Inferring Optimal Peer Assignment from Experimental Data

Debopam Bhattacharya *

Dartmouth College.

First draft: February 15, 2006; this draft: July 29, 2008.

Abstract

This paper studies the problem of optimally dividing individuals into peer groups in order to maximize social gains from heterogeneous peer effects. The specific setting analyzed here concerns efficient ways of allocating roommates in college dormitories. Using confidential data on a sample of Dartmouth College freshmen who were randomly assigned to be roommates, the paper derives efficient roommate pairing rules, based on demographic and academic background, which maximize different aggregate outcomes. Segregation by pre-college academic standing and by race are seen to minimize mean enrolment into sororities and maximize mean enrolment into fraternities. Segregation has no effect on mean and median freshman year GPA but increases the higher and decreases the lower percentiles of the GPA distribution for both men and women. Efficiency loss due to legal constraints on allocations (e.g., race-blindness) is shown to be significant and also larger for women for whom peer effects are more nonlinear. The paper develops large-sample inference methods for these optimal solutions and the resulting optimized values by using and extending insights from the mathematical programming literature. Applicability of these techniques extends beyond linear maximands such as the mean to other important policy objectives such as outcome quantiles which, though nonlinear, are shown to be quasi-convex in the allocation probabilities.

*Address for correspondence: 327, Rockefeller Hall, Dartmouth College, Hanover, NH 03755. e-mail: debopam@dartmouth.edu. This research was partly funded by the Junior Faculty Fellowship and the Rocky-Haney grant at Dartmouth College. Part of this research was completed when the author was a visiting research fellow at the Cowles Foundation for Economic Research at Yale University. The paper has greatly benefitted from the help and feedback of Bryan Graham and Bruce Sacerdote.

1 Introduction

A large body of research has emerged over the last two decades analyzing the phenomenon of peer effects in social environments using individual level data. This research has primarily focused on detecting peer influences on individual outcomes in specific empirical settings and occasionally examining their heterogeneity, i.e., how they vary with observable characteristics of individuals and their peers. Existence of such heterogeneity potentially implies an optimal way for a planner to form peer groups which maximizes the overall social gains from the resulting peer interactions. However, formal statistical analysis and application of such optimal allocation methods seem to have been largely ignored in the literature. The present paper examines this problem in a specific empirical setting– that of optimally assigning college freshmen to double-occupancy dorm rooms– using confidential data from two cohorts of Dartmouth College freshmen, previously analyzed in Sacerdote (2001). These freshmen were randomly paired up to be roommates and thus provide a useful dataset, purged of the confounding effects of endogenous peer selection, on which to base the analysis of optimal peer-group formation.

Methodologically, the problem of optimal peer assignment reduces to a constrained optimization problem. The objective function in this optimization problem is a functional (e.g., the mean or median) of the planner's utility distribution resulting from an assignment. Typically, the exact values of the functional corresponding to different assignments are unknown to the planner but are consistently estimable from outcome data on a sample of individuals who were randomly divided into peer groups. Because of sampling variability, the solutions and optimal values– the so called "value functions"– for the sample optimization problem would warrant a statistical distribution theory on which to base inference about the optimal allocation policy. The present paper formalizes this inferential methodology and uses it to study optimal peer assignment for the two cohorts of Dartmouth freshmen mentioned above.

The paper will focus on two types of objective functionals for the planner, viz. (i) the mean and (ii) any quantile of the planner's utility distribution. In decision theory, the maximization of expected utility is a fairly standard criterion for optimal choice under uncertainty. Manski (1988) proposed the use of quantile maximization as an alternative criterion which is consistent with several logically desirable properties of planner preferences over outcome distributions, that are not shared by mean utility maximization (elaborated below in section 5). The optimal allocation resulting out of these different criteria would be different in general.

The rest of the paper is structured as follows. Section 2 presents an overview of the dataset and summarizes the substantive and methodological contributions of the paper. Section 3 provides a brief review of the related literature. Sections 4 and 5 develop the technical apparatus used subsequently to analyze the problem. Section 4 focuses on mean maximization and section 5 considers quantile maximization. Section 6 presents the results from the empirical analysis and section 7 concludes. Proofs are collected in the appendix which also discusses a caveat related to the application.

2 Overview of Data, Methods and Results

The dataset used in this paper pertains to observed outcomes of a randomized roommate assignment process at Dartmouth College, a small liberal arts college located in the north-eastern state of New Hampshire in the US, for the graduating classes of 1997 and 1998. The dataset, compiled and previously analyzed by Sacerdote (2001), contains information on several background characteristics and a few academic and social outcomes of the students who were randomly paired up to be roommates in their freshman year. The random nature of roommate allocation in this case enabled Sacerdote (2001) to detect significant "causal" effects of roommate characteristics on individual outcomes, controlling for the individuals' own characteristics. Using the same dataset, the present paper analyzes the problem of optimally allocating freshmen based on the heterogeneity in peer effects detected in a first-stage analysis. The first stage analysis here, unlike Sacerdote (2001), is nonparametric in the sense that no (linear) structure is imposed on the equation linking individual outcomes to own and peer characteristics.

A detailed description of all variables in the dataset is provided in Sacerdote (2001). Among these variables, the two student outcomes considered in the present paper are (i) freshman year GPA as an index of academic outcome and (ii) eventual enrolment into a fraternity/sorority as a social outcome. Optimal allocation is analyzed for four different objectives– (i) maximizing the mean GPA, (ii) maximizing the lowest decile, median or highest decile of the GPA distribution, (iii) maximizing the mean incidence of fraternity joining and (iv) minimizing the mean incidence of fraternity joining. The two separate covariates that are used to design the optimal matching are (a) an index of prior academic standing, henceforth called "ACA", which is a weighted sum of the student's high school class rank and SAT scores (c.f. Sacerdote (2001) footnote 10) and (b) race which is either classified as minority (black, Hispanic, native Indian) and nonminority (all others) or as white and nonwhite. When ACA is used as the conditioning covariate, it is discretized into two, three or four categories and results are obtained separately for each classification. The analysis is done separately for men and women and restricted to individuals– 436 men and 428 women– who were assigned to double rooms.

The analysis reveals that roommate reallocation has very small effects on overall mean and median GPA but large effects on lower and upper quantiles of the GPA distribution for both male and female students. In particular, segregation by (observable) types leads to maximization of the upper and minimization of the lower quantiles of the GPA distribution. On the other hand, the effect of full segregation on fraternity joining is seen to be exactly opposite between male and female students. Segregation tends to maximize fraternity joining for men and minimize sorority joining for women. This result holds both when segregation is based on race and when it is based on prior academic standing. It is also seen that political constraints on allocations (e.g., race-blindness) can lead to significant efficiency loss, especially for women.

The inference methods employed to study the problem are based on insights from mathematical programming theory. In particular, the mean maximization problems reduce to standard linear programs (LP) with estimated objective functions. This causes (i) the effective parameter space to be discrete, (ii) the set of estimated solutions to converge to the set of true solutions arbitrarily fast and (iii) sampling error in estimating the optimal solution(s) to have no effect asymptotically on the distribution of the estimated maximum value. However, this latter distribution depends qualitatively on whether the solution to the corresponding population optimization problem is unique or not and generically, it will not be known a priori as to which of these is the case. So even if nonuniqueness is not of direct practical concern for a policymaker, i.e., she can pick any one of multiple optimal allocations, it affects the construction of confidence intervals (CI, henceforth) for the value function– an object of primary interest to her. Interestingly, it turns out that this programming theory analysis for the mean can also be applied to the problem of maximizing any quantile of the outcome distribution even though the quantile objective function is nonlinear in the allocation probabilities. The key idea behind this result is that the maximand in the quantile case is quasi-convex in the allocation probabilities, thereby reducing the effective parameter space to a discrete set, just as in the mean maximization case.

3 Related Literature

Since the present paper is substantively related to several areas of economics and statistics, this section attempts to give a brief overview of where it fits into each of these areas. Bibliographies of specific books and papers (usually the most recent ones) are cited below for more comprehensive lists of related research. The following discussion is divided by the strand of literature that the paper relates to.

Peer Effects: There is now a large literature on the analysis of peer effects in a variety of empirical settings– c.f. references in Graham (2008) or Mas and Moretti (2008) for a comprehensive bibliography. Several of these works also discuss heterogeneity in peer effects on a wide variety of outcomes. Some recent examples include Harris and Lopez-Valcarcel (2008) on smoking behavior, Fletcher (2007) on sexual initiation, Bandiera and Rasul (2006) on technology adoption and Hoxby (2002), Winston and Zimmerman (2003) and Graham (2008) on educational achievement. The latter set of papers also note that in the presence of heterogeneous peer effects, peer reallocation can potentially affect the overall mean of social outcomes. The present paper extends this body of research by formalizing the statistical analysis of optimal peer assignment in such situations.

Input Allocation: The work presented here was originally inspired by the input-matching problem of Graham, Imbens and Ridder (2005) (henceforth, GIR). GIR compared a fixed set of matching rules between two continuous inputs with fixed marginals. Calculating the optima, as outlined in the present paper, also enables one to compute the relative efficiency of a given allocation rule. Unlike GIR, however, this paper is concerned with discrete types which enables the use of mathematical programming. The discrete analysis is motivated by the facts that (i) survey data are often available only on discrete attributes, e.g., gender, race, marital status, (ii) conditioning allocation on a few discrete covariates is operationally simpler and (iii) unless the outcome varies a lot across small changes in a continuous covariate, basing allocation on a discretization of it is unlikely to cause significant suboptimality. In independent work, GIR (2007) have considered mean-based optimality for the case of two inputs with two or three levels each. In contrast, the present paper considers the maximization of both means and quantiles with general discrete valued covariates.

Treatment Choice: The paper is closely related to the treatment choice literature, c.f. Manski (2004), Dehejia (2005) and Hirano and Porter (2005)– in that it concerns designing optimal policies based on results of a random experiment. But it differs in that it analyzes a *constrained* decision-making problem where a large number of individuals have to be allocated simultaneously and not everyone can be assigned their first-best. The constraints fundamentally change the analytics of the problem. Unlike Manski (2004), however, the paper does not analyze the optimal choice of covariates and takes the set of covariates as given. In related work, Bhattacharya and Dupas (2008) explore efficient allocation of a binary treatment with continuously distributed covariates and a budget constraint limiting aggregate treatment coverage. Allowing for continuous covariates makes its analytics very different from the methods of this paper.

