspects such as beliefs about the value of strategic behavior.

Theoretical modeling of optimal application set decisions. In the theoretical literature, application portfolio selection has been studied in various settings. In general, this line of work considers a student choosing a limited subset from a larger set of schools of varying selectivities and utilities, under uncertainty of their strength as an applicant and/or chances of admission at each school.

Models of decentralized and/or simultaneous applications (such as the US college admissions system) often build on the canonical simultaneous search framework of Chade and Smith [25], in which every school has a (fixed, exogenous) acceptance probability, and admissions decisions are mutually independent. In this setting, the optimal size-k portfolio can be computed efficiently using a greedy algorithm or dynamic programming, and as k increases, it expands only upwards to include increasingly competitive schools. To model correlated admissions, Ali and Shorrer [11] introduce a threshold model in which the student has a single "common score" unknown to the student, and each school accepts students with scores above a (fixed, exogenous) threshold. Due to the use of the common score, admissions are completely correlated. Under this assumption, the optimal k-portfolio can be computed efficiently via dynamic programming, and demonstrates both upwards and downwards expansion (reflecting the notion of a reach-match-safety strategy).

However, field studies and lab experiments show that students often do not apply optimally in practice [40, 67]. A large literature has documented non-truthful behavior even with strategy-proof mechanisms [43], including with students choosing programs dominated by other options [44, 76]. One explanation that has been posited is *behavioral biases* such as loss aversion, but theoretical modeling of such behavioral phenomena has been limited. In a threshold-style model, Kleinberg et al. [51] show that rather than employing an optimal reach-match-safety strategy, biased students generally target less selective schools but may also occasionally overshoot (i.e. apply to schools that are *too* selective for them). In the stable matching context, Dreyfuss et al. [34] and Meisner and von Wangenheim [58] show that loss aversion may cause a student to misreport by inverting or excluding schools from their true ranking.

We adopt the common-score admissions model of Ali and Shorrer [11] in characterizing application rankings in terms of probability of admission at each program—and apply the framework to a rank-ordered deferred acceptance setting. Empirically, we find that students deviate substantially from reach-match-safety strategies, by applying to only selective schools, only safety schools, inverting selectivity ordering, and other deviations. We further calculate "optimal" (within the model of Ali and Shorrer [11]) local perturbations (changing one program) to student portfolios, and quantify their effect.

Structural modeling of ranking behavior A substantial literature has developed structural models to study applicant preferences from their ranking decisions [6, 9, 24]; see Agarwal and Somaini [7] for a review. Broadly, this literature attempts to model student preferences and behavior; as summarized by Agarwal and Somaini [7], the predominant approach is to model student behavior through a sequential search model, assuming either truthful, strategic, or boundedly rational applicants. For example, in the context of NYC middle school applications, Idoux [48] models students as making sequentially (greedy) optimal decisions given their search costs, allowing students to submit short lists and omit low probability-admissions schools. Larroucau and Rios [52] similarly find in Chile that modeling students as strategic (in a strategy-proof mechanism) better captures the data, and that preference over major subject differ by gender. Awadelkarim et al. [17] develop a choice model approach in which an applicant's choice for their kth rank school depends

on their first $k-1$ spots, to capture correlations due to latent preferences, such as an applicant who applies to one accessible school is likely to apply to another.

Ajayi and Sidibe [9] considers centralized high school assignment in Ghana (which uses serial dictatorship ordered by test scores), both through surveys and fitting dynamical search models using ranking data, in which students can endogenously decide whether to continue searching for more schools, given search costs. They analyze search, preference, and belief (of students uncertainty over school cutoffs and their own scores) parameters; stratifying by sex, they find that female students have weaker preferences for more competitive schools and stronger preferences for home economics versus business or general sciences programs. Calsamiglia et al. [24] and Agarwal and Somaini [6] compare welfare effects of a non-strategy-proof mechanism (the Boston Mechanism), allowing for a mixture of naive and sophisticated agents in a search model.

In contrast to this work, we do not estimate structural parameters from choice data, specifying how applicants may face search costs or make naive decisions given acceptance rate beliefs. Rather, we characterize heterogeneous student behavior in a "bottom-up" approach, and connect this heterogeneous behavior to individual-level outcomes and group-level disparities. As a result, our work cannot directly distinguish heterogeneous *preferences* from other heterogeneity (such as in beliefs, information, or search costs). On the other hand, our descriptive approach illustrates how *full portfolio* behaviors dictate undermatching. These behaviors are not captured by past work: for example, to our understanding, canonical sequential search models would not predict the observed level of *inversions* (ranking non-selective programs over selective ones), which explains about one-third of undermatching. As a result, our work provides empirical observation of heterogeneous behaviors that directly affect match outcomes. These behaviors may be modeled in future work. Moreover, our approach contributes a way to identify interventions that will most effectively decrease undermatching.

6 Data availability statement

The data was acquired through a data use agreement with the New York City Board of Education. Due to a non-disclosure agreement and FERPA-related protections, we cannot make the data publicly available. Interested researchers can apply to receive the data at https://infohub.nyced.org/in-our-schools/working-with-the-doe/research-irb/doing-research-in-new-york-city-public-schools.

References

[1] Atila Abdulkadiroglu, Parag A Pathak, Alvin E Roth, and Tayfun Sönmez. Changing the boston school choice mechanism: Strategy-proofness as equal access. *Mimeographed. Harvard University*, 2006.

[2] Atila Abdulkadiroglu, Joshua D Angrist, Yusuke Narita, and Parag A Pathak. Impact evaluation in matching markets with general tie-breaking. Technical report, National Bureau of Economic Research, 2018.

[3] Atila Abdulkadiroğlu, Parag A Pathak, and Alvin E Roth. The New York City High School Match. *American Economic Review*, 95(2):364–367, April 2005. ISSN 0002-8282. doi: 10.1257/000282805774670167. URL https://pubs.aeaweb.org/doi/10.1257/000282805774670167.

[4] 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, December 2017. ISSN 0002-8282. doi: 10.1257/aer.20151425. URL https://www.aeaweb.org/articles?id=10.1257/aer.20151425.

[5] Atila Abdulkadiroğlu, Parag A. Pathak, Jonathan Schellenberg, and Christopher R. Walters. Do Parents Value School Effectiveness? *American Economic Review*, 110(5):1502–1539, May 2020. ISSN 0002-8282. doi: 10.1257/aer.20172040. URL https://www.aeaweb.org/articles?id=10.1257/aer.20172040.

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

[7] Nikhil Agarwal and Paulo Somaini. Revealed preference analysis of school choice models. *Annual Review of Economics*, 12(1):471–501, 2020.

[8] Patrick Agte, Claudia Allende, Adam Kapor, Christopher Neilson, and Fernando Ochoa. Search and biased beliefs in education markets. Technical report, National Bureau of Economic Research, 2024.

[9] Kehinde Ajayi and Modibo Sidibe. School choice under imperfect information. *Economic Research Initiatives at Duke (ERID) Working Paper*, (294), 2020.

[10] Kehinde F Ajayi, Willa H Friedman, and Adrienne M Lucas. When information is not enough: Evidence from a centralized school choice system. Technical report, National Bureau of Economic Research, 2020.

[11] S Nageeb Ali and Ran I Shorrer. Hedging when applying: Simultaneous search with correlation. *American Economic Review (forthcoming)*, 2024.

[12] Claudia Allende, Francisco Gallego, and Christopher Neilson. Approximating the equilibrium effects of informed school choice. Technical report, 2019.

[13] Elaine M Allensworth, Paul T Moore, Lauren Sartain, and Marisa de la Torre. The educational benefits of attending higher performing schools: Evidence from chicago high schools. *Educational Evaluation and Policy Analysis*, 39(2):175–197, 2017.

[14] Maxwell Allman, Itai Ashlagi, Irene Lo, Juliette Love, Katherine Mentzer, Lulabel Ruiz-Setz, and Henry O'Connell. Designing school choice for diversity in the san francisco unified school district. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 290–291, 2022.

[15] Joshua D Angrist, Sarah R Cohodes, Susan M Dynarski, Parag A Pathak, and Christopher R Walters. Stand and deliver: Effects of boston's charter high schools on college preparation, entry, and choice. *Journal of Labor Economics*, 34(2):275–318, 2016.

[16] Felipe Arteaga, Adam J Kapor, Christopher A Neilson, and Seth D Zimmerman. Smart matching platforms and heterogeneous beliefs in centralized school choice*. *The Quarterly Journal of Economics*, 137(3):1791–1848, 03 2022. ISSN 0033-5533. doi: 10.1093/qje/qjac013. URL https://doi.org/10.1093/qje/qjac013.

[17] Amel Awadelkarim, Arjun Seshadri, Itai Ashlagi, Irene Lo, and Johan Ugander. Rank-heterogeneous preference models for school choice. In *Proceedings of the 29th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*, pages 47–56, 2023.

[18] Eduardo M Azevedo and Jacob D Leshno. A supply and demand framework for two-sided matching markets. *Journal of Political Economy*, 124(5):1235–1268, 2016.

[19] Michael N Bastedo and Allyson Flaster. Conceptual and methodological problems in research on college undermatch. *Educational Researcher*, 43(2):93–99, 2014.

[20] Diether W Beuermann, C Kirabo Jackson, Laia Navarro-Sola, and Francisco Pardo. What is a good school, and can parents tell? evidence on the multidimensionality of school output. *The Review of Economic Studies*, 90(1):65–101, 2023.

[21] Howard S Bloom and Rebecca Unterman. Can small high schools of choice improve educational prospects for disadvantaged students? *Journal of Policy Analysis and Management*, 33(2):290–319, 2014.

[22] Bliss Broyard. NYC high school admissions rewards parents like me, January 2023. URL https://ny.chalkbeat.org/2023/1/30/23574201/nyc-high-school-admissions-inequity-ethics. Section: Front Page.

[23] Caterina Calsamiglia and Maia Güell. Priorities in school choice: The case of the boston mechanism in barcelona. *Journal of Public Economics*, 163:20–36, 2018.

[24] 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.

[25] Hector Chade and Lones Smith. Simultaneous search. *Econometrica*, 74(5):1293–1307, 2006.

[26] Sarah R Cohodes, Sean P Corcoran, Jennifer L Jennings, and Carolyn Sattin-Bajaj. When Do Informational Interventions Work? Experimental Evidence from New York City High School Choice. *Educational Evaluation and Policy Analysis*, page 01623737231203293, 2022.

[27] Sean P Corcoran, Jennifer L Jennings, Sarah R Cohodes, and Carolyn Sattin-Bajaj. Leveling the playing field for high school choice: Results from a field experiment of informational interventions. Working Paper 24471, National Bureau of Economic Research, March 2018. URL http://www.nber.org/papers/w24471.

[28] Viola Corradini. Information and Access in School Choice Systems: Evidence from New York City. 2024. URL https://opportunityinsights.org/wp-content/uploads/2024/11/JMP_Corradini.pdf.

[29] Viola Corradini and Clemence Idoux. What matters in families' high school choices? results from a high school post-application survey of parents. Technical report, New York City Department of Education, May 2023. In collaboration with the NYCDOE Enrollment Research and Policy office, Carlos Flores, Brielle McDaniel, and Lianna Wright.

[30] Viola Corradini and Clemence M Idoux. Overcoming racial gaps in school preferences: The effect of peer diversity on school choice. Technical report, 2025.

[31] Jose Correa, Rafael Epstein, Juan Escobar, Ignacio Rios, Bastian Bahamondes, Carlos Bonet, Natalie Epstein, Nicolas Aramayo, Martin Castillo, Andres Cristi, et al. School choice in chile. In *Proceedings of the 2019 ACM Conference on Economics and Computation*, pages 325–343, 2019.

[32] David J Deming, Justine S Hastings, Thomas J Kane, and Douglas O Staiger. School choice, school quality, and postsecondary attainment. *American Economic Review*, 104(3):991–1013, 2014.

[33] Department of Education. 2023-24 educator guide to the high school quality guide. Technical report, New York City Department of Education, New York, NY, USA, January 2024. URL https://infohub.nyced.org/docs/default-source/default-document-library/2023-24-educator-guide-hs.pdf.

[34] Bnaya Dreyfuss, Ori Heffetz, and Matthew Rabin. Expectations-based loss aversion may help explain seemingly dominated choices in strategy-proof mechanisms. *American Economic Journal: Microeconomics*, 14(4):515–55, November 2022. doi: 10.1257/mic.20200259. URL https://www.aeaweb.org/articles?id=10.1257/mic.20200259.

[35] Umut Dur, Scott Duke Kominers, Parag A Pathak, and Tayfun Sönmez. Reserve design: Unintended consequences and the demise of boston's walk zones. *Journal of Political Economy*, 126(6):2457–2479, 2018.

[36] Itai Feigenbaum, Yash Kanoria, Irene Lo, and Jay Sethuraman. Dynamic matching in school choice: Efficient seat reassignment after late cancellations. *Management Science*, 66(11):5341–5361, 2020.

[37] Guillaume Haeringer and Flip Klijn. Constrained school choice. *Journal of Economic theory*, 144(5): 1921–1947, 2009.

[38] Dong Woo Hahm and Minseon Park. A dynamic framework of school choice: Effects of middle schools on high school choice. In *Proceedings of the 23rd ACM Conference on Economics and Computation*, pages 292–293, 2022.

[39] Chantal A Hailey. Racial preferences for schools: Evidence from an experiment with white, black, latinx, and asian parents and students. *Sociology of Education*, 95(2):110–132, 2022.