Set-identification: The optimization problems discussed in this paper can often have nonunique solutions– the leading case being where peer effects are linear– as discussed below. This implies a connection with the recent literature on set-valued identification in econometrics, c.f. Manski (2008) for a reasonably comprehensive bibliography. In fact, a variant of one key result from that literature is used below in proposition 1 as a step in developing the distribution theory. However, the focus of the present paper is on the distribution of the estimated maximized value when there are multiple solutions to the population maximization problem and less so on the set of maximizers, as in the set-identification literature. See section 4.2 below for more discussion on this.

Inference under parameter constraints: A sizeable literature in statistics and econometrics deals with statistical properties of solutions and values of certain constrained stochastic optimization problems (SOL, from now on). In econometrics, see Wolak (1989) and Rosen (2006). In statistics, the earliest reference appears to be Bartholomew (1957) and Aitchison and Silvey (1958). Some recent works include Shapiro (1989, 1993), King and Rockafellar (1993), Geyer (1994) and Shapiro and De Mello (2000) among others. Silvapulle and Sen (2005) provide a comprehensive textbook treatment of the literature. These works analyze statistical inference in presence of equality and inequality constraints on the parameters using Lagrangian-based techniques. The present paper differs from this literature in at least three respects. First, the objective functions here are either ratios of averages (c.f. (3) below) or more complicated functionals like quantiles (c.f. (6), (7) below) which cannot be expressed as simple population means, unlike in SOL. Second, nonuniqueness is usually not relevant for the SOL (c.f. assumptions A.3 in Shapiro (1989) or A in Geyer (1994)) because there one typically maximizes strictly concave functionals on convex parameter spaces, e.g., inequality constrained MLE's for normal means. In contrast, the problems in the present paper typically involve nonconcave objectives, so that nonuniqueness is always a theoretical possibility and this affects construction of CI for the value function. Nonuniqueness is also shown to be common in the central application of the present paper. Third, the key distributional results in the present paper are driven by discreteness of the effective parameter space. This discreteness is caused by **quasiconvex** maximands and does not occur when one is maximizing a strictly **concave** objective. An alternative way to describe the latter distinction is that Lagrangian/Kuhn-Tucker based methods are relevant in problems where optimality can occur at interior "tangency" points and where corner solutions (to the "population" problem) are not generic. The population optimization problems occurring in the present paper are of the LP type, where the *objective functions* are linear or quasiconvex and consequently optimality occurs generically at corners.

Quantile maximization: Statistical analysis of quantile maximization, although conceivably relevant to several of the broad themes cited above, does not appear to have any precedent in the corresponding literatures. Manski (1988) and Rostek (2005) have previously considered axiomatization and choice-theoretic interpretations of quantile maximization. The present paper differs from them because it analyzes the quantile maximization problem from the perspective of classical statistical inference.

Stochastic Programming: A significant body of research, usually labeled "stochastic programming", has developed efficient algorithms for solving multi-stage optimization problems in uncertain discrete environments. Textbooks like Birge and Louveaux (1997) and Bertsekas (2005) provide good overviews. In contrast, the standard double description algorithm, implemented via the "cdd" routine (coded by Komei Fukuda and described below in section 6) is sufficient for computation in the present paper and the focus instead is on formalizing statistical inference in "effectively" discrete environments when the data used in the optimization arise from sampling.

4 Mean Maximization

This section presents the formal description of the mean-maximizing allocation problem and develops the theory of inference which will be used to conduct the empirical analysis. To fix ideas initially, this section will use the example of college GPA as the outcome and consider roommate allocation based on race as the policy of interest. The theory will be developed subsequently in somewhat more general terms because the application has several subcases which differ by what covariates the room-types are based on and thus also by the number of possible types. The transition from the specific example to the general case is straightforward– it simply involves a renumbering of indices– but will be clarified below nonetheless.

Suppose that a college authority wants to improve average freshman year GPA of the incoming class, using dorm allocation as a policy instrument. The underlying behavioral assumption is that sharing a room with a "better" peer can potentially improve one's own outcome, where "better" could mean a high ability student, a student who is similar to her roommate etc. Scope for improvement exists if peer effects are nonlinear– i.e., the composite effect of own background and roommate's background on own outcome are not additively separable between the two. Otherwise, all assignments should yield the same total, and thus average, outcome.

Assume that every dorm room can accommodate 2 students and the college can assign individuals to dorms based on their race, classified as white (w), black (b), other (o). Denote the expected GPA of individuals living in each of 6 types of dorm rooms by $\gamma = (\gamma_{ww}, \gamma_{bb}, \gamma_{oo}, \gamma_{wo}, \gamma_{wb}, \gamma_{bo})'$. For instance, if one b type and one o type student were randomly picked from the population and assigned to be roommates for a year, then the expected value of the mean of their eventual GPA's is γ_{bo}. Further, let the marginal distribution of race for the incoming class be denoted by $\boldsymbol{\pi} = \{\pi_j\}_{j=w,b,o}$. Then an allocation is a vector $\mathbf{p} = \{p_{ij}\}'_{i,j=w,b,o}$, satisfying $p_{ij} \geq 0$ and $\mathbf{p'1} = 1$. Here p_{ij} $(= p_{ji})$ denotes the fraction of dorm rooms that have one student of type i and one of type j, with $i, j \in \{w, b, o\}$. Then the authority's problem is defined by the following Linear Programming (LP) problem.

$$\max_{\{p_{ij}\}} [\gamma_{ww}p_{ww} + \gamma_{bb}p_{bb} + \gamma_{oo}p_{oo} + \gamma_{wo}p_{wo} + \gamma_{wb}p_{wb} + \gamma_{bo}p_{bo}]$$

subject to

$$\begin{aligned} 2p_{ww} + p_{wo} + p_{wb} &= 2\pi_w \\ 2p_{bb} + p_{wb} + p_{bo} &= 2\pi_b \\ 2p_{oo} + p_{wo} + p_{bo} &= 2\pi_o = 2(1 - \pi_w - \pi_b) \\ p_{ij} &\geq 0,\ i, j \in \{w, b, o\}. \end{aligned} \tag{1}$$

The factor 2 appears on the RHS because the number of students is twice the number of rooms. Thus one can view the $\boldsymbol{\gamma}$'s as the preliminary parameters of interest and the solution to the LP problem and the resulting maximum value as functions of $\boldsymbol{\gamma}$'s which constitute the ultimate parameters of interest.

In general, the $\boldsymbol{\gamma}$'s will be unknown and so the above problem is infeasible. Now suppose a sample was drawn from the same population (e.g., the freshmen class in the previous year) and this pilot sample was randomly grouped into rooms. Suppose the planner has access to eventual freshman year GPA data for each member of this sample. Then the planner can estimate, say, γ_{wo} by $\hat{\gamma}_{wo}$, the mean individual GPA across the sample students who (randomly) lived in a $\{w, o\}$ type room. Replacing unknown $\boldsymbol{\gamma}$'s by the sample counterparts $\hat{\gamma}$, the planner now solves

$$\max_{\{p_{ij}\}} [\hat{\gamma}_{ww}p_{ww} + \hat{\gamma}_{bb}p_{bb} + \hat{\gamma}_{oo}p_{oo} + \hat{\gamma}_{wo}p_{wo} + \hat{\gamma}_{wb}p_{wb} + \hat{\gamma}_{bo}p_{bo}]$$

subject to the constraints in (1).

General Notation: Now, some general notation is introduced to help develop the subsequent exposition. Let M denote the number of different types of individuals and $m = \frac{M(M+1)}{2}$ be the corresponding number of types of rooms. Let the vector of mean outcomes for the m types of rooms, obtained from a random assignment of the entire population, be denoted by $\boldsymbol{\gamma} = (\gamma_j)_{j=1,..m}$, e.g., γ_1 is the expected outcome across individuals living in roomtype 1 corresponding to a random assignment of the entire population. Let the corresponding c.d.f.'s for outcomes be denoted by $\{F_j\}$, $j = 1, 2, ...m$. The constraint set is given by $\mathcal{P} = \{\mathbf{p} : \mathbf{Ap} = \pi, \mathbf{p} \geq 0\}$, where $\mathbf{A}$ is the known $M \times m$ matrix of constraint coefficients, $\mathbf{p}$ is an $m \times 1$ vector denoting a generic allocation and π is the $M \times 1$ known vector denoting the relative frequencies of the M types of individuals.

For instance, for the example in (1),

$$\begin{aligned}
M &= 3, m = 6, \\
\gamma_1 &\equiv \gamma_{ww}, \gamma_2 \equiv \gamma_{bb}, \gamma_3 \equiv \gamma_{oo}, ..., \gamma_6 = \gamma_{bo}, \\
\mathbf{A} &= \begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 & 0 & 1 \end{pmatrix}, \boldsymbol{\pi} = (\pi_w, \pi_b, \pi_o)'.
\end{aligned}$$

In this general set-up, the planner's problem, if she knew $\boldsymbol{\gamma}$, is referred to as the population problem and is given by the LP

$$\max_{\mathbf{p}} \boldsymbol{\gamma}'\mathbf{p} \text{ subject to } \mathbf{Ap} = \boldsymbol{\pi}, \, \mathbf{p} \geq \mathbf{0}. \tag{2}$$

Notice that the parameter space $\mathcal{P}$ is a compact simplex. Therefore, the maximized value of the objective in the optimization problem (2) is attained and is thus uniquely defined. But in general (2) can have multiple solutions, all of which yield the same maximum value (more on this below).

Typically, the π's will be known and $\boldsymbol{\gamma}$'s will not be known. A random sample is now drawn from the same population and one observes the outcomes resulting from random assignment of this sample to rooms. Conditional means calculated on the basis of this sample are denoted by $\hat{\boldsymbol{\gamma}} = \left(\hat{\gamma}_j\right)$, $j = 1, ..m$. Now, the planner solves the problem

$$\max_{\mathbf{p}} \hat{\boldsymbol{\gamma}}'\mathbf{p} \text{ subject to } \mathbf{Ap} = \boldsymbol{\pi},\ \mathbf{p} \geq \mathbf{0},$$

which will be called the sample problem.