[40] Rustamdjan Hakimov and Dorothea Kübler. Experiments on centralized school choice and college admissions: a survey. *Experimental Economics*, 24(2):434–488, 2021. URL https://EconPapers.repec.org/RePEc:kap:expeco:v:24:y:2021:i:2:d:10.1007_s10683-020-09667-7.

[41] Nikole Hannah-Jones. Choosing a school for my daughter in a segregated city. *The New York Times Magazine*, June 2016. URL https://www.nytimes.com/2016/06/12/magazine/choosing-a-school-for-my-daughter-in-a-segregated-city.html.

[42] Douglas N Harris and Matthew Larsen. What schools do families want (and why). *Education Research Alliance for New Orleans*, 66, 2015.

[43] Avinatan Hassidim, Déborah Marciano, Assaf Romm, and Ran I Shorrer. The mechanism is truthful, why aren't you? *American Economic Review*, 107(5):220–224, 2017.

[44] Avinatan Hassidim, Assaf Romm, and Ran I Shorrer. The limits of incentives in economic matching procedures. *Management Science*, 67(2):951–963, 2021.

[45] Justine Hastings, Thomas J Kane, and Douglas O Staiger. Heterogeneous preferences and the efficacy of public school choice. *NBER working paper*, 2145:1–46, 2009.

[46] Justine S Hastings, Christopher A Neilson, Anely Ramirez, and Seth D Zimmerman. (un) informed college and major choice: Evidence from linked survey and administrative data. *Economics of Education Review*, 51:136–151, 2016.

[47] Caroline M Hoxby and Christopher Avery. The missing "one-offs": The hidden supply of high-achieving, low income students. Working Paper 18586, National Bureau of Economic Research, December 2012. URL http://www.nber.org/papers/w18586.

[48] Clemence Idoux. Integrating New York City schools: The role of admission criteria and family preferences. March 2023. Working Paper, Version: March 11, 2023.

[49] C Kirabo Jackson, Shanette C Porter, John Q Easton, and Sebastián Kiguel. Who benefits from attending effective high schools? Technical report, National Bureau of Economic Research, 2020.

[50] Adam J. Kapor, Christopher A. Neilson, and Seth D. Zimmerman. Heterogeneous Beliefs and School Choice Mechanisms. *American Economic Review*, 110(5):1274–1315, May 2020. ISSN 0002-8282. doi: 10.1257/aer.20170129. URL https://www.aeaweb.org/articles?id=10.1257/aer.20170129.

[51] Jon Kleinberg, Sigal Oren, Emily Ryu, and Éva Tardos. Modeling reputation-based behavioral biases in school choice. *arXiv preprint arXiv:2403.04616*, 2024.

[52] Tomás Larroucau and Ignacio Rios. Do "short-list" students report truthfully? strategic behavior in the chilean college admissions problem. *Preprint, submitted September*, 1(10.13140), 2018.

[53] Tomás Larroucau, Ignacio Rios, Anaïs Fabre, and Christopher Neilson. College application mistakes and the design of information policies at scale. *Unpublished paper, Arizona State University, Tempe*, 2024.

[54] Mariana Laverde et al. Distance to schools and equal access in school choice systems. *Unpublished Manuscript*, 2022.

[55] Jacob D Leshno and Irene Lo. The cutoff structure of top trading cycles in school choice. *The Review of Economic Studies*, 88(4):1582–1623, 2021.

[56] Amelie Marian. Algorithmic transparency and accountability through crowdsourcing: A study of the nyc school admission lottery. In *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*, pages 434–443, 2023.

[57] Amelie Marian. Decoding the NYC School Admission Lottery Numbers, March 2023. URL https://medium.com/algorithms-in-the-wild/decoding-the-nyc-school-admission-lottery-numbers-bae7148e337d.

[58] Vincent Meisner and Jonas von Wangenheim. Loss aversion in strategy-proof school-choice mechanisms. *Journal of Economic Theory*, 207:105588, 2023.

[59] Yusuke Narita. Match or mismatch? learning and inertia in school choice. 2018.

[60] Lori Nathanson, Sean P Corcoran, and Christine Baker-Smith. High school choice in new york city: A report on the school choices and placements of low-achieving students, 2013.

[61] NYC DOE. Fall 2021 High School Data, 2021. URL https://infohub.nyced.org/reports/admissions-and-enrollment/directory-data. InfoHub Directory Data.

[62] Hessel Oosterbeek, Sándor Sóvágó, and Bas van der Klaauw. Preference heterogeneity and school segregation. *Journal of Public Economics*, 197:104400, 2021.

[63] Parag A Pathak. What really matters in designing school choice mechanisms. *Advances in Economics and Econometrics*, 1(12):176–214, 2017.

[64] Parag A Pathak and Peng Shi. How well do structural demand models work? counterfactual predictions in school choice. *Journal of Econometrics*, 222(1):161–195, 2021.

[65] Parag A Pathak and Tayfun Sönmez. Leveling the playing field: Sincere and sophisticated players in the boston mechanism. *American Economic Review*, 98(4):1636–1652, 2008.

[66] Sean F Reardon. School segregation and racial academic achievement gaps. *RSF: The Russell Sage Foundation Journal of the Social Sciences*, 2(5):34–57, 2016.

[67] Alex Rees-Jones and Ran Shorrer. Behavioral economics in education market design: A forward-looking review. *Journal of Political Economy Microeconomics*, 1(3):557–613, 2023.

[68] Alex Rees-Jones, Ran Shorrer, and Chloe J Tergiman. Correlation neglect in student-to-school matching. In *Proceedings of the 21st ACM Conference on Economics and Computation*, pages 467–468, 2020.

[69] Samantha Robertson, Tonya Nguyen, and Niloufar Salehi. Modeling assumptions clash with the real world: Transparency, equity, and community challenges for student assignment algorithms. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pages 1–14, 2021.

[70] Samantha Robertson, Tonya Nguyen, Cathy Hu, Catherine Albiston, Afshin Nikzad, and Niloufar Salehi. Expressiveness, cost, and collectivism: How the design of preference languages shapes participation in algorithmic decision-making. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23, New York, NY, USA, 2023. Association for Computing Machinery. ISBN 9781450394215. doi: 10.1145/3544548.3580996. URL https://doi.org/10.1145/3544548.3580996.

[71] Kihana Miraya Ross. On the road to brown and beyond: Troubling integration, desegregation, and segregation in the fight for black educational equity, opportunity, and justice, 2021.

[72] Alvin E Roth. Deferred acceptance algorithms: History, theory, practice, and open questions. *International Journal of Game Theory*, 36:537–569, 2008.

[73] Alvin E Roth. What have we learned from market design? *The economic journal*, 118(527):285–310, 2008.

[74] School Diversity Advisory Group. Making the grade: The path to real integration and equity for nyc public school students. Technical report, 2019. URL https://docs.wixstatic.com/ugd/1c478c_4de7a85cae884c53a8d48750e0858172.pdf.

[75] Julian Shen-Berro. Nyc education department developing tool to help applicants gauge high school admissions chances. *Chalkbeat New York*, August 2024. URL https://www.chalkbeat.org/newyork/2024/08/22/nyc-education-department-developing-tool-to-help-high-school-admissions/. Accessed: 2024-10-20.

[76] Ran I Shorrer and Sándor Sóvágó. Dominated choices in a strategically simple college admissions environment. *Journal of Political Economy Microeconomics*, 1(4):781–807, 2023.

[77] Sabina Tomkins, Joshua Grossman, Lindsay Page, and Sharad Goel. Showing high-achieving college applicants past admissions outcomes increases undermatching. *Proceedings of the National Academy of Sciences*, 120(45):e2306017120, 2023.

A Supplemental Materials

Contents

A.1 Computing Recommendations . . . 30
A.2 Additional details on interpretable sources of undermatching . . . 31
A.2.1 Rank order-related behaviors . . . 31
A.3 Undermatching by race and ethnicity, sex, and FRL status . . . 32
A.4 Analyses using other measures of program quality . . . 35
A.5 Matching and undermatching regressions . . . 48
A.5.1 Offer rate regressions . . . 48
A.5.2 Performance regressions . . . 51
A.5.3 Impact regressions . . . 54
A.5.4 Graduation rate regressions . . . 57
A.5.5 College enrollment rate regressions . . . 60
A.5.6 Aggregate quality regressions . . . 63
A.6 Consistency of program offer rates across years . . . 66
A.7 Program quality metrics by position in list . . . 66
A.8 The effect of lottery numbers . . . 66

A.1 Computing Recommendations

We now describe how to compute optimal recommendations under the model described in Section 4.5.

Consider a portfolio $R = (R_1, \cdots, R_k)$. We may extend the portfolio to $R = (R_0, \cdots, R_{k+1})$ where $R_0 = 0$ and $R_{k+1} = 1$. Choose $0 = i_0 < i_1 < \cdots < i_L$ such that $0 = R_{i_0} < R_{i_1} < \cdots < R_{i_L} = 1$ is the greedily-selected increasing subsequence of R. Then the first important observation is that

$$f(R) = \sum_{i=1}^{L} R_i(R_i - R_{i-1}). \tag{4}$$

Our approach is to find for each $m \in [k]$, r that minimizes $f(R(m, r))$. We can then choose the highest scoring pair (m, r).

For a given m, let ℓ be the largest integer such that $i_\ell < m$. Then set $R_{low} := R_{i_\ell}$. Clearly the optimal r must be larger than R_{low}, since otherwise the student will never match to that program. Now set ℓ' to be the smallest integer such that $i_{\ell'} > m$ and $R_{i_{\ell'}} > r$. Set $R_{high} := R_{i_{\ell'}}$ fixed. The key claim is that

$$f(R(m, r)) > f(R(m, (R_{low} + R_{high})/2)). \tag{5}$$

To see this, observe from (4) that

$$f(R(m, (R_{low} + R_{high})/2)) - f(R(m, r)) \leq g((R_{low} + R_{high})/2) - g(r) \tag{6}$$

where

$$g(x) := x(x - R_{low}) + R_{high}(R_{high} - x) \tag{7}$$

$$= R_{high}^2 + x(x - R_{low} - R_{high}). \tag{8}$$

(The inequality in (6) is due to the fact that the change from r to $(R_{low} + R_{high})/2$ may add an additional element into the increasing subsequence between $(R_{low} + R_{high})/2$ and R_{high}, which only reduces f.) The claim follows from the observation that $g(x)$ is minimized at $x = (R_{low} + R_{high})/2$.

Now observe that R_{high} must be in $\{R_{i_\ell+1}, R_{i_\ell+2}, \cdots, R_{i_L}\}$. We can then check $f(R(m, r))$ for $r = (R_{low} + R_{high})/2$ for all such R_{high}) to find the optimal choice of r.

(One can apply the same reasoning to find that the *optimal* portfolio of k programs consist of offer rates $\frac{1}{k+1}, \cdots, \frac{2}{k+1}, \cdots, \frac{k}{k+1}$.)

A.2 Additional details on interpretable sources of undermatching

Here, we elaborate on results described in Section 2.2.

No reach program. An extreme case of not reaching is when a student lists a program with a 100 percent offer rate first. The behavior guarantees that the student will match to this program, regardless of the remainder of their ranking. We find that Black and Hispanic students are much more likely to apply to a 100% offer rate program first in their list, guaranteeing their placement at the program, in comparison to Asian and white applicants (25.3% and 18.4% vs. 5.1% and 5.2%).

This type of behavior falls under a broader class of behavior, in which a student does not apply to selective programs early in their list. For example, relative to Asian and white students, Black and Hispanic students on average list a first choice with a higher offer rate (A: 21.5%, B: 48.9%, H: 44.3%, w: 26.1%) and are more likely to match to their top-ranked program (A: 32.0%, B: 54.3%, H: 52.5%, w: 40.5%). Intuitively, students who matched to their top-ranked program could have afforded to "reach further" by applying to a more selective program.

No match program. Students who applied to programs that were too selective (too selective in hindsight, as the student did not match there) followed by a program that was unselective (100% offer rate). Such students may have matched to a more selective program had they included a program with a "medium" selectivity, above the unselective program to which they matched. This behavior is (relatively speaking) prevalent across all groups (A: 12.4%, B: 20.5%, H: 17.1%, w: 9.3%).

No safety program. One measure of failing to apply to safety programs is simply that a student did not match to any of their choices; overall, 8% of students do not match to any program. Another measure of not listing a safety program is whether the student did not list any program with a high offer rate. Asian and white applicants are much more likely to not apply to any programs with an offer rate of at least 90% than Black and Hispanic students (56.3% and 62.1% vs. 18.0% and 26.2%). 21.3% of these students fail to match, compared to less than 0.1% of students who apply to a program with an offer rate of at least 90%. We note that one limitation of this analysis is that we do not yet have access to where the student ultimately enrolled; it is possible that such students did not list any safety programs because they preferred their outside option (e.g., a private school) above any such program. Our measure of undermatching does not reflect such circumstances; assuming that otherwise privileged students are more likely to have access to such outside options, our work would then underestimate undermatching disparities.

A.2.1 Rank order-related behaviors

Implicit in our above analysis is the role of deferred acceptance: students match to the program that they ranked highest in their list and which accepted them. Then, students may have counterfactually been accepted to a program they ranked lower down their list, if they had instead ranked it higher. Therefore, the rank order is important. Even if a student lists a portfolio of reach, match, and safety programs, the ordering of these programs can determine the practical effect of the portfolio. We now study ranking based behaviors directly.

Ranking Inversions: Listing selective programs below nonselective ones. As illustrated in an example portfolio in Figure 5, students may match to a less selective program, when there was a program later in their list that was more selective and that they would have been admitted to. We find that 27.0% of students list at least one program in their list after their matched program that is both more selective and at least as nearby. This number is somewhat higher for Black and Hispanic students (29.7%, 29.4%). As shown in Figure 2a, on average, students of all groups had a program in their portfolio (that they would have matched to) with a 10 p.p. lower offer rate than the program they ultimately matched to. (While inversions are prevalent, on average, students list higher quality programs earlier in their list; see Appendix A.7.)