4.1 Utility-based analysis

The above analysis can be extended without any substantive changes to the situation where the planner puts different utility weights on different values of the outcome. If the planner's utility from outcomes is captured by the function $u(\cdot)$, simply define $\tilde{\gamma}_j$ to be the mean of the planner's *utility realizations* based on outcomes of individuals in roomtype j. Then the the objective function for mean utility maximization is simply $\sum_j p_j \tilde{\gamma}_j$, which is still linear in the $\mathbf{p}$'s and the LP structure of the problem is preserved. A similar argument will work for quantiles based on c.d.f.'s of the utility distribution. Consequently, the rest of the paper will use the simplest utility specification, viz. $u(y) = y$, with the understanding that one can easily perform an analogous analysis for any other reasonable utility function of interest.

4.2 Parameters of Interest

Given a social objective, a key parameter of interest is the maximal allocation(s). But the resulting value function is also of independent policy interest, especially when all technologically feasible allocations are not feasible due to political or legal reasons. For example, race-based allocations are often viewed as unfair and therefore banned. Then the unconstrained maximal *solution(s)* is not of direct interest but the (unique) maximum *value* is still useful as a "first-best" benchmark. In particular, the difference between the maximum value for a legally unconstrained problem and that of a constrained one gives an estimate of the efficiency costs of the constraint. In the allocation problem (1) for instance, if race-based allocation is not permitted and the planner allocates students randomly (since no other covariate is observed in this example), then the corresponding expected value is given by $v_r = \boldsymbol{\gamma}'\mathbf{p}_r$ where $\mathbf{p}_r = (\pi_w^2, \pi_b^2, \pi_o^2, 2\pi_w\pi_o, 2\pi_w\pi_b, 2\pi_o\pi_b)$. The efficiency cost of non-discrimination is therefore $(v - v_r)$ where $v = \boldsymbol{\gamma}'\mathbf{p}^*(\boldsymbol{\gamma}, \boldsymbol{\pi})$ and $\mathbf{p}^*(\boldsymbol{\gamma}, \boldsymbol{\pi})$ solves the LP (1). Such considerations motivate focusing the present paper on inferring the value function. Also, methodologically, inference on the maximized value in a set-identified situation appears to be new in contrast to inference on the set of solutions, on which there is now a well-developed literature in econometrics.

4.3 Plug-in Estimation

Note that both the optimal allocation(s) $\mathbf{p}^*$ and the corresponding maximum value v^* in the population problem (1) can be viewed as functionals of the underlying parameters $F = \{F_j(\cdot)\}_{j=1,...m}$ defined above, i.e., $v^* = v(F)$. The approach followed in this paper is to use consistent estimates $\hat{F}$ of F to estimate v by $\hat{v} = v\left(\hat{F}\right)$ and show that such a $\hat{v}$ is consistent and has an asymptotic distribution which can be simulated. The latter derivation can be viewed as a "delta method" for the value function in constrained optimization problems with potentially multiple solutions.

From a decision theoretic standpoint, this plug-in rule asymptotically minimizes maximum regret among all feasible allocations. To see this, let $\mathbf{z}_n$ be the data and $\boldsymbol{\gamma} \in \Gamma$ be a "state of nature" where Γ is a bounded subset of the Euclidean space. Then in the mean maximization case, the maximum regret of a feasible decision rule $\mathbf{p}(\cdot)$, mapping $\mathbf{z}_n$ to an allocation, is given by $\sup_{\boldsymbol{\gamma}\in\Gamma}\left[\max_{\mathbf{p}\in\mathcal{P}} \boldsymbol{\gamma}'\mathbf{p} - \boldsymbol{\gamma}' E\left(\mathbf{p}\left(\mathbf{z}_n\right)\right)\right]$. Since $\mathbf{p}(\cdot)$ has to be feasible, it must be that $A\mathbf{p}(\mathbf{z}_n) = \pi$ and $\mathbf{p}(\cdot) \geq 0$, i.e., $\mathbf{p}(\mathbf{z}_n) \in \mathcal{P}$. Therefore, $\max_{\mathbf{p}\in\mathcal{P}} \boldsymbol{\gamma}'\mathbf{p} \geq \boldsymbol{\gamma}'\mathbf{p}(\mathbf{z}_n)$ with probability 1 for every $\boldsymbol{\gamma}$. Consequently,

$$\min_{\mathbf{p}(\cdot)\in\mathcal{P}} \sup_{\boldsymbol{\gamma}\in\Gamma} \left[\max_{p\in\mathcal{P}} \boldsymbol{\gamma}'\mathbf{p} - E\left(\boldsymbol{\gamma}'\mathbf{p}\left(\mathbf{z}_n\right)\right)\right] \geq 0.$$

Regret of plug-in rule is given by

$$\sup_{\boldsymbol{\gamma}\in\Gamma}\left[\max_{\mathbf{p}:\mathbf{A}\mathbf{p}=\pi} \boldsymbol{\gamma}'\mathbf{p} - E\left(\boldsymbol{\gamma}' \arg\max_{\mathbf{p}:\mathbf{A}\mathbf{p}=\pi} \hat{\boldsymbol{\gamma}}'\mathbf{p}\right)\right] \overset{n\to\infty}{\to} \sup_{\boldsymbol{\gamma}\in\Gamma}\left[\max_{\mathbf{p}:\mathbf{A}\mathbf{p}=\pi} \boldsymbol{\gamma}'\mathbf{p} - \boldsymbol{\gamma}' E\left(\arg\max_{\mathbf{p}:\mathbf{A}\mathbf{p}=\pi} \boldsymbol{\gamma}'\mathbf{p}\right)\right] = 0,$$

where the convergence follows from compactness of $\mathcal{P}$, $\sup_{p\in\mathcal{P}}\left|\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right)'\mathbf{p}\right| \overset{P}{\to} 0$ together with a dominated convergence theorem given that $\hat{\boldsymbol{\gamma}}$, $\boldsymbol{\gamma}$ are bounded with probability 1. A more rigorous analysis of decision issues related to choice of conditioners is outside the scope of the present paper. See, e.g., Manski (2004) for a decision theoretic analysis of the unconstrained treatment choice problem.

4.4 Discussion

This subsection briefly discusses four issues which qualify the present paper's approach.

First, the analysis in this paper assumes that the underlying structural parameters (the F's) are unaffected by the allocation and also by the π's. The latter can fail to hold in example (1), if for instance, w types are more accommodating of b types if b types are a small minority but more antagonistic, otherwise.

Second, the estimates of the underlying parameters are based on the random allocation of a pilot subgroup to design a policy for a different subgroup which is assumed to come from the same population. This suggests using more covariates to define the types in order to make estimates robust to heterogeneity between the target and the pilot. Even when heterogeneity is not a concern, conditioning on more covariates is likely to yield a larger maximum value. However, as one increases the number of covariates on which to match, the data requirement becomes more severe due to familiar small-cell problems. The present paper does not analyze the optimal trade-off along this margin and works with a pre-specified set of types/covariates.

A third point concerns an alternative strategy where one specifies an explicit model of outcomes as a function of the matches, estimates it using data from randomization and then performs the optimization by using those estimates as inputs. When covariates are discrete and *all interactions are allowed for*, subgroup level means used in the present approach are one-to-one linear combinations of coefficients on the covariates and their interactions in the alternative model-based approach. While the model-based approach can accommodate prior information, e.g., that some interactions have no or little effect, such information is likely to be viewed as somewhat ad-hoc and overtly restrictive. On the other hand, pretesting these restrictions will typically offset the gain in precision from estimating a restricted model. Second, a model based approach will not solve the small-cell problem because small cells imply that coefficients on the corresponding interaction will be estimated imprecisely. Third, unlike the mean maximization case, a model-based approach for quantile maximization is problematic because, e.g.,

$$median\left(y\right) \neq \int median\left(y|x\right) dF\left(x\right).$$

It is not clear how one may use a model-based analysis for maximizing a quantile. Even with continuous covariates, a model-based approach is not particularly helpful because the number of feasible allocations is very large and a frugal model for outcome is unlikely to restrict that set sufficiently unless very strong assumptions are imposed on the model. In view of the above considerations, this paper takes the view that conditioning policy on a few covariates and leaving the model unrestricted otherwise is the right approach.

Fourth, notice that an implicit assumption here is that agents cannot reassign themselves after the initial allocation and before realization of the outcome– i.e., there is perfect compliance. So agents' own preferences over characteristics of partners are irrelevant here and the planner is fully paternalistic. This makes these problems distinct from the two-sided matching literature in Economics (c.f. Roth and Sotomayor (1990)).

4.5 Distribution Theory

This subsection develops the methodology to be used in the application for conducting inference on optimal allocations and the resulting value functions.

4.5.1 LP formulation

Notice that the constraint set $\mathcal{P}$, which is the same for the sample and population problems, is a convex polytope with extreme points corresponding to the set of basic solutions (e.g., Luenberger (1984) page 19). Let $S = \{\mathbf{z}_1, ...\mathbf{z}_{|S|}\}$ denote the set of basic feasible solutions with $|S| \leq \begin{pmatrix} m \\ M \end{pmatrix}$. Then the fundamental theorem of LP implies that

$$v = \max\{\mathbf{z}_1'\boldsymbol{\gamma}, ...\mathbf{z}_{|S|}'\boldsymbol{\gamma}\} \text{ and } \hat{v} = \max\{\mathbf{z}_1'\hat{\boldsymbol{\gamma}}, ...\mathbf{z}_{|S|}'\hat{\boldsymbol{\gamma}}\}.$$

Moreover, the compact and convex constraint set $\mathcal{P}$ equals the convex hull of set S (e.g., Luenberger (1984), page 470, theorem 3). This would imply that for identifying the optimal allocation (and, as will be shown below, for conducting asymptotic inference), one can simply concentrate on the finite set S rather than the much larger constraint set $\mathcal{P}$.