List length and effective list length. In New York City, students historically were allowed to apply to at most 12 programs. (Starting in 2024-2025, students are allowed to list as many programs as they like). As shown in Figure 6, students from across groups list around 8 programs on average (A: 8.6, B: 9.0, H: 8.5, w: 7.6); only 38.8% of students have full lists. Of students with full lists, 5.1% do not match, suggesting that some students would directly benefit from being able to apply to more programs; this number is 1.7% and 2.7% for Black and Hispanic students, and 9.6% and 10.8% for Asian and white applicants.

However, as discussed above, list *length* does not matter by itself, rank *order* also does. We define *effective list length* as the number of programs up and including the first nonselective program (with 100% offer rate) in the list, or the total number of listed programs if the student does not list any unselective program. This measure captures the idea that it does not matter what a student lists after listing a fully nonselective program. Listing such programs after a nonselective program can also be seen as an extreme form of correlation neglect, cf. [68]. While having similar or longer length lists on average, Black and Hispanic applicants have smaller *effective* list lengths (4.1 and 4.7, versus 7.0 for Asian and 6.4 for white applicants).

A.3 Undermatching by race and ethnicity, sex, and FRL status

We provide additional visualizations of disparities in undermatching by race and ethnicity, sex, and FRL eligibility status.

Figure 8 shows undermatching as stratified jointly by race/ethnicity and FRL eligibility; neither characteristic alone fully explains undermatching, and there are disparities both by race/ethnicity and FRL eligibility. Figure 9 shows histogram and cumulative distribution functions of undermatching by race/ethnicity, sex, and FRL status; crucially, there is both substantial inter- and intra- group variance.

Figure 10 (for all students) and Figure 11 (limited to Asian students) shows undermatching stratified by the student's at-home language. There are substantial differences in both offer rates and undermatching by at-home language. In general, those who speak Mandarin, Chinese/Any, or Cantonese have the lowest undermatching, and this is true both for all students and when limited to just Asian students (who comprise the majority of several of the non-English at-home language groups such as Mandarin, Chinese/Any, Bengali, and Cantonese).



Figure 8: *Undermatching across race/ethnicity and additionally stratified by FRL eligibility.*

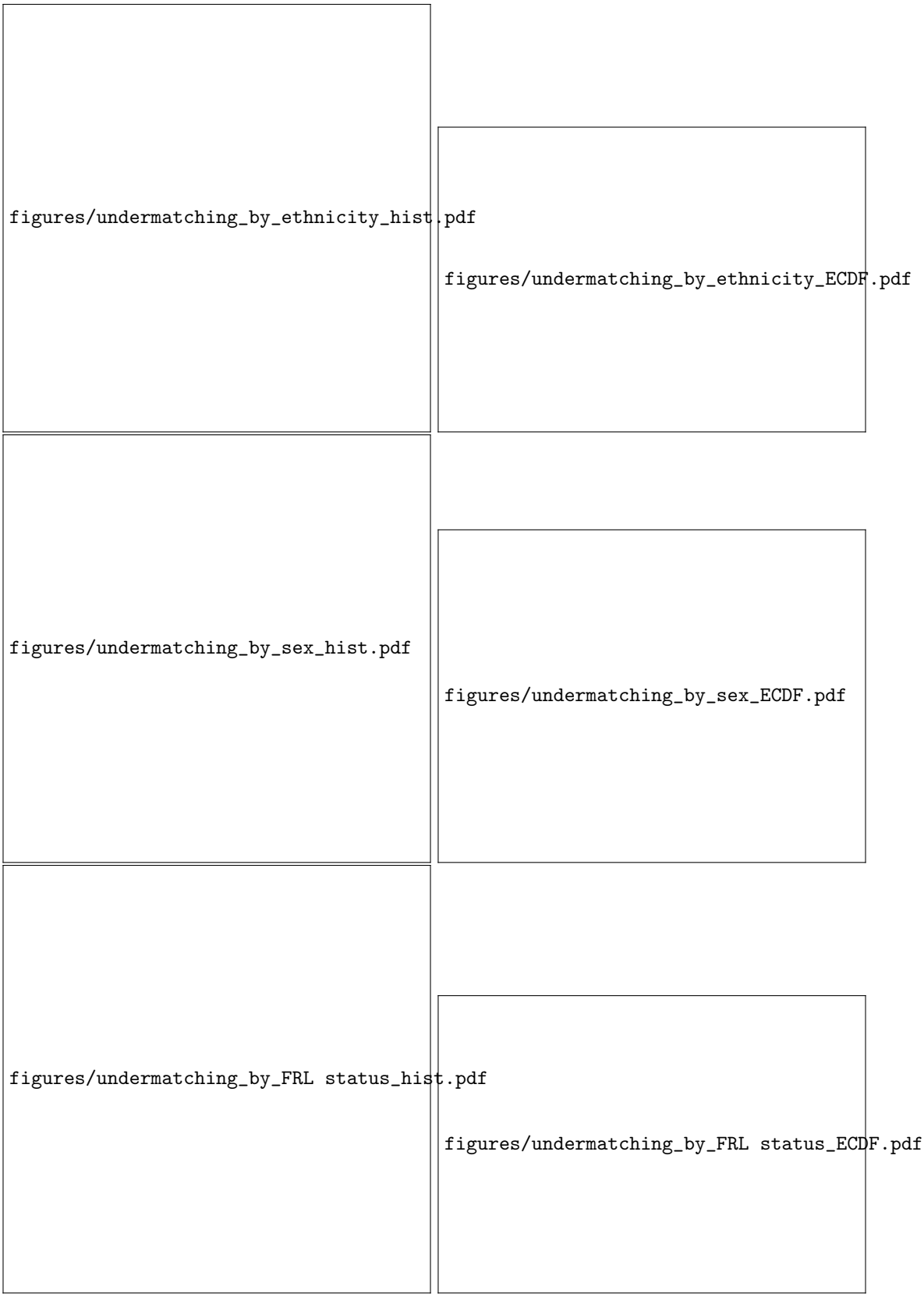


Figure 9: *Histograms and empirical cumulative distributions of undermatching to further visualize disparities by race/ethnicity, sex, and FRL eligibility. Crucially for our personalized approach, there is both substantial inter- and intra-group variance.*

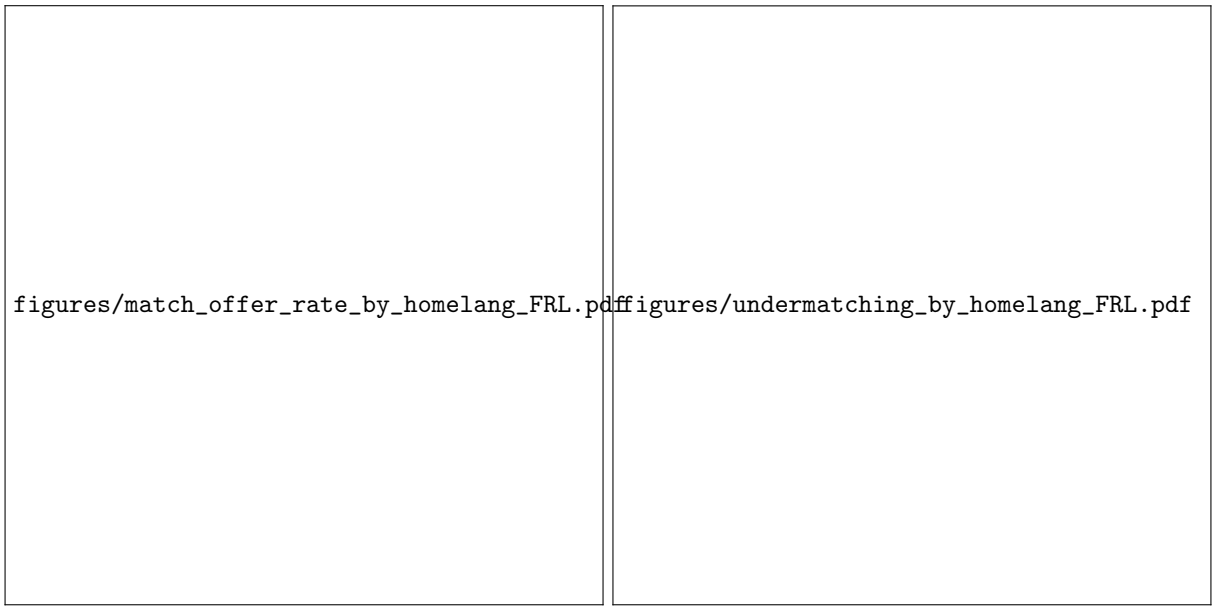


Figure 10: *Match selectivity and and undermatching by language spoken at home, measured by program offer rate. Error bars show 95% nonparametric bootstrap CI.*

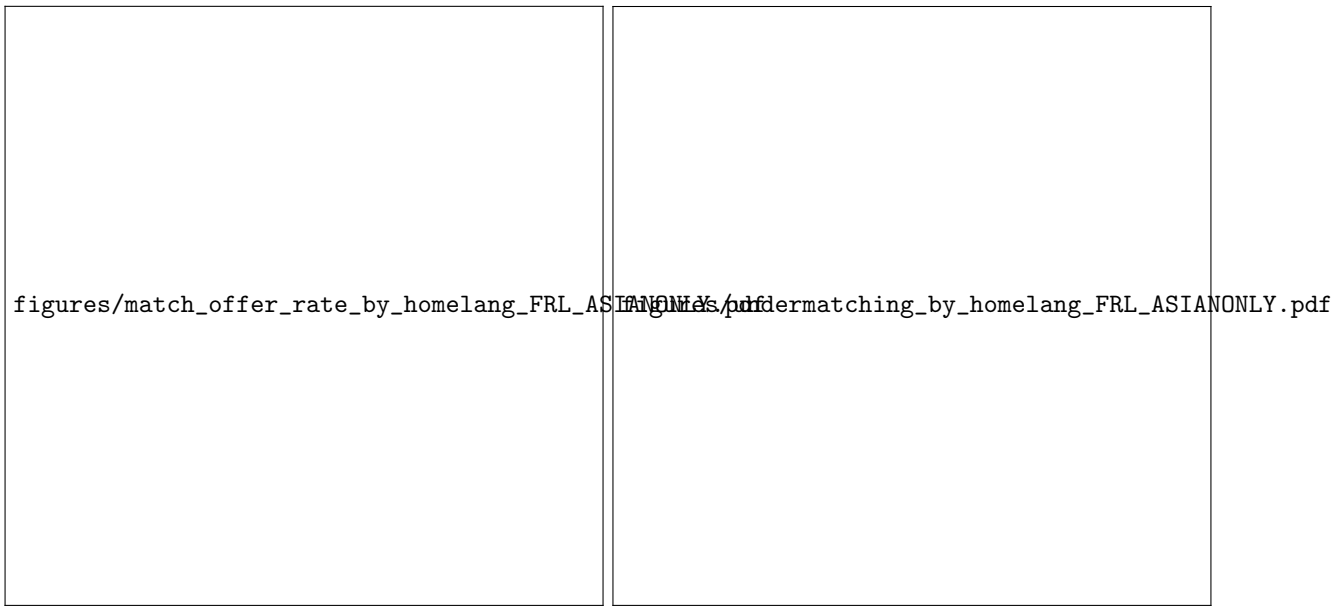


Figure 11: *Match selectivity and and undermatching by language spoken at home (measured by program offer rate), restricted to Asian students only. Error bars show 95% nonparametric bootstrap CI.*

A.4 Analyses using other measures of program quality

Here we replicate the main text analysis using alternate metrics of program quality: performance, impact, graduation rate, college enrollment rate, and an aggregate measure of NYC provided scores, to help reflect the dimensions identified as important by Corradini and Idoux [30]. The metric is calculated as the average of a school performance metric (the average of the DOE's measure of student performance and impact, both standardized) and the standardized average of six DOE survey measures of school quality (rigorous instruction, collaborative teachers, supportive environment, effective school leadership, strong family-community ties, and trust).

Figure 12 replicates Figure 2, our primary undermatching and offer rate figure. Figure 13 further splits up undermatching by race/ethnicity and FRL eligibility, replicating Figure 8. Figure 14 replicates Figure 3, undermatching by zip code. Figure 15 replicates Figure 4, undermatching by application features such as lottery number, grade tier, and best counterfactual program.

Figure 16 shows the mean improvement by group on each metric when single recommendations are made to improve match selectivity. Figure 18 shows the same when each student applies to a "optimal" portfolio according to the model of Ali and Shorrer [11].

In general, all results qualitatively replicate, as these metrics are all closely correlated with offer rate, as shown in Figure 22. Note that for an unmatched student, we impute their matched metrics to correspond to their manual placement program assignment (rather than a fixed constant like an offer rate of 1).

Figure 15, undermatching for other metrics by application characteristics, shows additional nuance from Figure 4. The metrics differ in whether undermatching is highest for the most competitive or least competitive students. For example, there are many programs with high graduation rates (compared to the number of programs with high selectivity); thus, even students with uncompetitive lottery numbers *could* be admitted to at least one program with high graduation rates—but they may not be able to find those programs ex ante. Generally, however, disparities between groups tend to be highest for the most competitive applicants.



Figure 12: *Replication of Figure 2. Performance metrics and undermatching across race/ethnicity, measured by program performance, impact, graduation rate, college enrollment rate, and an aggregate score of performance and environment scores, and additionally stratified by FRL eligibility. Observe that undermatching patterns by alternate metrics of program selectivity/quality closely replicate the analyses using offer rate. Error bars show 95% nonparametric bootstrap CI.*



Figure 13: *Undermatching across race/ethnicity and additionally stratified by FRL eligibility. Observe that these alternate metrics of program selectivity/quality closely replicate the analyses using program offer rate in Figure 8. Error bars show 95% nonparametric bootstrap CI.*

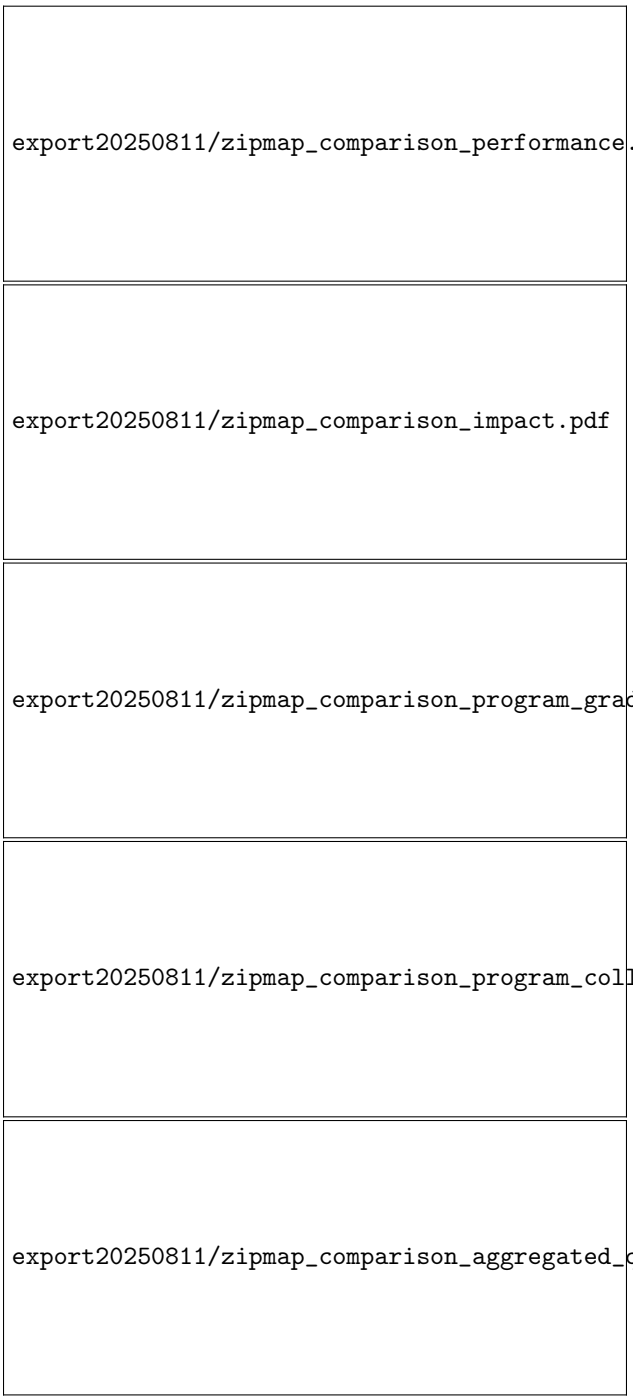


Figure 14: *Replication of Figure 3. Patterns are qualitatively identical. Error bars show 95% nonparametric bootstrap CI.*

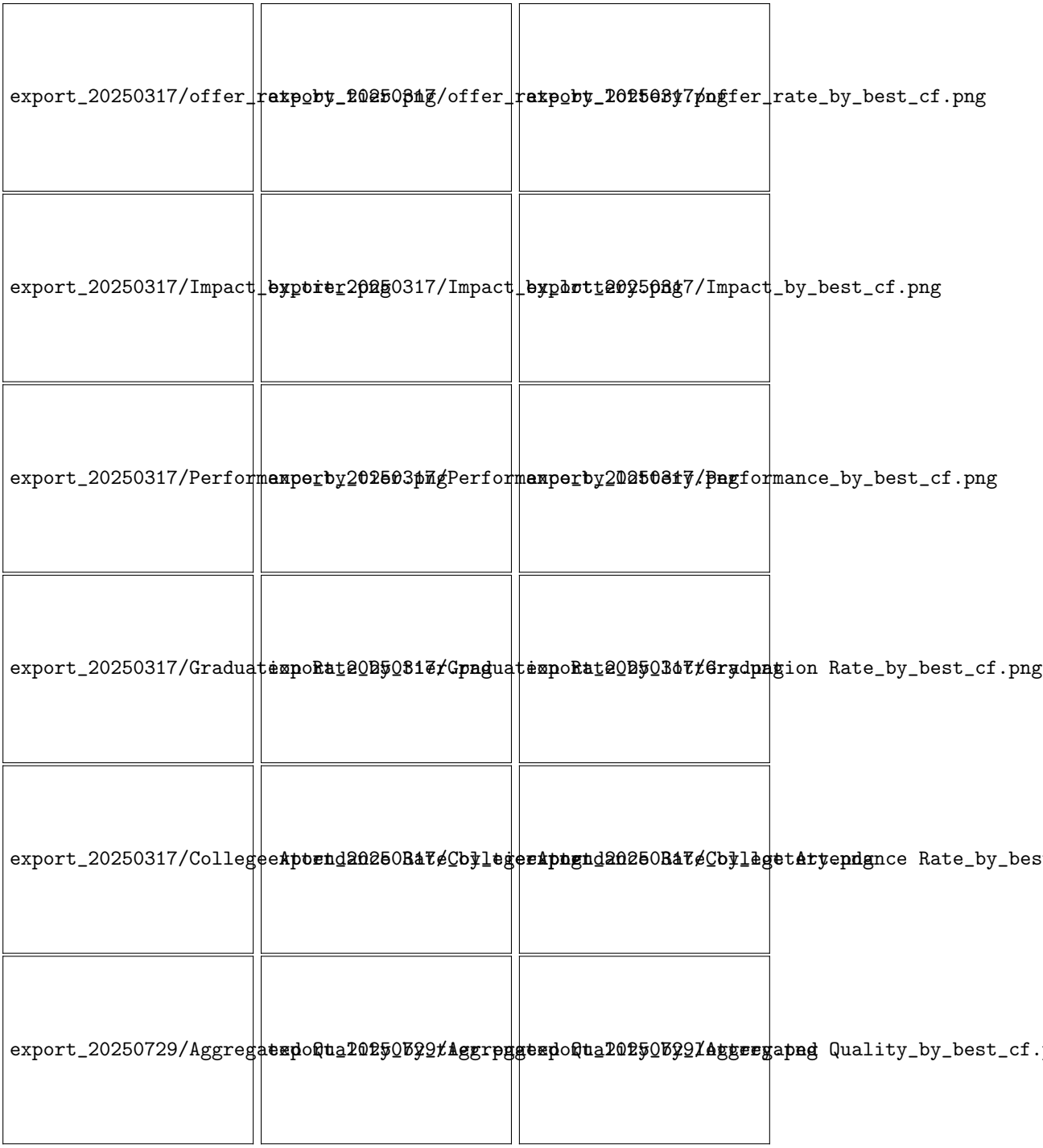


Figure 15: *Replication of Figure 4 for other metrics of undermatching and student grade tier. Trends and disparities in undermatching across student grade tier, lottery number, and offer rate of most selective counterfactual. Each row is a different measure of undermatching. From top to bottom: offer rate, impact, performance, graduation rate, college rate. Error bars show 95% nonparametric bootstrap CI.*



Figure 16: *Improvement in match after single recommendation on each metric. Note that the recommendations are all optimized to improve offer rate in expectation (The same recommendations are analyzed in each figure). The recommendations would reduce undermatching with respect to all the metrics for almost all groups (except several metrics for white applicants not eligible for Free/reduced lunch). Error bars show 95% nonparametric bootstrap CI.*

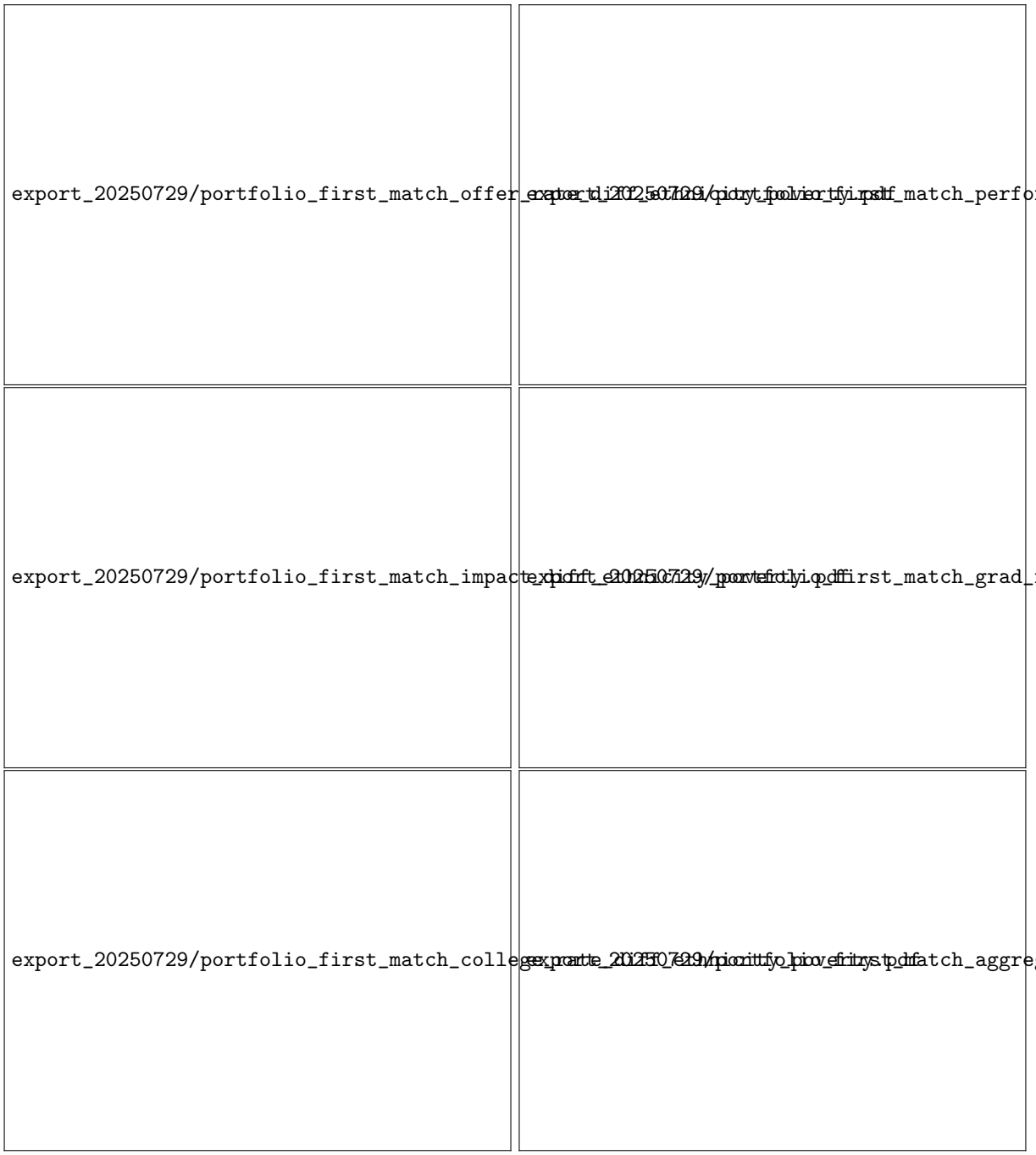


Figure 17: *Improvement in match after single recommendation, restricted to reaching further in the first choice. Note that the recommendations are all optimized to improve offer rate in expectation (The same recommendations are analyzed in each figure). The recommendations would reduce undermatching with respect to all the metrics for almost all groups (except several metrics for white applicants not eligible for Free/reduced lunch). Error bars show 95% nonparametric bootstrap CI.*



Figure 18: *Improvement in match after change to 12 recommended programs. Note that the recommendations are all optimized to improve offer rate in expectation. For the other metrics, higher numbers are better. Interestingly, while suggesting one change improves almost all metrics for all groups, moving to a fully optimal ranking (using population level admissions information) worsens outcomes for Asian and white applicants according to the other metrics. Error bars show 95% nonparametric bootstrap CI.*



Figure 19: *Change in equilibrium match after single recommendation on each metric. Note that the recommendations are all optimized to improve offer rate in expectation (The same recommendations are analyzed in each figure). The recommendations would reduce undermatching with respect to all the metrics for almost all groups (except several metrics for white applicants not eligible for Free/reduced lunch). Error bars show 95% nonparametric bootstrap CI.*



Figure 20: *Change in equilibrium match after single recommendation, restricted to reaching further in the first choice. Note that the recommendations are all optimized to improve offer rate in expectation (The same recommendations are analyzed in each figure). The recommendations would reduce undermatching with respect to all the metrics for almost all groups (except several metrics for white applicants not eligible for Free/reduced lunch). Error bars show 95% nonparametric bootstrap CI.*



Figure 21: *Change in equilibrium match after change to 12 recommended programs. Note that the recommendations are all optimized to improve offer rate in expectation. For the other metrics, higher numbers are better. Interestingly, while suggesting one change improves almost all metrics for all groups, moving to a fully optimal ranking (using population level admissions information) worsens outcomes for Asian and white applicants according to the other metrics. Error bars show 95% nonparametric bootstrap CI.*