4.5.2 Uniqueness of Population Solution

For the population solution, there are three possible scenarios: (i) the degenerate case: $\mathbf{z}_1'\boldsymbol{\gamma} = \mathbf{z}_2'\boldsymbol{\gamma} = ... = \mathbf{z}_{|S|}'\boldsymbol{\gamma}$, (ii) the intermediate case: $\exists\ \mathbf{z}_1 \neq \mathbf{z}_2 \neq \mathbf{z}_3 \in \mathcal{P}$ such that $\mathbf{z}_1, \mathbf{z}_2 \in \{\mathbf{z}^* \in S : \mathbf{z}^{*\prime}\boldsymbol{\gamma} = \max_{\mathbf{z}\in S}(\boldsymbol{\gamma}'\mathbf{z}) \equiv v\}$ and $\boldsymbol{\gamma}'\mathbf{z}_1 = \boldsymbol{\gamma}'\mathbf{z}_2 = v > \boldsymbol{\gamma}'\mathbf{z}_3$ and (iii) the unique case: $\boldsymbol{\gamma}'\mathbf{z}_1 > \boldsymbol{\gamma}'\mathbf{z}_j$ for all $j \neq 1$. It should be noted that uniqueness refers to the solution and not the maximum value since the latter will be unique if it is finite. The general situation can be depicted as $\boldsymbol{\gamma}'\mathbf{z}_1 = \boldsymbol{\gamma}'\mathbf{z}_2 = ...\boldsymbol{\gamma}'\mathbf{z}_J = v$ and $\boldsymbol{\gamma}'\mathbf{z}_j < v$ for all $J < j \leq |S|$ with $J = |S|$ being case (i), $J = 1$ being case (iii) and $1 < J < |S|$ being the general case (ii).

In our example (1), case (i) will arise if, for instance, $\gamma_{ij} = (\gamma_{ii} + \gamma_{jj})/2$ for $i \neq j$, $i, j \in \{w, b, o\}$. Case (ii) will arise if, e.g., $\gamma_{ww} = \gamma_{wb} = \gamma_{bb} = a > b = \gamma_{wo} = \gamma_{bo} = \gamma_{oo}$. In that case, the value of the objective function will be $a - (a-b)\pi_o - (p_{wo} + p_{bo})(a-b)/2$

and so any allocation which gives nonzero weight to one or more of p_{wo}, p_{bo} will yield a strictly lower objective function than one which sets p_{wo}, p_{bo} equal to zero. Moreover, any allocation with $p_{wo} = p_{bo} = 0$, $p_{oo} = \pi_o$ and satisfies $2p_{ww} + p_{wb} = 2\pi_w$ and $2p_{bb} + p_{wb} = 2\pi_b$ with $p_{ww}, p_{wb}, p_{bb} \geq 0$ will be optimal and there will be at least two of them, viz. $p_{ww} = \pi_w$, $p_{bb} = \pi_b$, $p_{wb} = 0$ and $p_{wb} = \min\{\pi_w, \pi_b\}$, $p_{ww} = \frac{1}{2}(2\pi_w - p_{wb})$ and $p_{bb} = \frac{1}{2}(2\pi_b - p_{wb})$. Finally, case (iii) will arise if, e.g., $\gamma_{wb} < \min\{\gamma_{ww}, \gamma_{bb}\}$, $\gamma_{wo} < \min\{\gamma_{ww}, \gamma_{oo}\}$ and $\gamma_{bo} < \{\gamma_{oo}, \gamma_{bb}\}$.

It is easy to see that when $M = 2$ (considered in GIR (2006)), one can have either situation (i) or situation (iii) but not situation (ii). In other words, a nontrivial set-identified situation, of the type discussed here, will not arise there.

Since the $\hat{\boldsymbol{\gamma}}$'s will be asymptotically normally distributed, the sample solution will likely be unique, at least for large enough n, no matter whether the population solution is unique or not. Otherwise there will exist $\mathbf{z}_1 \neq \mathbf{z}_2$ subject to $\hat{\boldsymbol{\gamma}}'(\mathbf{z}_1 - \mathbf{z}_2) = 0$ which has 0 probability if $\hat{\boldsymbol{\gamma}}$ is continuously distributed. If the sample size is small, so that the distribution of $\hat{\boldsymbol{\gamma}}$ is far from continuous and the sample solution is nonunique, one can pick any one of the sample solutions and call it $\hat{\mathbf{z}}$.

4.5.3 The asymptotic behavior of $\hat{\boldsymbol{\gamma}}$:

Let us re-index room-type by $k = 1, ...m$ and rewrite

$$\hat{\gamma}_k = \frac{\frac{1}{n}\sum_{i=1}^{n} D_{ik} y_i}{\frac{1}{n}\sum_{i=1}^{n} D_{ik}} \equiv \frac{\bar{y}_k}{\bar{d}_k} \text{ and } \gamma_k = \frac{E(\bar{y}_k)}{E(\bar{d}_k)} = \frac{\mu_k}{\delta_k} \tag{3}$$

where D_{ik} is a dummy which equals 1 if the ith sampled individual, following random allocation, is in room type k, $i = 1, ...n$ and $k = 1, ...m$. The expectation terms in the above display correspond to the combined experiment of drawing one random sample and making one random allocation of this drawn sample. The rest of this paper will assume that $\delta_k > C > 0$ for all k, where C is a positive constant. Then $\hat{\boldsymbol{\gamma}}$ is consistent for $\boldsymbol{\gamma}$ and $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ converges in distribution to a normal $N(0, \boldsymbol{\Sigma})$ which follows from classical weak laws and CLT's under standard conditions, since $\hat{\boldsymbol{\gamma}}$'s are ratios of means. It can also be shown that asymptotically, $\hat{\gamma}_l$ and $\hat{\gamma}_k$ will be uncorrelated for $l \neq k$ even if both their numerators and both their denominators are correlated. The intuitive explanation for this is that the different components of $\hat{\boldsymbol{\gamma}}$ pertain to non-overlapping subpopulations.

4.5.4 Asymptotic Properties of Solution(s): Consistency and Rate

Let $v = \sup_{\mathbf{z}\in S} \boldsymbol{\gamma}'\mathbf{z}$ and define the set Θ_0 to be the set of solutions, i.e.,

$$\Theta_0 = \{\mathbf{z} : \mathbf{z} \in S,\ \boldsymbol{\gamma}'\mathbf{z} = v\} \equiv \{\mathbf{z}_1, ...\mathbf{z}_J\}.$$

The following proposition states and establishes consistency of the sample solution in the Hausdorff sense. This proposition is analogous to Manski and Tamer's (2002) result on consistency for set identified parameters specialized to a discrete parameter space. Let $\hat{\mathbf{z}}$ be as in section 4.2.

Proposition 1 *(consistency of solution): Assume that* $\sqrt{n}\left(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}\right) \xrightarrow{d} w \equiv N\left(0, \boldsymbol{\Sigma}\right)$. *Then (i)* $\Pr\left(\hat{\mathbf{z}} \in \Theta_0\right)$ *converges to 1. Define*

$$\hat{\Theta}_n = \left\{\mathbf{z}^* \in S : \hat{\boldsymbol{\gamma}}'\mathbf{z}^* \geq \max_{\mathbf{z}\in S} \hat{\boldsymbol{\gamma}}'\mathbf{z} - c_n\right\},$$

where c_n *is a sequence of positive constants with* $\sqrt{n}c_n \to \infty$*. Then (ii)* $\Pr\left(\Theta_0 \subset \hat{\Theta}_n\right) \to 1$*. Moreover, if* $c_n \to 0$*, then (iii)* $\Pr\left(\hat{\Theta}_n \subset \Theta_0\right) \to 1$.

Proof. Similar to Manski and Tamer (2002). ∎

Note that here one gets $\Pr\left(\hat{\Theta}_n = \Theta_0\right) \to 1$, which is a stronger version of the conclusion $d\left(\hat{\Theta}_n, \Theta_0\right) \xrightarrow{P} 0$ in Manski and Tamer (2002), where $d\left(\cdot\right)$ is the Hausdorff distance. This stronger version holds because the parameter space is finite.

Rate of convergence: Proposition 1 implies that the sample maxima $\hat{\mathbf{z}}$ is included in the set Θ_0 with probability approaching 1. The following corollary shows that the rate of convergence is arbitrarily fast. This result will be used below when establishing a distribution theory for the estimated maximum value. In particular, it will imply the somewhat surprising result that this asymptotic distribution is not affected by the sampling error in the estimation of the optimal solution.

Let the notation $I\left(A\right)$ denote the indicator for event A.

Corollary 1 *(rate of convergence): Under the hypothesis of proposition 1, for* ***any*** *sequence* $a_n \uparrow \infty$, $\Pr\left\{a_n I\left(\hat{\mathbf{z}} \notin \Theta_0\right) \neq 0\right\} \to 0$ *as* $n \to \infty$*.*

Proof.

$$\Pr\left\{a_n I\left(\hat{\mathbf{z}} \notin \Theta_0\right) \neq 0\right\} = \Pr\left[I\left(\hat{\mathbf{z}} \notin \Theta_0\right) \neq 0\right] = \Pr\left[I\left(\hat{\mathbf{z}} \notin \Theta_0\right) = 1\right] = \Pr\left[\hat{\mathbf{z}} \notin \Theta_0\right] \to 0.$$

■

This result is somewhat nonstandard and very different from the exponential rates derived in, e.g., Shapiro and de Mello (2000). For instance if there is a unique solution $\mathbf{z}_0$, then the Shapiro and de Mello results say that $\Pr(\hat{\mathbf{z}} \neq \mathbf{z}_0) \to 0$ at the rate of $e^{-\rho n}$ for some $\rho > 0$ while what is shown in the corollary above is that for each $j = 1, 2, ...m$, we have that $\Pr\left(a_n\left(\hat{\mathbf{z}}^{(j)} - \mathbf{z}_0^{(j)}\right) = 0\right) \to 0$ for any $a_n \to \infty$, where $\hat{\mathbf{z}}^{(j)}$ and $\mathbf{z}_0^{(j)}$ are the jth coordinates of the $m \times 1$ vectors $\hat{\mathbf{z}}$ and $\mathbf{z}_0$. Corollary 1 is sufficient for deriving an asymptotic distribution theory for $\hat{v}$. Furthermore, the corollary applies under bounded second moments alone which is much weaker than thin tail requirements for the large deviation type results in Shapiro and de Mello (2000).

Confidence set construction for the set of solutions is not attempted here since the methods for these are now well-known. Moreover, here one gets $\Pr\left(\hat{\Theta}_n = \Theta_0\right) \to 1$ and not just that $\Pr\left(\hat{\Theta}_n \supseteq \Theta_0\right) \to 1$ and so constructing a separate confidence set for Θ_0 with coverage probability less than 1 seems to be a somewhat unnecessary exercise.

4.5.5 Asymptotic Properties of Maximum value

This subsection derives a distributional theory for the estimated maximum value which will be used to construct a CI for the population maximum. It will be shown that asymptotically, the sampling error in estimation of optimal solutions has no effect on the distribution of the estimated optimal value– a consequence of the discreteness of the parameter space. This distribution will also be nonstandard in that it will depend on uniqueness of the solution.

Proposition 2 *(consistency of estimated maximum): If* $\sqrt{n}(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = O_p(1)$, *and* $\hat{v} = \max_{\mathbf{p} \in \mathcal{P}} \hat{\boldsymbol{\gamma}}'\mathbf{p} \equiv \max_{\mathbf{z} \in S} \hat{\boldsymbol{\gamma}}'\mathbf{z}$, *then*

$$\hat{v} - v = o_p(1)$$

where $v = \max_{\mathbf{p} \in \mathcal{P}} \boldsymbol{\gamma}'\mathbf{p}$.

Proof. Appendix. Note that the proof allows for a set-identified solution. ■

A distribution theory for the estimated maximum value is now presented. The key features of this distribution are that (i) it is not affected by sampling error in the estimation of the sample maximum $\hat{\mathbf{z}}$ and (ii) it depends qualitatively on the uniqueness of the population solution.

Proposition 3 *Assume that* $\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)\xrightarrow{d}w$ *and that elements of* $\hat{\boldsymbol{\gamma}}$ *are bounded with probability 1. Then*

$$\sqrt{n}\left(\hat{v}-v\right)=\left[\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\hat{\mathbf{z}}-\sqrt{n}\left(v-\boldsymbol{\gamma}'\hat{\mathbf{z}}\right)\right]I\left(\hat{\mathbf{z}}\in\Theta_0\right)+o_p\left(1\right), \tag{4}$$

with $\Pr\left(\hat{\mathbf{z}}\in\Theta_0\right)\to 1$*. Consequently,*

$$\sqrt{n}\left(\hat{v}-v\right)=\max_{\mathbf{z}\in\Theta_0}\left\{w'\mathbf{z}\right\}+o_p\left(1\right)\equiv\max_{1\leq j\leq J}\left\{w'\mathbf{z}_j\right\}+o_p\left(1\right). \tag{5}$$

Proof. Appendix. ■

Observe that in (4), terms involving $I\left(\hat{\mathbf{z}}\notin\Theta_0\right)$ are at most $o_p\left(1\right)$ and hence the sampling error in the estimation of $\hat{\mathbf{z}}$ does not affect the distribution of $\sqrt{n}\left(\hat{v}-v\right)$. As can be seen in the proof, this is a consequence of a discrete parameter space which ensures that $\Pr\left(\sqrt{n}I\left(\hat{\mathbf{z}}\notin\Theta_0\right)\neq 0\right)\to 0$. In particular, if $|\Theta_0|=1$, then $\sqrt{n}\left(\hat{v}-v\right)=\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}_0+o_p\left(1\right)$. In the general case where $|\Theta_0|>1$, the distribution of $\sqrt{n}\left(\hat{v}-v\right)$ will not be centered at 0.

Simulating the CI: Notice that the distribution of $\max_{\mathbf{z}\in\Theta_0}\left[\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}\right]$ is not pivotal. However, since one can estimate the asymptotic covariance matrix of $\hat{\boldsymbol{\gamma}}$, i.e., $\boldsymbol{\Sigma}$, one can simulate the distribution of $\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)$. One can then replace the unknown Θ_0 everywhere in proposition 2 by the estimated set $\hat{\Theta}_n$, i.e., first draw a w from the distribution of $\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)$ and then calculate $\max_{\mathbf{z}\in\hat{\Theta}_n}\left\{w'\mathbf{z}\right\}$. The following proposition shows that approximating $\max_{\mathbf{z}\in\Theta_0}\left\{\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}\right\}$ by $\max_{\mathbf{z}\in\hat{\Theta}_n}\left\{\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}\right\}$ does not entail any additional estimation error.

Proposition 4 *Under the assumptions of proposition 1, one has that*

$$\max_{\mathbf{z}\in\hat{\Theta}_n}\left\{\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}\right\}=\max_{\mathbf{z}\in\Theta_0}\left\{\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}\right\}+o_p\left(1\right).$$

Proof. Appendix ■

The upper and lower α percentiles of this simulated distribution, denoted by c_u and c_l, can be used to construct a CI for v, given by $CI_{nd}=\left(\hat{v}-\frac{c_u}{\sqrt{n}},\hat{v}-\frac{c_l}{\sqrt{n}}\right)$. Note that due to the asymptotic bias of $\hat{v}$ in the general case where $|\Theta_0|>1$, this CI will not be centered at $\hat{v}$.

One can calculate the bias of $\hat{v}$ by $\hat{\beta}$, the mean of the simulated distribution of $\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)$, and form a bias-corrected estimate of $\hat{v}$ as $\hat{v}_{BC}=\hat{v}-n^{-1/2}\hat{\beta}$.

4.6 Extensions

Assortative matching: Solution to the LP problem provides asymptotically a higher expected mean than any other allocation mechanism– like positive assortative (PA) and negative assortative (NA) matching or random assignment, considered in GIR (2005). This is because all these other allocations have to be feasible and the LP one maximizes the mean among all feasible allocations. Using the usual delta method and the asymptotic distributions derived above, one can also form CIs for relative efficiency of alternative allocations compared to the optimal one.

Degeneracy: Degeneracy, of the type mentioned in section 4.5.2, makes the allocation problem vacuous and so one should first check that this is not the case. In the actual application described in section 6 below, degeneracy *cannot* be rejected in any case where the outcome of interest is student GPA.

A test of degeneracy can be formulated as follows. Note that the null of degeneracy is equivalent to $\max_{\mathbf{z}\in S} \boldsymbol{\gamma}'\mathbf{z} = \min_{\mathbf{z}\in S} \boldsymbol{\gamma}'\mathbf{z}$. Under the null of degeneracy, the test statistic $T_n = \sqrt{n}\,(\max_{\mathbf{z}\in S} \hat{\boldsymbol{\gamma}}'\mathbf{z} - \min_{\mathbf{z}\in S} \hat{\boldsymbol{\gamma}}'\mathbf{z})$ will be distributed as $U - V$ where U, V are defined by $\sqrt{n}\max_{\mathbf{z}\in S}\left((\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\,\mathbf{z}\right) \xrightarrow{d} U$ and $\sqrt{n}\min_{\mathbf{z}\in S}\left((\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'\,\mathbf{z}\right) \xrightarrow{d} V$. Therefore the asymptotic null distribution of T_n can be simulated by the distribution of $\max_{\mathbf{z}\in S} w'\mathbf{z} - \min_{\mathbf{z}\in S} w'\mathbf{z}$ where w is a draw from a $N(0, \boldsymbol{\Sigma})$. It is straightforward that this test will have nontrivial power against $1/\sqrt{n}$ alternatives and will be consistent against fixed alternatives. Observe that implementing this test does not require calculating the extreme points in the mean case. One can simply solve the original LP problems to get T_n and then replace $\hat{\boldsymbol{\gamma}}$ by w in the LP problem to get $\max_{\mathbf{z}\in S} w'\mathbf{z}$ and $\min_{\mathbf{z}\in S} w'\mathbf{z}$ which are the same as $\max_{\mathbf{z}\in \mathcal{P}} w'\mathbf{z}$ and $\min_{\mathbf{z}\in \mathcal{P}} w'\mathbf{z}$ respectively.

There are other ways of testing degeneracy in the mean case, e.g., by jointly testing the coefficients of all second and higher order interactions in a mean regression of outcome on own and roommate characteristics in example 1. But these do not work for the quantile case, while the above test based on max-min is equally applicable there. Note also that the proposed test is of the hypothesis that all *feasible* allocations and not necessarily all allocations yield the same aggregate outcome. In contrast, a test based on joint significance of all interaction coefficients tests a *sufficient* condition for degeneracy.

Pretest CI: One approach to inference in an application would be to first test for degeneracy and upon rejection, construct the appropriate CIs. The overall CI is, therefore, given by

$$\hat{I} = I\,(T_n \geq c)\,CI_{nd} + I\,(T_n < c)\,CI_d,$$

where c is the critical value used in the test of degeneracy, described above, $CI_{nd} = (c_l, c_u)$ is the CI constructed upon rejection of degeneracy. The CI_d is the CI corresponding to the case when degeneracy cannot be rejected, denoted by (d_l, d_u). It is not hard to show that this pretest CI will have a coverage probability of $1 - \alpha - \alpha'$ where α' is the size of the test of degeneracy which rejects the null of degeneracy when $T_n > c$ and α is the nominal level used in constructing CI_d and CI_{nd}.

5 Quantile Maximization

The analysis for mean maximization described above is now extended to maximization of a quantile for the planner's utility distribution. This section motivates the exercise of quantile maximization and establishes the distributional analysis necessary for conducting inference on the maximal allocation and maximum value in the quantile maximization problem. What makes quantile maximization analytically a nonstandard exercise is that it cannot be interpreted as an M-estimation problem, unlike, e.g., quantile regressions.

To formalize the quantile maximization problem, let $F^{(\mathbf{p})}(\cdot)$ denote the overall c.d.f. of individual outcomes induced by the allocation $\mathbf{p}$, i.e., $F^{(\mathbf{p})}(a) \equiv \sum_{j=1}^{m} p_j F_j(a)$, where $F_j(a)$ is the distribution of individual outcomes across room type j, $j = 1, 2, ...m$. Then the τth quantile $\mu_{\mathbf{p}}^{(\tau)}$ of the overall outcome distribution induced by the allocation $\mathbf{p}$ may be defined as

$$\mu_{\mathbf{p}}^{(\tau)} \equiv \left[F^{(\mathbf{p})}\right]^{-1}(\tau) = \inf_{\mu} \left(\sum_{j=1}^{m} p_j F_j(\mu) \geq \tau \right), \tag{6}$$

where $0 \leq \tau \leq 1$. Then, analogous to (2) for mean maximization, the population problem for maximizing the τth quantile has the form

$$\max_{p} \mu_{p}^{(\tau)} \text{ subject to } \mathbf{Ap} = \boldsymbol{\pi},\ \mathbf{p} \geq \mathbf{0}. \tag{7}$$

The corresponding sample problem will replace $\mu_{\mathbf{p}}^{(\tau)}$ by the sample analog

$$\hat{\mu}_{\mathbf{p}}^{(\tau)} = \inf_{\mu} \left(\sum_{j=1}^{m} p_j \hat{F}_j(\mu) \geq \tau \right) \text{ where } \hat{F}_j(\mu) = \frac{\sum_{k=1}^{n} I(y_k \leq \mu) \times 1(k,j)}{\sum_{k=1}^{n} 1(k,j)},$$

where $1(k,j)$ equals 1 if individual k belongs to group type j and is zero otherwise.