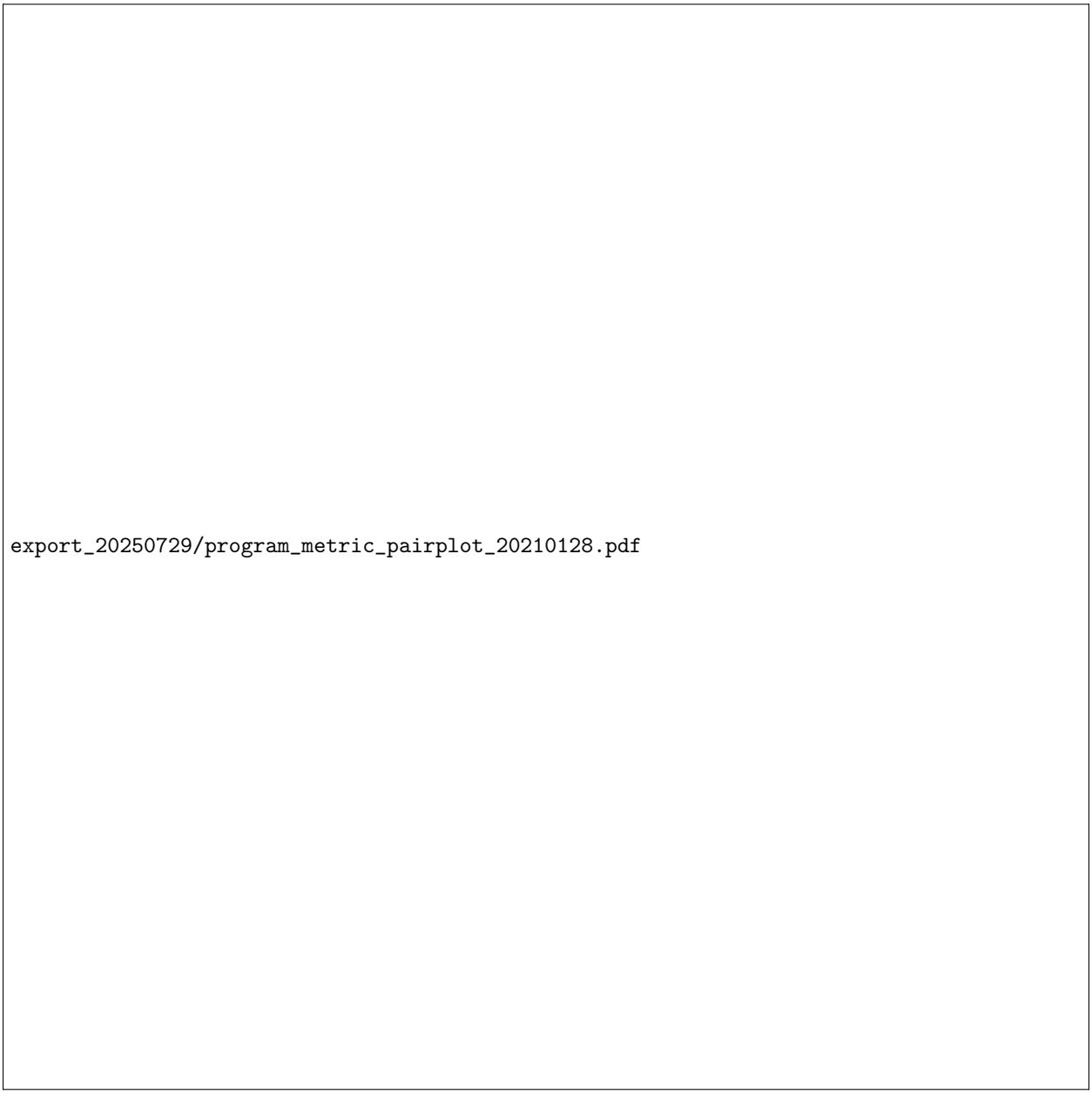


Figure 22: *Program offer rate correlates with performance, impact, graduation rate, college enrollment rate, and an aggregate score of performance and environment scores.*

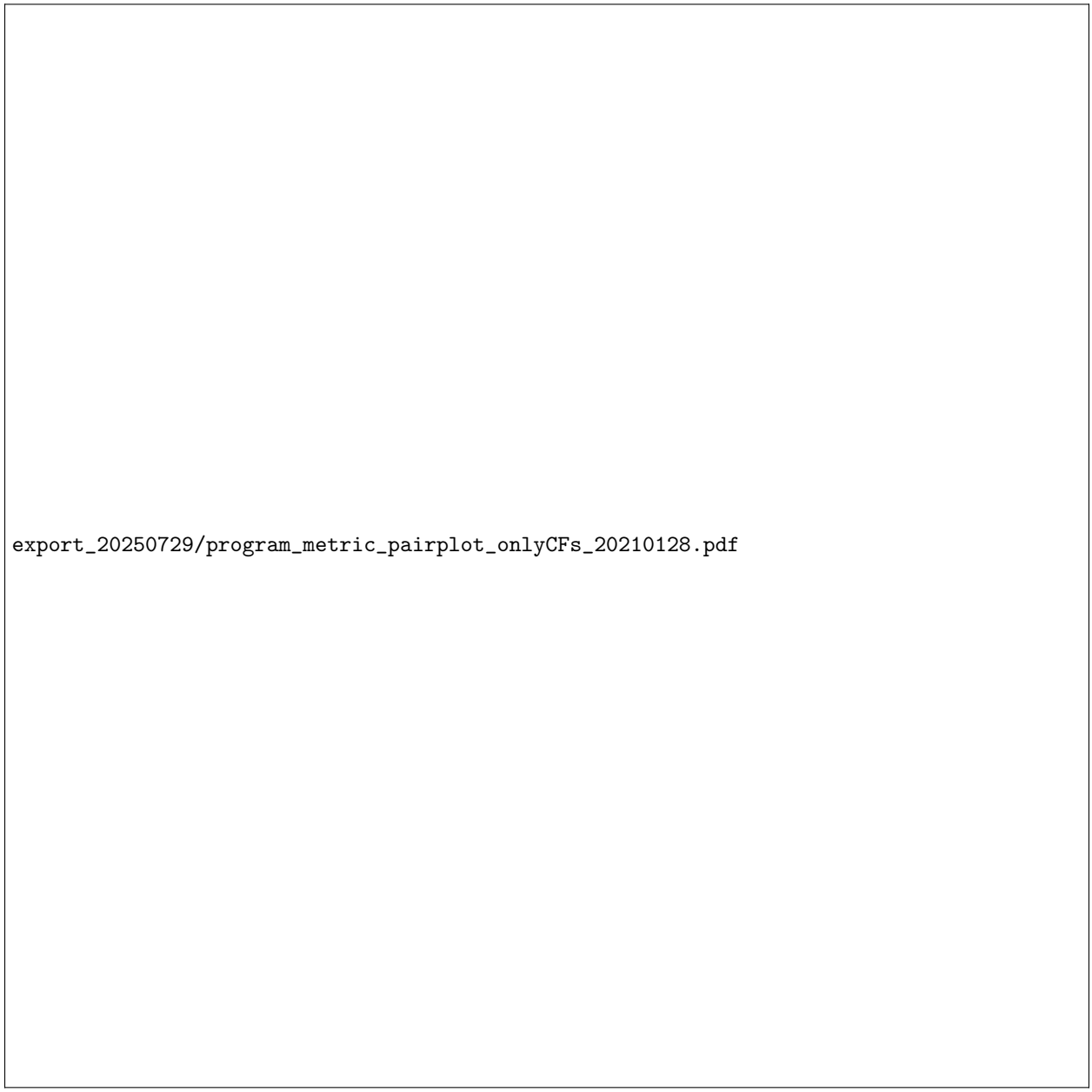


Figure 23: *Program offer rate correlates with performance, impact, graduation rate, college enrollment rate, and an aggregate score of performance and environment scores, among programs for which counterfactuals were computed.*

A.5 Matching and undermatching regressions

Here, we show the full regressions in which we assess the degree to which student and portfolio features explain undermatching.

Out of the 58,901 GE students in the 2022 cohort, 1,570 students did not have middle school grades available; 251 students did not have geographic/transit time data available; 981 students were matched or manually placed to a program without an available performance/impact score (thus accounting for the lower number of observations in these regressions). Prior to regressing, all numerical features were independently standardized (subtracting the mean and dividing by the standard deviation).

In addition to the interpretable portfolio behaviors we analyze in the main text, we also include a feature based on a clustering of portfolios. To compute these clusters, we represented each student's portfolio in terms of its *sorted* offer rates. If a student listed fewer than 12 programs, we appended 1's to the end of the portfolio, so that all portfolios were represented by a vector in $[0, 1]^{12}$. We then applied K-means clustering to these portfolios. These clusters capture the behavior of students in terms of the extent to which they applied to reach, match, and safety programs. They do not, however, account for ordering effects.

Note that, unlike in the writing and the figures, here we do not normalize undermatching to be defined as to be positive. To be consistent with the regressions for explaining each program quality metric (instead of undermatching), in our regressions against undermatching we define undermatching to be quality-metric(actual match) - quality-metric(best counterfactual). Thus, for offer rate (where lower offer rates is assumed to be higher quality), undermatching is positive, and higher values indicates more undermatching. For the remaining metrics (where higher values are higher quality in the metric), undermatching is negative, and more negative values indicate more undermatching. In all cases, higher magnitude undermatching is worse.

Finally, we also include regressions where the portfolio features (no-match and no-safety) are defined in a way that does not depend on the match outcome itself, to prevent any leakage between features and the target. Recall that in the main text, 'no safety' was defined as an applicant going unmatched, and 'no match' program was defined as an applicant being matched to a program with offer rate 1, after getting rejected by more selective programs. We further include regressions with just outcome independent measures:

- No safety (outcome independent): The ranked list does not contain any program with offer rate greater than 90%.
- No match program (outcome independent): The ranked list does not contain any program with offer rate between 30% and 60%.

Results are qualitatively similar; explained variance is lower, but still substantially higher than when using demographics alone.

A.5.1 Offer rate regressions

	coef	**std err**	**t**	**P> |t|**	**[0.025**	**0.975]**
Intercept	0.5377	0.004	123.638	0.000	0.529	0.546
male	0.0132	0.003	4.709	0.000	0.008	0.019
Black	0.1370	0.005	30.366	0.000	0.128	0.146
Hispanic	0.0935	0.004	23.458	0.000	0.086	0.101
Multiple races/ethnicities or Other	0.0745	0.007	11.414	0.000	0.062	0.087
white	0.0075	0.005	1.531	0.126	-0.002	0.017
FRL status	0.0310	0.003	9.280	0.000	0.024	0.038
avg MS grades	-0.0742	0.001	-49.524	0.000	-0.077	-0.071

Table 2: *OLS Regressions for match offer rate regressed against individual characteristics.* $R^2 = 0.096$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.2032	0.004	56.237	0.000	0.196	0.210
male	-0.0036	0.002	-1.546	0.122	-0.008	0.001
Black	0.0929	0.004	24.780	0.000	0.086	0.100
Hispanic	0.0689	0.003	20.812	0.000	0.062	0.075
Multiple races/ethnicities or Other	0.0658	0.005	12.136	0.000	0.055	0.076
white	0.0024	0.004	0.585	0.558	-0.006	0.010
FRL status	0.0204	0.003	7.340	0.000	0.015	0.026
avg MS grades	-0.0269	0.001	-21.632	0.000	-0.029	-0.024

Table 3: *Undermatching measured by program offer rate, regressed against individual characteristics (demographics and grades only).* $R^2 = 0.038$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.1235	0.005	22.740	0.000	0.113	0.134
male	-0.0068	0.002	-3.691	0.000	-0.010	-0.003
Black	0.0033	0.003	1.062	0.288	-0.003	0.009
Hispanic	0.0208	0.003	7.615	0.000	0.015	0.026
Multiple races/ethnicities or Other	0.0234	0.004	5.451	0.000	0.015	0.032
white	0.0061	0.003	1.895	0.058	-0.000	0.012
Portfolio cluster 1	0.0385	0.006	6.484	0.000	0.027	0.050
Portfolio cluster 2	0.0796	0.009	8.872	0.000	0.062	0.097
Portfolio cluster 3	0.0082	0.004	1.982	0.047	9.22e-05	0.016
Portfolio cluster 4	0.0737	0.009	7.766	0.000	0.055	0.092
Portfolio cluster 5	0.0910	0.011	8.569	0.000	0.070	0.112
Portfolio cluster 6	0.0598	0.007	8.218	0.000	0.046	0.074
Portfolio cluster 7	0.0191	0.005	3.881	0.000	0.009	0.029
Portfolio cluster 8	0.0023	0.005	0.488	0.626	-0.007	0.012
no reach	0.2236	0.004	62.984	0.000	0.217	0.231
no match (match-dependent)	0.2613	0.003	91.653	0.000	0.256	0.267
no safety (match-dependent)	0.3285	0.004	87.809	0.000	0.321	0.336
FRL status	-0.0025	0.002	-1.138	0.255	-0.007	0.002
avg MS grades	-0.0015	0.001	-1.494	0.135	-0.004	0.000
list length	-0.0298	0.003	-10.638	0.000	-0.035	-0.024
effective list length	-0.0245	0.002	-11.942	0.000	-0.029	-0.021
portfolio min offer rate	-0.0394	0.003	-14.810	0.000	-0.045	-0.034
portfolio avg offer rate	0.0676	0.003	23.163	0.000	0.062	0.073
inversions (binary, match-dependent)	0.0882	0.001	92.208	0.000	0.086	0.090
mean transit time	0.0445	0.001	45.044	0.000	0.043	0.046