5.1 Motivation for quantile maximization

From a purely statistical standpoint, quantiles are robust to outliers unlike means, so that quantile based decisions are based on targeting a more "representative" section of the population. Secondly, as argued in Manski (1988), quantile maximization is robust to ordinal transformations of the planner's utility. In particular, suppose that the planner's utility from an outcome is captured by the function $u(\cdot)$ and the c.d.f. of the realizations of the planner's utility under two distinct allocations $\mathbf{p}$ and $\mathbf{q}$ are $F^{(\mathbf{p})}(\cdot)$ and $F^{(\mathbf{q})}(\cdot)$, respectively. Assume that the planner's preferences between $\mathbf{p}$ and $\mathbf{q}$ are based on the αth quantile of $F^{(\mathbf{p})}(\cdot)$ and $F^{(\mathbf{q})}(\cdot)$, respectively, i.e., $\mathbf{p} \preceq \mathbf{q}$ if and only if $\left[F^{(\mathbf{p})}\right]^{-1}(\alpha) \leq \left[F^{(\mathbf{q})}\right]^{-1}(\alpha)$. Then if $\mathbf{q} \succeq \mathbf{p}$ under utility $u(\cdot)$, then $\mathbf{q}$ will continue to be weakly preferred over $\mathbf{p}$ under a monotone increasing transformation of $u(\cdot)$. This follows from the well-known fact that quantiles "go through" monotone transformations. That monotone transformations of utility should reflect identical preferences is generally considered a desirable property of preference patterns in traditional theory of choice. Preferences represented by expected, rather than quantile, utility do not possess this robustness property. Manski (1988) discusses these issues in greater details. Interestingly, the ordering of planners by which quantile they want to maximize orders them exactly by their tolerance of risk (c.f. Manski (1988), section 4). In particular, judging allocations in terms of the lower quantiles of the utility distributions which they generate reflects a greater degree of risk aversion. This shows that quantile utility maximization can be an interesting parallel criterion for planner choice problems under uncertainty. Relative to expected utility maximization, it has additional robustness properties and yet retains all of the key elements that make mean maximization a well-defined criterion for choice under uncertainty. It is also interesting to note that quantiles will be finite for any choice of a planner's utility function whereas for expected utility to be finite the utility distribution must have thin enough tails. The only apparent drawback of quantile maximization (c.f. Manski (1988), section 7), viz. its analytic intractability for analyzing *compound* lotteries, which is irrelevant to the empirical problem the present paper aims to study.

Policy-makers are typically concerned about distributional equity to some extent and maximizing the lower quantiles of the distribution is one of several alternative ways of achieving this (maximizing expectations of concave transformations of outcomes can also address this goal, as a referee has correctly pointed out). It is also interesting to compare the allocations that lead to maximizing upper and lower quantiles with those that maximize the mean or median, since this reveals whether any trade-off exists between productive efficiency and distributional equity. Since this paper is concerned with situations where output (e.g., GPA) cannot be redistributed, the distributional consequences of alternative allocations become all the more important.

5.2 Formal analysis of quantile maximization

This subsection presents the formal statistical analysis of quantile maximization in the context of the above allocation problem. For simplicity of notation the analysis in this section focuses on the median (i.e $\tau = 0.5$ and so the superscript on μ is dropped) but the methods presented here apply to any fixed quantile.

The first proposition will establish that the extreme point idea, which is key to the mean analysis above, extends to the median. The underlying reason is that if $\mathbf{r} = \lambda\mathbf{p} + (1-\lambda)\,\mathbf{q} \in \mathcal{P}$ with $\lambda \in (0,1)$, then the median corresponding to the allocation $\mathbf{r}$ lies "between" the medians corresponding to the allocations $\mathbf{p}$ and $\mathbf{q}$. This is geometrically illustrated in Fig. 1 where $F^{(\mathbf{p})}$, $F^{(\mathbf{q})}$ and $F^{(\mathbf{r})}$ are the c.d.f.'s corresponding to allocations $\mathbf{p}$, $\mathbf{q}$ and $\mathbf{r} = \lambda\mathbf{p} + (1-\lambda)\,\mathbf{q}$, respectively. This insight will be key to the rest of this section.

A formal statement and its proof now follow.

Proposition 5 *Let* $\mathbf{p}, \mathbf{q} \in \mathcal{P}$. *Let* $\mathbf{r} = \lambda\mathbf{p} + (1-\lambda)\,\mathbf{q} \in \mathcal{P}$ *with* $\lambda \in (0,1)$. *Then* $\mathbf{r}$ *cannot be the unique solution to the population problem (7).*

Proof. Appendix ■

So quantile maximization cannot occur at an interior point of $\mathcal{P}$ because an interior point is a strict convex combination of the extreme points of $\mathcal{P}$. Thus it follows from the above proposition that for quantile maximization, just like in the mean case, one can focus on the extreme points of the constraint set which can be calculated a priori. Further, notice that the only property of the F_j's used in the above proof is that they are nondecreasing. So exactly the same result will hold for the sample problem since the estimates $\hat{F}_j$ are also nondecreasing.

A general characterization: One can generalize the previous proposition to any functional $\Lambda(\cdot)$ of the C.D.F. which is quasi-convex, i.e., satisfies

$$\Lambda\left(F^{(\lambda\mathbf{p}+(1-\lambda)\mathbf{q})}\right) \leq \max\left\{\Lambda\left(F^{(\mathbf{p})}\right), \Lambda\left(F^{(\mathbf{q})}\right)\right\}$$

for every $\lambda \in [0,1]$. Means and quantiles are the two leading examples which are obvious policy targets. But inter-quantile differences do not satisfy this property in general.

Asymptotic properties for quantile maxima and value: Essentially all the results for the mean case will hold here because of the analogous discretization of the parameter space. One can therefore replicate all the propositions for the mean in the quantile case simply by setting

$$\Theta_0 = \left\{ \mathbf{z} = \{\mathbf{z}_1, ...\mathbf{z}_J\} : \mathbf{z} \in S,\, v = \mu_{\mathbf{z}_j} \right\} \text{ and } \hat{\Theta}_n = \left\{ \mathbf{z}^* \in S : \hat{\mu}_{\mathbf{z}^*} \geq \max_{\mathbf{z} \in S} \hat{\mu}_{\mathbf{z}} - c_n \right\}, \quad (8)$$

and replacing $\boldsymbol{\gamma}'\mathbf{z}$ by $\mu_{\mathbf{z}}$ which equals the median corresponding to the c.d.f. $\mathbf{z}'\mathbf{F}$, where $\mathbf{F}(\cdot)= (F_j\,(\cdot))_{j=1,...m}$.

The only step left to be specified is the asymptotic distribution of $\sqrt{n}\left(\hat{\mu}_{\mathbf{p}} - \mu_{\mathbf{p}}\right)$ for any fixed $\mathbf{p}$. To this end, define $\alpha_{\mathbf{p}j} = E\left(D_{ij} I\left(Y_i \leq \mu_{\mathbf{p}}\right)\right)$, $f_j\,(\cdot) = \frac{\partial}{\partial x} F_j\,(\cdot)$, assuming that these quantities exist where evaluated. It then follows from standard Taylor expansions and results on asymptotic normality of medians that

$$\begin{aligned} \sqrt{n}\left(\hat{\mu}_{\mathbf{p}} - \mu_{\mathbf{p}}\right) &= \frac{1}{\sqrt{n}} \sum_{i=1}^{n} v_i + o_p\,(1) \text{ where} \\ v_i &= -\left(\sum_{j=1}^{m} \mathbf{p}_j f_j\left(\mu_{\mathbf{p}}\right) \right)^{-1} \sum_{j=1}^{m} p_j D_{ij} \left(\frac{I\left(Y_i \leq \mu_{\mathbf{p}}\right)}{\delta_j} - \frac{\alpha_{pj}}{\delta_j^2} \right). \end{aligned}$$

6 Results

The methodology is now applied to calculate optimal allocation of freshmen to dorm rooms based on the data from Dartmouth College, which was introduced in section 2. Table 1 reports the summary statistics for the relevant variables in the dataset. In the allocation problems considered below, the variables GPA and fraternity/sorority joining are used as outcomes and the variables ACA and those pertaining to race are used as covariates on which the allocation is conditioned.

First consider the case where the policy covariate is ACA. This variable assumes values between 151 and 231 in the data. Discreteness is imposed by dividing the sample into several ranges of ACA and results are shown for 2, 3 and 4 categories. The cut-off points were chosen to equalize the number of individuals across the categories. The corresponding results are shown in table 2 separately for men and women when the outcome of interest is mean freshman year GPA and in table 3A, 3B for the outcome "joining a fraternity/sorority in junior year".

When constructing CIs based on the discussion in section 4.5, a sequence $c_n = k \ln(n) / \sqrt{n}$ for three values of k- 0.01, 0.1 and 1 were used. For the application studied in this paper, n is approximately 400, $\ln(400)$ is about 6 and $\sqrt{n}$ is about 20. So this range of values of k seem reasonable and implies values of c_n equal to 0.003, 0.03 and 0.3. To compute the extreme points where necessary, the "cdd" routines written by Komei

Fukuda and coauthors were used. These routines are downloadable free-of-charge from the website: http://www.ifor.math.ethz.ch/~fukuda/cdd_home/

Column 1 in table 2 reports the p-value for a test of degeneracy (for the GPA outcome) described in section 4.6. As is seen in table 2, degeneracy cannot be rejected anywhere. In table 3A, the same exercise is repeated for the outcome "joining a Greek organization". Degeneracy is rejected at 5% or 10% level of significance for a small (2) number of categories but not for the larger ones. This is a consequence of the fact that one loses precision as the number of categories rises. Columns (3), (4) and (5) respectively report the number of vertices (equal to $|S|$ in our notation above), the tolerance level c_n and the cardinality of Θ_n for each choice of c_n. The corresponding 95% CI are reported in column (7). Note that in all tables 95% is to be interpreted as a "pretest" one, as discussed in section 4.6.

Table 3B describes the nature of the maximizing allocations corresponding to the cases where degeneracy was rejected at 5% or 10% in table 3A. Columns 2 and 3 of table 3B report the fractions of rooms with two highest ACA types and two lowest ACA types. For instance, in the first row (0.51, 0.49) means that the allocation which achieves the maximum probability of fraternity enrolment is where 51% of the rooms have two high types and 49% of the rooms have two low-types (implying that no room is mixed). For men, as can be seen from the last two columns of table 3A, the minimum probability of joining a fraternity is achieved when no room has two students from the bottom category and few rooms with two students from the very top category. This happens because a very low type experiences a larger decline in its propensity to join a Greek house when it moves in with a high type relative to how much increase the high type experiences when he moves in with a low ACA type. Overall, this can be interpreted as a recommendation for more mixed rooms for men, if the planner wants to reduce the probability of joining Greek houses.

For women, the picture is exactly the opposite– more segregation in terms of previous academic achievement seems to produce lower enrollment into Greek houses. One likely explanation (and this is simply a conjecture) might be that women utilize Greek houses in different ways than men and look to them for psychological comfort. When forced to live with someone very different from her, she seeks a comfort-group outside her room and becomes more likely to join a sorority which contains more women like herself. For men, peers like oneself reinforce one's tendencies and this effect dominates.

Note that by imposing an arbitrary discreteness on ACA, one is getting a smaller maximum or higher minimum than if a finer partitioning was allowed. On the other hand, a finer partition implies that within-category means or c.d.f.'s will be estimated less precisely and the asymptotic approximations will be poor, since the number of observations per category will tend to be very small (maybe even zero for some cells). Thus there is a trade-off here between getting a higher maximum and being able to estimate the corresponding allocation (recall the second point in section 2.4 above). Indeed, as the number of ACA categories is increased from 2 to 4 in table 3A, tests of degeneracy fail to be rejected precisely due to imprecise estimation even if with 4 categories, we have a larger maximum and smaller minimum than with 2 categories.

Optimization exercises with race are reported in tables 4 and 5. The results here are sensitive to the definition of race. First consider allocation based on the dichotomous covariate whether the student belongs to an under-represented minority (Black, Hispanic and Native Indian) or not. Table 4 shows that when the outcome variable is GPA, degeneracy cannot be rejected. Corresponding results are reported in table 5 when the outcome is is joining a Greek house. There seems to be no degeneracy and optimal solutions are very similar to the ones obtained with ACA as the covariate. The mean probability of joining a sorority is minimum when segregation is maximum and that of joining a fraternity is minimum when dorms are almost completely mixed with no two individuals from the minorities staying with one another. When race is classified as white and others, degeneracy cannot be rejected for either GPA or joining fraternities. Other categorizations such as "non-blacks" and "blacks" caused problems due to very small cell frequencies.

Tables 6A-B show the efficiency loss arising out of constraining the allocation. The outcome here is joining Greek houses and the minimal allocations are compared with a completely random allocation which is blind to student characteristics. If students are classified by ACA as H and L with aggregate proportions π_H and π_L, then under random allocation, expected fractions of the three types of rooms will be $p_{LL} = \pi_L^2$, $p_{HL} = 2\pi_H\pi_L$ and $p_{HH} = \pi_H^2$. One can then compare the value function arising from minimizing the probability of joining frats/sororities with the one arising from the random allocation and calculate the difference. This is an estimate of efficiency loss due to the inability to condition on ACA. Tables 6A-B report the efficiency loss and a 95% CI for it, when the conditioning is on categories of ACA (table 6A) and race (table 6B). It is seen that the efficiency loss is significantly different from zero in all these cases with the loss being larger for women. The reason, of course, is that the extent of nonlinearity in peer effects are larger for women.

Finally, table 7 investigates the nature of maximizing allocations when the outcomes of interest are quantiles of the first year GPA distribution. Table 7 considers the case of 2 categories- H (top half) and L(bottom half)– for ACA and report the share of HH, HL and LL type rooms in the maximal allocations. For instance, the column corresponding to 10%-ile shows that the sample maximum value of 2.63 is attained when the allocation HH equals 3.6% and HL equals 96.4% with no LL type room. Analogous numbers are reported for the mean, medians and 90th percentiles. A test of degeneracy reveals that both the 10th percentile and 90th percentile problems are nondegenerate– i.e., the maximum and minimum values are distinct but both the mean and median problems are degenerate, i.e., one cannot reject that all allocations yield the same mean and median. The results reported in table 7 reveal that segregation by ACA leads to maximization of higher percentiles and minimization of lower percentiles for both men and women– i.e., segregation seems to benefit students at the upper end of the distribution at the cost of those at the lower end.

A caveat to the above application (discussed more elaborately in appendix 2) is that room assignment at Dartmouth was random, conditional on three self-reported binary covariates– smoking habit, cleanliness and music playing– in addition to gender. Estimates of mean outcomes within roomtypes defined by all these categories lead to severe small cell problems. On the other hand, ignoring them can potentially lead to suboptimal allocations unless a (different) type of separability restriction holds for the production function (see appendix 2). The separability restrictions are more plausible for some outcomes and less so for others. The bottom line, however, is that such restrictions are much more likely to hold when the outcome is joining a Greek organization– the case where optimal allocations were found to be nondegenerate. Thus results for that outcome, reported in tables 3A-B, 5A-B, 6A-B, are potentially more robust than those for GPA maximization to ignoring these additional covariates.

7 Summary and Conclusion

This paper empirically studies the problem of optimally dividing individuals into peer groups in order to maximize social gains from heterogeneous peer effects. The specific setting analyzed in the paper concerns optimal ways of pairing up roommates in college dormitories. Using data on a sample of Dartmouth College freshmen, previously analyzed by Sacerdote (2001) to detect causal effects of peer characteristics, the paper estimates optimal roommate allocation based on race and pre-college academic standing. In order to conduct inference on these sample-based optimal solutions and the resulting optimized values, the paper draws on and extends insights from the mathematical programming literature. The methods of inference are nonstandard owing to (i) discreteness of the effective parameter space and (ii) the possibility that the population optimization problems have non-unique solutions. These inferential methods are applicable to both mean maximization problems which have LP structures and, more interestingly, to quantile maximization where the objective function is shown to be quasi-convex in the allocation probabilities. Developing a general asymptotic theory of inference for value functions in presence of nonunique population solutions is outside the scope of the present paper, given its applied focus, and is left to future research.

Empirical analysis of the Dartmouth data based on these methods reveals several interesting findings. In particular, segregation by previous background is seen to affect social behavior of average men and women in exactly opposite ways. For women, this leads to lowest propensity of joining social organizations like sororities whereas for men, the propensity to join fraternities is maximized. When the outcome of interest is freshman year GPA, it is seen that for both men and women, segregation by prior academic achievement leads to maximizing higher percentiles and minimizing lower percentiles of the GPA distribution without affecting the mean or the median. Thus, a policy-maker who wants to reduce GPA inequality and discourage Greek enrolment, faces a policy dilemma in the case of women. This is because segregation increases inequality in GPA but reduces mean enrolment into sororities. For men, segregation both increases inequality and increases the probability of joining fraternities and so the policy recommendation is clear.

Finally, it is also seen that efficiency loss due to political constraints on allocations (e.g., race-blindness) can be significant and this loss tends to be larger for women for whom peer effects appear to be more nonlinear, compared to men. These empirically observed men-women differences in Greek house joining is an interesting fact and is open to further investigation using alternative datasets. It is worth noting that there do not appear to be any gender differences in the nature of peer-effects on freshman year GPA.

Randomized field experiments are becoming increasingly popular among social scientists, especially applied microeconomists, c.f. Duflo (2005). They provide the ideal source of data to which these methods can be applied for designing optimal policies. In nonrandomized observational studies, one could use instrumental variables to estimate underlying (average) structural functions which could, in turn, be utilized to compute optimal allocations.