Table 4: *OLS Regressions for undermatching measured by program offer rate, regressed against individual and (match-dependent) portfolio characteristics.* $R^2 = 0.409$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.2177	0.007	30.319	0.000	0.204	0.232
male	-0.0020	0.002	-0.924	0.355	-0.006	0.002
Black	-0.0059	0.004	-1.595	0.111	-0.013	0.001
Hispanic	0.0096	0.003	2.997	0.003	0.003	0.016
Multiple races/ethnicities or Other	0.0279	0.005	5.554	0.000	0.018	0.038
white	-0.0012	0.004	-0.321	0.749	-0.009	0.006
Portfolio cluster 1	0.0336	0.007	4.505	0.000	0.019	0.048
Portfolio cluster 2	0.0685	0.011	5.999	0.000	0.046	0.091
Portfolio cluster 3	0.0124	0.005	2.499	0.012	0.003	0.022
Portfolio cluster 4	-0.0038	0.012	-0.311	0.756	-0.028	0.020
Portfolio cluster 5	0.1151	0.013	8.865	0.000	0.090	0.140
Portfolio cluster 6	0.0519	0.009	5.595	0.000	0.034	0.070
Portfolio cluster 7	0.0160	0.006	2.639	0.008	0.004	0.028
Portfolio cluster 8	-0.0464	0.006	-8.191	0.000	-0.057	-0.035
no reach	0.1369	0.004	35.362	0.000	0.129	0.144
no match (match-independent)	0.0440	0.003	14.184	0.000	0.038	0.050
no safety (match-independent)	-0.0334	0.004	-8.453	0.000	-0.041	-0.026
FRL status	-0.0021	0.003	-0.795	0.426	-0.007	0.003
avg MS grades	-0.0068	0.001	-5.683	0.000	-0.009	-0.004
list length	-0.0023	0.003	-0.689	0.491	-0.009	0.004
effective list length	-0.0687	0.002	-27.554	0.000	-0.074	-0.064
portfolio min offer rate	-0.0102	0.003	-2.964	0.003	-0.017	-0.003
portfolio avg offer rate	0.0231	0.004	5.653	0.000	0.015	0.031
inversions (max diff, match-independent)	0.0303	0.002	14.704	0.000	0.026	0.034
mean transit time	0.0447	0.001	38.739	0.000	0.042	0.047

Table 5: *OLS Regressions for undermatching measured by program offer rate, regressed against individual and match-independent portfolio characteristics.* $R^2 = 0.192$.

A.5.2 Performance regressions

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.6695	0.002	430.654	0.000	0.666	0.673
male	-0.0047	0.001	-4.737	0.000	-0.007	-0.003
Black	-0.0488	0.002	-30.495	0.000	-0.052	-0.046
Hispanic	-0.0485	0.001	-34.207	0.000	-0.051	-0.046
Multiple races/ethnicities or Other	-0.0181	0.002	-7.741	0.000	-0.023	-0.014
white	-0.0069	0.002	-3.900	0.000	-0.010	-0.003
FRL status	-0.0206	0.001	-17.369	0.000	-0.023	-0.018
avg MS grades	0.0313	0.001	59.341	0.000	0.030	0.032

Table 6: *OLS Regressions for match performance regressed against individual characteristics.* $R^2 = 0.139$

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0801	0.001	-55.891	0.000	-0.083	-0.077
male	0.0012	0.001	1.263	0.206	-0.001	0.003
Black	-0.0334	0.001	-22.650	0.000	-0.036	-0.031
Hispanic	-0.0253	0.001	-19.336	0.000	-0.028	-0.023
Multiple races/ethnicities or Other	-0.0128	0.002	-5.903	0.000	-0.017	-0.009
white	0.0021	0.002	1.271	0.204	-0.001	0.005
FRL status	-0.0147	0.001	-13.438	0.000	-0.017	-0.013
avg MS grades	0.0077	0.000	15.919	0.000	0.007	0.009

Table 7: *OLS Regressions for undermatching measured by program performance, regressed against individual characteristics (demographics and grades only).* $R^2 = 0.036$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0689	0.002	-28.527	0.000	-0.074	-0.064
male	0.0011	0.001	1.302	0.193	-0.001	0.003
Black	-0.0003	0.001	-0.205	0.837	-0.003	0.002
Hispanic	-0.0066	0.001	-5.401	0.000	-0.009	-0.004
Multiple races/ethnicities or Other	0.0012	0.002	0.605	0.545	-0.003	0.005
white	0.0011	0.001	0.763	0.445	-0.002	0.004
Portfolio cluster 1	-0.0081	0.003	-3.077	0.002	-0.013	-0.003
Portfolio cluster 2	-0.0075	0.004	-1.871	0.061	-0.015	0.000
Portfolio cluster 3	-0.0002	0.002	-0.081	0.935	-0.004	0.003
Portfolio cluster 4	-0.0036	0.004	-0.853	0.394	-0.012	0.005
Portfolio cluster 5	-0.0023	0.005	-0.486	0.627	-0.012	0.007
Portfolio cluster 6	-0.0108	0.003	-3.348	0.001	-0.017	-0.004
Portfolio cluster 7	-0.0054	0.002	-2.458	0.014	-0.010	-0.001
Portfolio cluster 8	0.0046	0.002	2.185	0.029	0.000	0.009
no reach	-0.0701	0.002	-44.648	0.000	-0.073	-0.067
no match (match-dependent)	-0.0817	0.001	-64.658	0.000	-0.084	-0.079
no safety (match-dependent)	-0.0768	0.002	-43.092	0.000	-0.080	-0.073
FRL status	-0.0060	0.001	-6.104	0.000	-0.008	-0.004
avg MS grades	-0.0010	0.000	-2.239	0.025	-0.002	-0.000
list length	0.0060	0.001	4.838	0.000	0.004	0.008
effective list length	-0.0009	0.001	-0.936	0.349	-0.003	0.001
portfolio min offer rate	0.0084	0.001	7.122	0.000	0.006	0.011
portfolio avg offer rate	-0.0239	0.001	-18.483	0.000	-0.026	-0.021
inversions (binary, match-dependent)	-0.0198	0.000	-46.890	0.000	-0.021	-0.019
mean transit time	-0.0173	0.000	-39.315	0.000	-0.018	-0.016

Table 8: *OLS Regressions for undermatching measured by program performance, regressed against individual and (match-dependent) portfolio characteristics.* $R^2 = 0.240$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0989	0.003	-34.048	0.000	-0.105	-0.093
male	8.697e-05	0.001	0.100	0.920	-0.002	0.002
Black	0.0015	0.001	0.991	0.321	-0.001	0.004
Hispanic	-0.0039	0.001	-2.979	0.003	-0.006	-0.001
Multiple races/ethnicities or Other	0.0002	0.002	0.100	0.921	-0.004	0.004
white	0.0027	0.002	1.768	0.077	-0.000	0.006
Portfolio cluster 1	-0.0060	0.003	-2.005	0.045	-0.012	-0.000
Portfolio cluster 2	-0.0054	0.005	-1.178	0.239	-0.014	0.004
Portfolio cluster 3	-0.0004	0.002	-0.215	0.830	-0.004	0.003
Portfolio cluster 4	0.0031	0.005	0.631	0.528	-0.007	0.013
Portfolio cluster 5	-0.0119	0.005	-2.259	0.024	-0.022	-0.002
Portfolio cluster 6	-0.0082	0.004	-2.175	0.030	-0.016	-0.001
Portfolio cluster 7	-0.0036	0.002	-1.458	0.145	-0.008	0.001
Portfolio cluster 8	0.0139	0.002	6.065	0.000	0.009	0.018
no reach	-0.0407	0.002	-26.180	0.000	-0.044	-0.038
no match (match-independent)	-0.0039	0.001	-3.061	0.002	-0.006	-0.001
no safety (match-independent)	0.0177	0.002	11.101	0.000	0.015	0.021
FRL status	-0.0066	0.001	-6.330	0.000	-0.009	-0.005
avg MS grades	0.0008	0.000	1.676	0.094	-0.000	0.002
list length	0.0003	0.001	0.200	0.842	-0.002	0.003
effective list length	0.0124	0.001	12.298	0.000	0.010	0.014
portfolio min offer rate	0.0032	0.001	2.332	0.020	0.001	0.006
portfolio avg offer rate	-0.0120	0.002	-7.298	0.000	-0.015	-0.009
inversions (max diff, match-independent)	-0.0074	0.001	-8.929	0.000	-0.009	-0.006
mean transit time	-0.0174	0.000	-37.191	0.000	-0.018	-0.016

Table 9: *OLS Regressions for undermatching measured by program performance, regressed against individual and match-independent portfolio characteristics.* $R^2 = 0.144$.

A.5.3 Impact regressions

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.5515	0.001	448.302	0.000	0.549	0.554
male	0.0034	0.001	4.310	0.000	0.002	0.005
Black	-0.0244	0.001	-19.228	0.000	-0.027	-0.022
Hispanic	-0.0158	0.001	-14.037	0.000	-0.018	-0.014
Multiple races/ethnicities or Other	-0.0092	0.002	-4.985	0.000	-0.013	-0.006
white	-0.0077	0.001	-5.549	0.000	-0.010	-0.005
FRL status	-0.0111	0.001	-11.854	0.000	-0.013	-0.009
avg MS grades	0.0193	0.000	46.153	0.000	0.018	0.020

Table 10: *OLS Regressions for match impact regressed against individual characteristics.* $R^2 = 0.067$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0872	0.001	-65.342	0.000	-0.090	-0.085
male	0.0077	0.001	8.992	0.000	0.006	0.009
Black	-0.0485	0.001	-35.350	0.000	-0.051	-0.046
Hispanic	-0.0322	0.001	-26.489	0.000	-0.035	-0.030
Multiple races/ethnicities or Other	-0.0260	0.002	-12.909	0.000	-0.030	-0.022
white	-0.0052	0.002	-3.449	0.001	-0.008	-0.002
FRL status	-0.0096	0.001	-9.463	0.000	-0.012	-0.008
avg MS grades	0.0046	0.000	10.256	0.000	0.004	0.006

Table 11: *OLS Regressions for undermatching measured by program impact, regressed against individual characteristics (demographics and grades only).* $R^2 = 0.044$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0758	0.002	-32.639	0.000	-0.080	-0.071
male	0.0062	0.001	7.926	0.000	0.005	0.008
Black	-0.0228	0.001	-17.043	0.000	-0.025	-0.020
Hispanic	-0.0193	0.001	-16.555	0.000	-0.022	-0.017
Multiple races/ethnicities or Other	-0.0153	0.002	-8.192	0.000	-0.019	-0.012
white	-0.0081	0.001	-5.818	0.000	-0.011	-0.005
Portfolio cluster 1	-0.0107	0.003	-4.199	0.000	-0.016	-0.006
Portfolio cluster 2	-0.0107	0.004	-2.784	0.005	-0.018	-0.003
Portfolio cluster 3	-0.0021	0.002	-1.196	0.232	-0.006	0.001
Portfolio cluster 4	-0.0031	0.004	-0.769	0.442	-0.011	0.005
Portfolio cluster 5	-0.0100	0.005	-2.191	0.028	-0.019	-0.001
Portfolio cluster 6	-0.0107	0.003	-3.445	0.001	-0.017	-0.005
Portfolio cluster 7	-0.0068	0.002	-3.221	0.001	-0.011	-0.003
Portfolio cluster 8	-0.0059	0.002	-2.907	0.004	-0.010	-0.002
no reach	-0.0481	0.002	-31.855	0.000	-0.051	-0.045
no match (match-dependent)	-0.0594	0.001	-48.872	0.000	-0.062	-0.057
no safety (match-dependent)	-0.0302	0.002	-17.593	0.000	-0.034	-0.027
FRL status	-0.0042	0.001	-4.480	0.000	-0.006	-0.002
avg MS grades	9.751e-05	0.000	0.223	0.824	-0.001	0.001
list length	-0.0012	0.001	-1.002	0.316	-0.004	0.001
effective list length	0.0012	0.001	1.423	0.155	-0.000	0.003
portfolio min offer rate	0.0032	0.001	2.809	0.005	0.001	0.005
portfolio avg offer rate	-0.0120	0.001	-9.645	0.000	-0.014	-0.010
inversions (binary, match-dependent)	-0.0166	0.000	-40.799	0.000	-0.017	-0.016
mean transit time	-0.0208	0.000	-49.191	0.000	-0.022	-0.020

Table 12: *OLS Regressions for undermatching measured by program impact, regressed against individual and (match-dependent) portfolio characteristics.* $R^2 = 0.193$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0987	0.003	-36.309	0.000	-0.104	-0.093
male	0.0057	0.001	6.968	0.000	0.004	0.007
Black	-0.0210	0.001	-15.170	0.000	-0.024	-0.018
Hispanic	-0.0170	0.001	-14.011	0.000	-0.019	-0.015
Multiple races/ethnicities or Other	-0.0147	0.002	-7.632	0.000	-0.018	-0.011
white	-0.0071	0.001	-4.894	0.000	-0.010	-0.004
Portfolio cluster 1	-0.0096	0.003	-3.397	0.001	-0.015	-0.004
Portfolio cluster 2	-0.0107	0.004	-2.465	0.014	-0.019	-0.002
Portfolio cluster 3	-0.0022	0.002	-1.179	0.238	-0.006	0.001
Portfolio cluster 4	-0.0005	0.005	-0.105	0.917	-0.009	0.009
Portfolio cluster 5	-0.0189	0.005	-3.843	0.000	-0.029	-0.009
Portfolio cluster 6	-0.0094	0.004	-2.685	0.007	-0.016	-0.003
Portfolio cluster 7	-0.0057	0.002	-2.499	0.012	-0.010	-0.001
Portfolio cluster 8	-0.0008	0.002	-0.377	0.706	-0.005	0.003
no reach	-0.0271	0.001	-18.607	0.000	-0.030	-0.024
no match (match-independent)	0.0006	0.001	0.547	0.584	-0.002	0.003
no safety (match-independent)	0.0189	0.001	12.666	0.000	0.016	0.022
FRL status	-0.0046	0.001	-4.657	0.000	-0.006	-0.003
avg MS grades	0.0010	0.000	2.210	0.027	0.000	0.002
list length	-0.0051	0.001	-4.098	0.000	-0.007	-0.003
effective list length	0.0108	0.001	11.459	0.000	0.009	0.013
portfolio min offer rate	-0.0002	0.001	-0.125	0.901	-0.003	0.002
portfolio avg offer rate	-0.0042	0.002	-2.705	0.007	-0.007	-0.001
inversions (max diff, match-independent)	-0.0053	0.001	-6.800	0.000	-0.007	-0.004
mean transit time	-0.0206	0.000	-47.233	0.000	-0.022	-0.020

Table 13: *OLS Regressions for undermatching measured by program impact, regressed against individual and match-independent portfolio characteristics.* $R^2 = 0.139$.

A.5.4 Graduation rate regressions

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.9294	0.001	1019.083	0.000	0.928	0.931
male	-0.0040	0.001	-6.945	0.000	-0.005	-0.003
Black	-0.0189	0.001	-20.214	0.000	-0.021	-0.017
Hispanic	-0.0204	0.001	-24.550	0.000	-0.022	-0.019
Multiple races/ethnicities or Other	-0.0074	0.001	-5.385	0.000	-0.010	-0.005
white	-0.0050	0.001	-4.790	0.000	-0.007	-0.003
FRL status	-0.0098	0.001	-14.096	0.000	-0.011	-0.008
avg MS grades	0.0135	0.000	43.786	0.000	0.013	0.014

Table 14: *OLS Regression Results for match graduate rate regressed against individual characteristics.* $R^2 = 0.080$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0474	0.001	-52.735	0.000	-0.049	-0.046
male	-0.0022	0.001	-3.916	0.000	-0.003	-0.001
Black	-0.0271	0.001	-29.412	0.000	-0.029	-0.025
Hispanic	-0.0222	0.001	-27.115	0.000	-0.024	-0.021
Multiple races/ethnicities or Other	-0.0131	0.001	-9.681	0.000	-0.016	-0.010
white	-0.0036	0.001	-3.533	0.000	-0.006	-0.002
FRL status	-0.0081	0.001	-11.788	0.000	-0.009	-0.007
avg MS grades	0.0102	0.000	33.612	0.000	0.010	0.011

Table 15: *OLS Regressions for undermatching measured by program graduation rate, regressed against individual characteristics (demographics and grades only).* $R^2 = 0.071$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0448	0.002	-28.530	0.000	-0.048	-0.042
male	-0.0018	0.001	-3.359	0.001	-0.003	-0.001
Black	-0.0094	0.001	-10.357	0.000	-0.011	-0.008
Hispanic	-0.0117	0.001	-14.808	0.000	-0.013	-0.010
Multiple races/ethnicities or Other	-0.0055	0.001	-4.392	0.000	-0.008	-0.003
white	-0.0033	0.001	-3.505	0.000	-0.005	-0.001
Portfolio cluster 1	-0.0004	0.002	-0.223	0.823	-0.004	0.003
Portfolio cluster 2	0.0061	0.003	2.350	0.019	0.001	0.011
Portfolio cluster 3	-0.0024	0.001	-1.969	0.049	-0.005	-1.13e-05
Portfolio cluster 4	-5.351e-05	0.003	-0.020	0.984	-0.005	0.005
Portfolio cluster 5	0.0093	0.003	3.020	0.003	0.003	0.015
Portfolio cluster 6	0.0021	0.002	1.010	0.313	-0.002	0.006
Portfolio cluster 7	-0.0013	0.001	-0.925	0.355	-0.004	0.001
Portfolio cluster 8	0.0017	0.001	1.248	0.212	-0.001	0.004
no reach	-0.0421	0.001	-41.283	0.000	-0.044	-0.040
no match (match-dependent)	-0.0434	0.001	-52.761	0.000	-0.045	-0.042
no safety (match-dependent)	-0.0510	0.001	-44.437	0.000	-0.053	-0.049
FRL status	-0.0034	0.001	-5.317	0.000	-0.005	-0.002
avg MS grades	0.0044	0.000	15.074	0.000	0.004	0.005
list length	-0.0007	0.001	-0.863	0.388	-0.002	0.001
effective list length	-0.0036	0.001	-6.182	0.000	-0.005	-0.002
portfolio min offer rate	0.0004	0.001	0.564	0.573	-0.001	0.002
portfolio avg offer rate	-0.0123	0.001	-14.669	0.000	-0.014	-0.011
inversions (binary, match-dependent)	-0.0063	0.000	-22.980	0.000	-0.007	-0.006
mean transit time	-0.0063	0.000	-22.023	0.000	-0.007	-0.006

Table 16: *OLS Regressions for undermatching measured by program graduation rate, regressed against individual and (match-dependent) portfolio characteristics.* $R^2 = 0.212$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.0598	0.002	-32.322	0.000	-0.063	-0.056
male	-0.0023	0.001	-4.248	0.000	-0.003	-0.001
Black	-0.0089	0.001	-9.389	0.000	-0.011	-0.007
Hispanic	-0.0106	0.001	-12.835	0.000	-0.012	-0.009
Multiple races/ethnicities or Other	-0.0068	0.001	-5.229	0.000	-0.009	-0.004
white	-0.0025	0.001	-2.575	0.010	-0.004	-0.001
Portfolio cluster 1	0.0006	0.002	0.317	0.751	-0.003	0.004
Portfolio cluster 2	0.0069	0.003	2.343	0.019	0.001	0.013
Portfolio cluster 3	-0.0025	0.001	-1.954	0.051	-0.005	7.23e-06
Portfolio cluster 4	0.0008	0.003	0.241	0.809	-0.005	0.007
Portfolio cluster 5	0.0039	0.003	1.174	0.240	-0.003	0.010
Portfolio cluster 6	0.0033	0.002	1.363	0.173	-0.001	0.008
Portfolio cluster 7	-0.0004	0.002	-0.240	0.811	-0.003	0.003
Portfolio cluster 8	0.0066	0.001	4.536	0.000	0.004	0.009
no reach	-0.0258	0.001	-26.014	0.000	-0.028	-0.024
no match (match-independent)	-0.0022	0.001	-2.692	0.007	-0.004	-0.001
no safety (match-independent)	0.0068	0.001	6.661	0.000	0.005	0.009
FRL status	-0.0036	0.001	-5.414	0.000	-0.005	-0.002
avg MS grades	0.0058	0.000	18.969	0.000	0.005	0.006
list length	-0.0032	0.001	-3.853	0.000	-0.005	-0.002
effective list length	0.0036	0.001	5.623	0.000	0.002	0.005
portfolio min offer rate	-0.0012	0.001	-1.358	0.174	-0.003	0.001
portfolio avg offer rate	-0.0062	0.001	-5.916	0.000	-0.008	-0.004
inversions (max diff, match-independent)	-0.0027	0.001	-5.057	0.000	-0.004	-0.002
mean transit time	-0.0066	0.000	-22.065	0.000	-0.007	-0.006

Table 17: *OLS Regressions for undermatching measured by program graduation rate, regressed against individual and match-independent portfolio characteristics.* $R^2 = 0.192$.

A.5.5 College enrollment rate regressions

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.7463	0.002	401.068	0.000	0.743	0.750
male	-0.0123	0.001	-10.348	0.000	-0.015	-0.010
Black	-0.0791	0.002	-41.358	0.000	-0.083	-0.075
Hispanic	-0.0752	0.002	-44.384	0.000	-0.078	-0.072
Multiple races/ethnicities or Other	-0.0354	0.003	-12.640	0.000	-0.041	-0.030
white	-0.0212	0.002	-10.050	0.000	-0.025	-0.017
FRL status	-0.0210	0.001	-14.791	0.000	-0.024	-0.018
avg MS grades	0.0383	0.001	60.631	0.000	0.037	0.039

Table 18: *OLS Regressions for match college rate regressed against individual characteristics.* $R^2 = 0.164$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.1095	0.002	-51.110	0.000	-0.114	-0.105
male	-0.0041	0.001	-2.991	0.003	-0.007	-0.001
Black	-0.1039	0.002	-47.246	0.000	-0.108	-0.100
Hispanic	-0.0948	0.002	-48.642	0.000	-0.099	-0.091
Multiple races/ethnicities or Other	-0.0587	0.003	-18.225	0.000	-0.065	-0.052
white	-0.0204	0.002	-8.392	0.000	-0.025	-0.016
FRL status	-0.0181	0.002	-11.062	0.000	-0.021	-0.015
avg MS grades	0.0209	0.001	28.742	0.000	0.019	0.022

Table 19: *OLS Regressions for undermatching measured by program college rate, regressed against individual characteristics (demographics and grades only).* $R^2 = 0.108$

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.1176	0.004	-32.248	0.000	-0.125	-0.110
male	-0.0051	0.001	-4.107	0.000	-0.007	-0.003
Black	-0.0492	0.002	-23.450	0.000	-0.053	-0.045
Hispanic	-0.0640	0.002	-34.971	0.000	-0.068	-0.060
Multiple races/ethnicities or Other	-0.0365	0.003	-12.508	0.000	-0.042	-0.031
white	-0.0214	0.002	-9.749	0.000	-0.026	-0.017
Portfolio cluster 1	0.0100	0.004	2.507	0.012	0.002	0.018
Portfolio cluster 2	0.0171	0.006	2.843	0.004	0.005	0.029
Portfolio cluster 3	0.0012	0.003	0.415	0.678	-0.004	0.007
Portfolio cluster 4	0.0096	0.006	1.518	0.129	-0.003	0.022
Portfolio cluster 5	0.0175	0.007	2.450	0.014	0.003	0.031
Portfolio cluster 6	0.0120	0.005	2.460	0.014	0.002	0.022
Portfolio cluster 7	0.0040	0.003	1.216	0.224	-0.002	0.010
Portfolio cluster 8	8.87e-05	0.003	0.028	0.978	-0.006	0.006
no reach	-0.0895	0.002	-37.790	0.000	-0.094	-0.085
no match (match-dependent)	-0.0973	0.002	-50.853	0.000	-0.101	-0.094
no safety (match-dependent)	-0.0796	0.003	-29.830	0.000	-0.085	-0.074
FRL status	-0.0062	0.001	-4.143	0.000	-0.009	-0.003
avg MS grades	0.0080	0.001	11.652	0.000	0.007	0.009
list length	-0.0127	0.002	-6.789	0.000	-0.016	-0.009
effective list length	-0.0033	0.001	-2.424	0.015	-0.006	-0.001
portfolio min offer rate	-0.0037	0.002	-2.083	0.037	-0.007	-0.000
portfolio avg offer rate	-0.0295	0.002	-15.052	0.000	-0.033	-0.026
inversions (binary, match-dependent)	-0.0240	0.001	-37.599	0.000	-0.025	-0.023
mean transit time	-0.0316	0.001	-47.566	0.000	-0.033	-0.030

Table 20: *OLS Regressions for undermatching measured by program college rate, regressed against individual and (match-dependent) portfolio characteristics.* $R^2 = 0.281$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.1491	0.004	-34.818	0.000	-0.158	-0.141
male	-0.0061	0.001	-4.755	0.000	-0.009	-0.004
Black	-0.0466	0.002	-21.367	0.000	-0.051	-0.042
Hispanic	-0.0603	0.002	-31.678	0.000	-0.064	-0.057
Multiple races/ethnicities or Other	-0.0368	0.003	-12.142	0.000	-0.043	-0.031
white	-0.0193	0.002	-8.429	0.000	-0.024	-0.015
Portfolio cluster 1	0.0100	0.004	2.247	0.025	0.001	0.019
Portfolio cluster 2	0.0161	0.007	2.372	0.018	0.003	0.029
Portfolio cluster 3	6.285e-05	0.003	0.021	0.983	-0.006	0.006
Portfolio cluster 4	0.0252	0.007	3.487	0.000	0.011	0.039
Portfolio cluster 5	0.0040	0.008	0.521	0.603	-0.011	0.019
Portfolio cluster 6	0.0121	0.006	2.189	0.029	0.001	0.023
Portfolio cluster 7	0.0042	0.004	1.158	0.247	-0.003	0.011
Portfolio cluster 8	0.0121	0.003	3.587	0.000	0.005	0.019
no reach	-0.0544	0.002	-23.677	0.000	-0.059	-0.050
no match (match-independent)	-0.0093	0.002	-5.014	0.000	-0.013	-0.006
no safety (match-independent)	0.0176	0.002	7.482	0.000	0.013	0.022
FRL status	-0.0067	0.002	-4.349	0.000	-0.010	-0.004
avg MS grades	0.0101	0.001	14.216	0.000	0.009	0.011
list length	-0.0185	0.002	-9.494	0.000	-0.022	-0.015
effective list length	0.0113	0.001	7.651	0.000	0.008	0.014
portfolio min offer rate	-0.0114	0.002	-5.573	0.000	-0.015	-0.007
portfolio avg offer rate	-0.0177	0.002	-7.234	0.000	-0.022	-0.013
inversions (max diff, match-independent)	-0.0116	0.001	-9.456	0.000	-0.014	-0.009
mean transit time	-0.0315	0.001	-45.667	0.000	-0.033	-0.030

Table 21: *OLS Regressions for undermatching measured by program college rate, regressed against individual and match-independent portfolio characteristics.* $R^2 = 0.225$.

A.5.6 Aggregate quality regressions

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.3114	0.009	33.355	0.000	0.293	0.330
male	0.0156	0.006	2.616	0.009	0.004	0.027
Black	-0.1616	0.010	-16.800	0.000	-0.181	-0.143
Hispanic	-0.0739	0.009	-8.674	0.000	-0.091	-0.057
Multiple races/ethnicities or Other	-0.0409	0.014	-2.909	0.004	-0.069	-0.013
white	0.0070	0.011	0.660	0.509	-0.014	0.028
FRL status	-0.0805	0.007	-11.296	0.000	-0.094	-0.067
avg MS grades	0.1251	0.003	39.379	0.000	0.119	0.131

Table 22: *OLS Regressions for match aggregate quality regressed against individual characteristics.* $R^2 = 0.054$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.8148	0.010	-79.158	0.000	-0.835	-0.795
male	0.0491	0.007	7.463	0.000	0.036	0.062
Black	-0.3037	0.011	-28.654	0.000	-0.324	-0.283
Hispanic	-0.1176	0.009	-12.537	0.000	-0.136	-0.099
Multiple races/ethnicities or Other	-0.1388	0.016	-8.943	0.000	-0.169	-0.108
white	0.0583	0.012	5.007	0.000	0.035	0.081
FRL status	-0.0730	0.008	-9.290	0.000	-0.088	-0.058
avg MS grades	0.0580	0.003	16.589	0.000	0.051	0.065

Table 23: *OLS Regressions for undermatching measured by program aggregate quality, regressed against individual characteristics (demographics and grades only).* $R^2 = 0.041$

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.7279	0.018	-39.874	0.000	-0.764	-0.692
male	0.0395	0.006	6.392	0.000	0.027	0.052
Black	-0.1420	0.011	-13.478	0.000	-0.163	-0.121
Hispanic	-0.0455	0.009	-4.953	0.000	-0.063	-0.027
Multiple races/ethnicities or Other	-0.0700	0.015	-4.781	0.000	-0.099	-0.041
white	0.0309	0.011	2.815	0.005	0.009	0.052
Portfolio cluster 1	-0.0491	0.020	-2.461	0.014	-0.088	-0.010
Portfolio cluster 2	-0.0469	0.030	-1.556	0.120	-0.106	0.012
Portfolio cluster 3	-0.0071	0.014	-0.509	0.611	-0.034	0.020
Portfolio cluster 4	-0.1962	0.032	-6.170	0.000	-0.259	-0.134
Portfolio cluster 5	-0.0319	0.036	-0.894	0.371	-0.102	0.038
Portfolio cluster 6	-0.0362	0.024	-1.484	0.138	-0.084	0.012
Portfolio cluster 7	-0.0153	0.016	-0.928	0.354	-0.048	0.017
Portfolio cluster 8	-0.0759	0.016	-4.794	0.000	-0.107	-0.045
no reach	-0.3329	0.012	-28.031	0.000	-0.356	-0.310
no match (match-dependent))[T.True]	-0.4326	0.010	-45.281	0.000	-0.451	-0.414
no safety (match-dependent))[T.True]	-0.2216	0.013	-16.467	0.000	-0.248	-0.195
FRL status	-0.0407	0.007	-5.470	0.000	-0.055	-0.026
avg MS grades	0.0320	0.003	9.299	0.000	0.025	0.039
list length	-0.0134	0.009	-1.434	0.151	-0.032	0.005
effective list length	0.0006	0.007	0.088	0.930	-0.013	0.014
portfolio min offer rate	0.0708	0.009	7.950	0.000	0.053	0.088
portfolio avg offer rate	-0.0732	0.010	-7.479	0.000	-0.092	-0.054
inversions	-0.1131	0.003	-35.364	0.000	-0.119	-0.107
mean transit time	-0.1392	0.003	-41.905	0.000	-0.146	-0.133

Table 24: *OLS Regressions for undermatching measured by program aggregate quality, regressed against individual and (match-dependent) portfolio characteristics.* $R^2 = 0.162$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	-0.9050	0.021	-42.656	0.000	-0.947	-0.863
male	0.0360	0.006	5.675	0.000	0.024	0.048
Black	-0.1283	0.011	-11.851	0.000	-0.149	-0.107
Hispanic	-0.0270	0.009	-2.864	0.004	-0.046	-0.009
Multiple races/ethnicities or Other	-0.0643	0.015	-4.281	0.000	-0.094	-0.035
white	0.0387	0.011	3.434	0.001	0.017	0.061
Portfolio cluster 1	-0.0337	0.022	-1.529	0.126	-0.077	0.009
Portfolio cluster 2	-0.0310	0.034	-0.918	0.359	-0.097	0.035
Portfolio cluster 3	-0.0057	0.015	-0.387	0.699	-0.034	0.023
Portfolio cluster 4	-0.1484	0.036	-4.136	0.000	-0.219	-0.078
Portfolio cluster 5	-0.0860	0.038	-2.244	0.025	-0.161	-0.011
Portfolio cluster 6	-0.0148	0.027	-0.542	0.588	-0.069	0.039
Portfolio cluster 7	-0.0040	0.018	-0.226	0.821	-0.039	0.031
Portfolio cluster 8	-0.0374	0.017	-2.237	0.025	-0.070	-0.005
no reach	-0.1754	0.011	-15.425	0.000	-0.198	-0.153
no match (match-independent)	-0.0175	0.009	-1.899	0.058	-0.036	0.001
no safety (match-independent)	0.1546	0.012	13.309	0.000	0.132	0.177
FRL status	-0.0426	0.008	-5.586	0.000	-0.058	-0.028
avg MS grades	0.0391	0.004	11.129	0.000	0.032	0.046
list length	-0.0398	0.010	-4.135	0.000	-0.059	-0.021
effective list length	0.0642	0.007	8.733	0.000	0.050	0.079
portfolio min offer rate	0.0349	0.010	3.464	0.001	0.015	0.055
portfolio avg offer rate	-0.0048	0.012	-0.398	0.691	-0.028	0.019
inversions (max diff, match-independent)	-0.0454	0.006	-7.485	0.000	-0.057	-0.033
mean transit time	-0.1379	0.003	-40.423	0.000	-0.145	-0.131

Table 25: *OLS Regressions for undermatching measured by program aggregate quality, regressed against individual and match-independent portfolio characteristics.* $R^2 = 0.117$.

A.6 Consistency of program offer rates across years

Figure 24 shows, for each program its offer rate for the 2021 admissions cycle versus that in the 2022 cycle. Program offer rates are largely consistent across years. In particular, 85% of programs with offer rate 100% in 2021 also have offer rate 100% in 2022. This broad consistency helps explain the effectiveness of recommendations using prior years' offer rate information available at the time of the feedback.

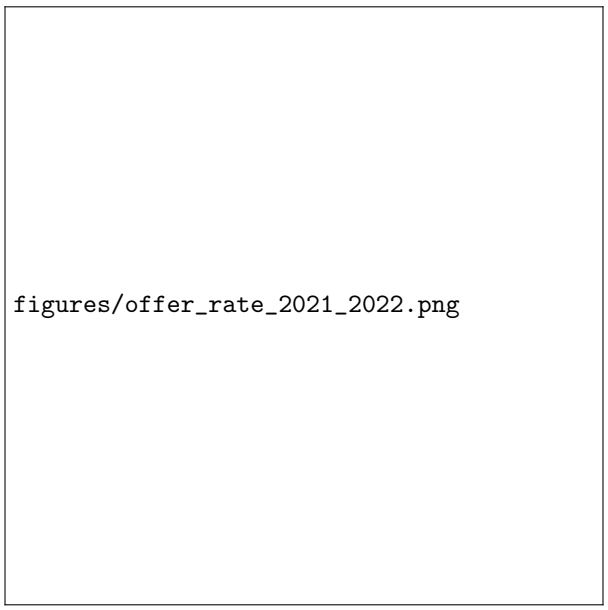


Figure 24: *Comparison of program offer rates between 2021 and 2022. Note that the correlation is overall high, particularly among highly selective (low offer rate) and nearly unselective (offer rate close to 100%) programs.*

A.7 Program quality metrics by position in list

Figure 25 shows, for each position k in applicant rank lists, the average program quality (by various metrics) for a program ranked at that position. We disambiguate this average by student characteristics as well as by the overall number of programs that a student ranked. Notably, on average, there is a reach-match-safety pattern; on average, the last program listed by a student is substantially less selective than the other programs listed.

A.8 The effect of lottery numbers

Here, we consider the effect of lottery numbers on four portfolio statistics: number of choices, offer rate of the first choice, mean offer rate, and maximum offer rate. In each regression, we include interactions with race and ethnicity.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	8.4978	0.069	123.235	0.000	8.363	8.633
Black	0.3920	0.095	4.146	0.000	0.207	0.577
Hispanic	0.0139	0.085	0.164	0.869	-0.152	0.180
Multiple races/ethnicities or Other	0.0737	0.136	0.544	0.587	-0.192	0.339
unknown	-1.2066	2.817	-0.428	0.668	-6.728	4.315
white	-0.9580	0.106	-9.043	0.000	-1.166	-0.750
tiebreaker num	0.2104	0.121	1.741	0.082	-0.027	0.447
tiebreaker num : Black	-0.0542	0.165	-0.329	0.742	-0.378	0.269
tiebreaker num : Hispanic	-0.2748	0.148	-1.859	0.063	-0.565	0.015
tiebreaker num : Multiple	-0.2949	0.236	-1.250	0.211	-0.757	0.167
tiebreaker num : unknown	3.4771	6.382	0.545	0.586	-9.032	15.987
tiebreaker num: white	-0.0483	0.185	-0.261	0.794	-0.411	0.315

Table 26: *Effect of lottery numbers on number of choices.* $R^2 = 0.012$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.2192	0.006	37.159	0.000	0.208	0.231
Black	0.2616	0.008	32.347	0.000	0.246	0.277
Hispanic	0.2201	0.007	30.422	0.000	0.206	0.234
Multiple races/ethnicities or Other	0.0588	0.012	5.074	0.000	0.036	0.082
unknown	0.2460	0.241	1.021	0.307	-0.226	0.718
white	0.0399	0.009	4.407	0.000	0.022	0.058
tiebreaker num	-0.0071	0.010	-0.691	0.490	-0.027	0.013
tiebreaker num : Black	0.0227	0.014	1.610	0.107	-0.005	0.050
tiebreaker num: Hispanic	0.0142	0.013	1.125	0.261	-0.011	0.039
tiebreaker num: Multiple	-0.0103	0.020	-0.511	0.609	-0.050	0.029
tiebreaker num: unknown	-0.4608	0.546	-0.844	0.399	-1.531	0.609
tiebreaker num: white	0.0113	0.016	0.713	0.476	-0.020	0.042

Table 27: *Effect of lottery numbers on offer rate of first choice.* $R^2 = 0.108$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.5088	0.004	114.794	0.000	0.500	0.518
Black	0.1591	0.006	26.174	0.000	0.147	0.171
Hispanic	0.1376	0.005	25.316	0.000	0.127	0.148
Multiple races/ethnicities or Other	0.0182	0.009	2.085	0.037	0.001	0.035
unknown	0.0350	0.181	0.193	0.847	-0.320	0.390
white	0.0604	0.007	8.872	0.000	0.047	0.074
tiebreaker num	-0.0143	0.008	-1.842	0.065	-0.030	0.001
tiebreaker num: Black	0.0126	0.011	1.193	0.233	-0.008	0.033
tiebreaker num: Hispanic	0.0229	0.010	2.411	0.016	0.004	0.042
tiebreaker num: Multiple	0.0289	0.015	1.905	0.057	-0.001	0.059
tiebreaker num: unknown	-0.2379	0.410	-0.580	0.562	-1.042	0.566
tiebreaker num: white	0.0110	0.012	0.925	0.355	-0.012	0.034

Table 28: *Effect of lottery numbers on mean offer rate of choices.* $R^2 = 0.070$.

	coef	std err	t	P> |t|	[0.025	0.975]
Intercept	0.9227	0.003	319.563	0.000	0.917	0.928
Black	0.0574	0.004	14.503	0.000	0.050	0.065
Hispanic	0.0503	0.004	14.213	0.000	0.043	0.057
Multiple races/ethnicities or Other	-0.0205	0.006	-3.611	0.000	-0.032	-0.009
unknown	0.0544	0.118	0.461	0.645	-0.177	0.286
white	-0.0062	0.004	-1.407	0.159	-0.015	0.002
tiebreaker num	-0.0109	0.005	-2.154	0.031	-0.021	-0.001
tiebreaker num : Black	0.0106	0.007	1.540	0.124	-0.003	0.024
tiebreaker num : Hispanic	0.0119	0.006	1.917	0.055	-0.000	0.024
tiebreaker num : Multiple	0.0279	0.010	2.822	0.005	0.009	0.047
tiebreaker num : unknown	-0.0087	0.267	-0.033	0.974	-0.533	0.515
tiebreaker num : white	0.0157	0.008	2.023	0.043	0.000	0.031

Table 29: *Effect of lottery numbers on max offer rate of choices.* $R^2 = 0.035$.



Figure 25: *Trends and disparities in program quality by position in ranked list. Left shows trends in the quality of programs, aggregated according to the total number of programs listed by a student. Middle and right breaks down trends by FRL status and ethnicity, restricted to students who listed 12 programs.*