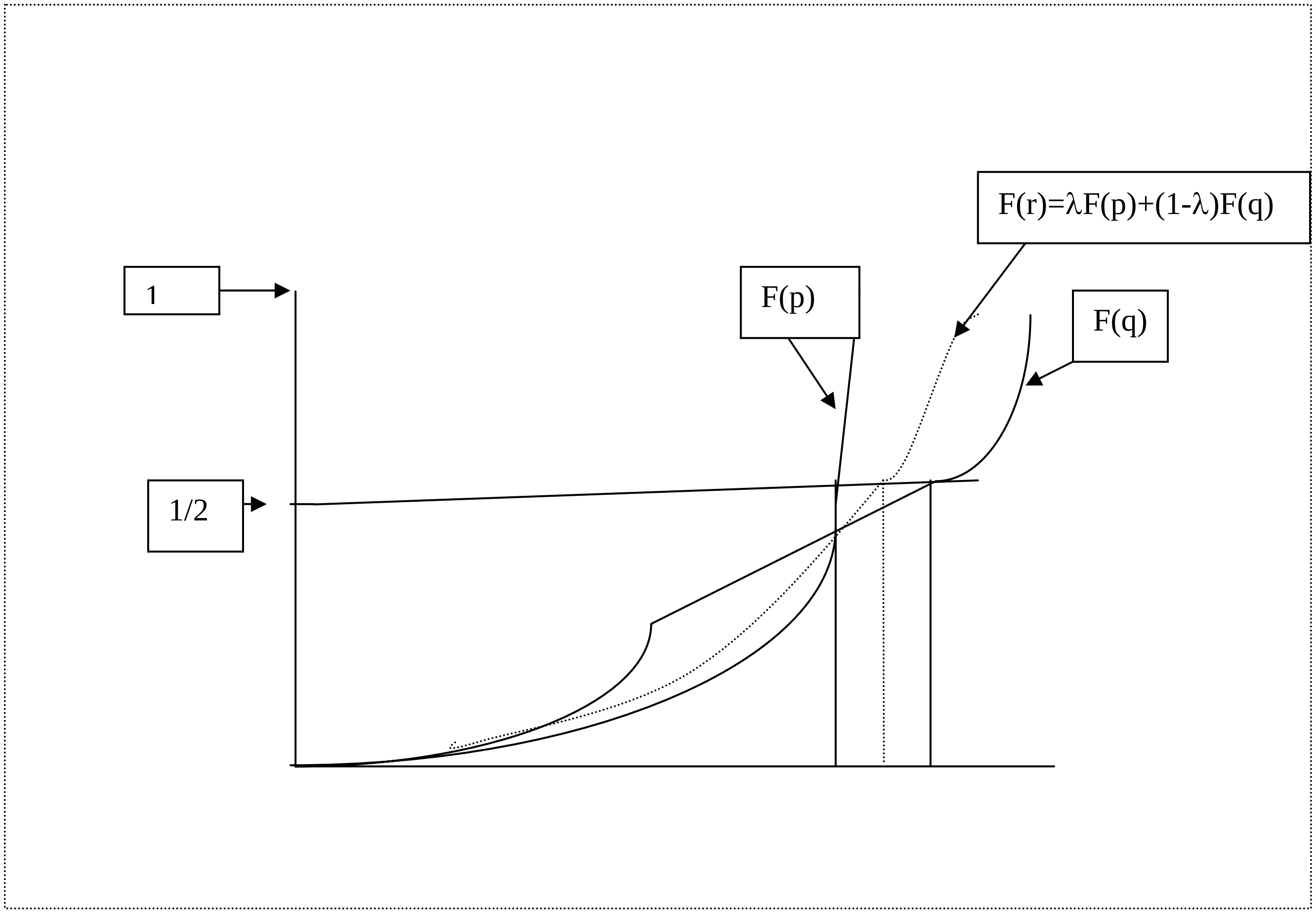

Fig. 1: Quasi-convexity of quantiles

Note: This figure illustrates proposition 5. The bolder lines represent two c.d.f.'s corresponding to F(p) and F(q) and the lighter line represents F(r) which is simply λF(p)+(1-λ)F(q). The median for the distribution F(r) (i.e. the point on the horizontal axis where F(r) equals 0.5) lies between the medians of F(p) and F(q). The result holds for any other quantile as well.

Table 1: Sample Characteristics

Variable	Mean	SD	Range
Women (N=428)			
ACA	202.68	12.65	156, 228
White	0.69	0.46	0, 1
Non-minority	0.87	0.33	0, 1
Freshman_GPA	3.23	0.39	1.56, 4.0
Sorority	0.47	0.5	0, 1
Men (N=436)			
ACA	205.55	12.99	151, 231
White	0.72	0.45	0, 1
Non-minority	0.91	0.29	0, 1
Freshman_GPA	3.153	0.45	1.15, 3.9
Fraternity	0.53	0.48	0,1

Table 2: Degeneracy, Outcome=GPA, Conditioning covariate=ACA

(0)	(1)	(2)	(3)
# categories	p-value	Sample Max	Sample Min
Men (mean=3.15)			
2	0.58	3.167	3.14
3	0.95	3.17	3.11
4	0.87	3.205	3.095
Women (mean=3.235)			
2	0.75	3.238	3.234
3	0.93	3.25	3.15
4	0.81	3.275	3.125

Notes: This table shows the results of trying to maximize and minimize freshman year GPA by allocating students based on ACA. Column (0) lists the number of categories into which ACA is divided prior to attempting reallocation. Column (1) reports p-values for test of degeneracy, column (2) shows the maximized and column (3) shows the minimized values in the sample. From column (1), it is seen that degeneracy cannot be rejected anywhere.

Table 3A: Outcome=Joining Frat, Conditioning covariate=ACA

(0) # Categories	(1) p-value	(2) Sample Min	(3) Max-Min	(4) # vertices	(5) Toler	(6) Cardinality	(7) "95%" CI
Men (mean=0.53)							
2	0.07	0.487	0.084	2	.	.	0.483, 0.490^^
3	0.68	0.51**	0.05	8	.	.	
4	0.215	0.463	0.163	37			
Women (mean=0.47)							
2	0.053	0.423	0.09	2			0.419, 0.427^^
3	0.029	0.38	0.16	8	0.003	1	0.377, 0.385
					0.03	2	0.376, 0.384
					0.3	8	0.377, 0.386
4	0.35	0.399^	0.106	41			

** Bias-corrected=0.506, ^ Bias-corrected=0.403, ^^ Normal CI

Notes: This table shows the results of minimizing the probability of joining fraternity/sorority by reallocation based on the student's ACA. Each row in column 0 reports the number of categories into which ACA is divided before the reallocation optima is sought. Column 1 contains p-value for test of degeneracy. A small p-value implies that degeneracy can be rejected. Column 2 contains estimated minimum value of the probability of joining a frat/sorority. Column 4 contains the number of vertices of the convex polytope that is the feasible set of allocations. Column 5 contains the tolerance level used in the definition of the estimated set of maxima (denoted by theta_hat_n in the text) and column 6 contains the corresponding number of vertices (denoted J in the text) in the estimated set. Column 6 contains an appropriate 2-sided 95% confidence interval- when the cardinality of the estimated set is 1, it is the usual CI based on normal approximation; when the cardinality excceds 1, it is the CI described just below proposition 4. When the p-value is large (i.e. degeneracy cannot be rejected), no further analysis is pursued.

Table 3B: Outcome=Joining frat, Conditioning covariate=ACA

(0)	(1)	(2)	(3)
# Categories	p-value	HH and LL: Max	HH and LL: Min
Men (mean=0.53)			
2	0.07	0.51, 0.49	0.037, 0.00
Women (mean=0.47)			
2	0.05	0.037, 0	0.52, 0.48
3	0.04	0.00, 0.00	0.11, 0.28

Notes: This table shows the nature of allocations which maximize and minimize the probabilities of joining a frat/sorority corresponding to the cases from table 4A where degeneracy was rejected. Column 1 reproduces the p-values from table 4A. Column 2 (3) shows the proportion of rooms which contain 2 highest type and 2 lowest type members corresponding to the allocation which maximizes (minimizes) the sample probability of enrolling in a frat/sorority. Column 2, last entry shows for example that for the probability maximizing allocation, no room should contain two high types and two low type women when all women are classified into high, medium and low types. This table suggests that segregation leads to maximization of fraternity joining for men and minimization of sorority joining for women.

Table 4: Outcome=GPA, Conditioning covariate =Race: Non-minorities (L), minorities (H)

(0)	(1)	(2)	(3)
	p-value	Sample Max	Sample Min
Men (mean=3.15)	0.611	3.15	3.149
Women (mean=3.237)	0.76	3.238	3.234

Notes: This table shows the results of maximizing mean freshman year GPA by reallocation based on the student's race. The race of each student is classified as minority (Black, Hispanic, Native Indian) and non-minority (all others). Column 1 contains the p-value for test of degeneracy. Degeneracy cannot be rejected anywhere.

.

Table 5: Outcome =Joining Frat, Conditioning covariate =Race: Non-minorities (L), minorities (H)

(0)	(1)	(2)	(3)	(4)	(5)	(6)
	# of vertices	p-value	Sample Max	CI	HL: Max	HL: Min
Men (mean=0.53)	2	0.06	0.55	0.5505, 0.5557	0.0	0.23
Women (mean=0.47)	2	0.03	0.48	0.476, 0.482	0.22	0.0

Notes: This table shows the results of maximizing the probability of joining fraternity/sorority by reallocation based on the student's race. The race of each student is classified as minority (Black, Hispanic, Native Indian) and non-minority (all others). Column 1 contains the number of extreme points of the feasible set. Column 2 contains p-value for test of degeneracy. A small p-value implies that degeneracy can be rejected. Column 3 contains estimated maximum value of the probability of joining a frat/sorority. Column 4 contains the 2-sided 95% confidence interval for the maximum value. Columns 5 and 6 show the nature of allocations which maximize and minimize the probabilities of joining a frat/sorority. Column 5 (6) shows the proportion of rooms which contain 1 monority and 1 non-minority student, corresponding to the allocation which maximizes (minimizes) the sample probability of enrolling in a frat/sorority. Column 5, first entry shows for example that for the probability-maximizing allocation, no room should be mixed.

Table 6A: Efficiency Loss: Outcome=Joining Frat , Conditioning covariate =ACA

	(1)	(2)	(3)	(4)
# categories	Sample Min Value	Random Allocation Value	Efficiency Loss	Confidence Interval
Men (mean=0.53)				
2	0.487	0.528	0.041	0.038, 0.043
Women (mean=0.47)				
2	0.465	0.518	0.052	0.050, 0.055

Notes: The table shows estimates of efficiency loss when room allocation cannot be based on ACA. Column 1 lists the minimized value when ACA is classified as high or low. Column 2 lists the estimate of expected probability of joining a fraternity/sorority when room allocation is fully random. Column 3 lists the increase in probability which is the efficiency loss of not being able to condition allocation on ACA and column 4 provides a 95% confidence interval for this loss.

Table 6B: Efficiency Loss: Outcome=Joining Frat , Conditioning covariate: Minorities/Non-Minorities

	(1)	(2)	(3)	(4)
	Sample Min Value	Random Allocation Value	Efficiency Loss	Confidence Interval
Men (mean=0.53)	0.525	0.527	0.002	0.0011, 0.0038
Women (mean=0.47)	0.441	0.474	0.033	0.0333, 0.0336

Notes: The table shows estimates of efficiency loss when room allocation cannot be based on race. Column 1 lists the minimized value when race is classified as minority or non-minority. Column 2 lists the estimate of expected probability of joining a fraternity/sorority when room allocation is fully random. Column 3 lists the increase in probability which is the efficiency loss of not being able to condition allocation on race and column 4 provides a 95% CI for this loss.

Table 7: Maximal allocations: Outcome=GPA, Conditioning covariate=ACA, # Categories=2

(0)	(1)	(2)	(3)	(4)
	Mean	10 %-ile	50%-ile	90%-ile
Men (sample mean=3.15)				
Value	3.167	2.63	3.22	3.74
HH	0.036	0.036	0.036	0.518
HL	0.963	0.963	0.963	0
LL	0	0	0	0.482
Women (sample mean=3.235)				
Value	3.24	2.71	3.3	3.7
HH	0.518	0.037	0.52	0.52
HL	0	0.962	0	0
LL	0.481	0	0.48	0.48

Notes: This table shows the nature of allocations which maximize and minimize various summary measures of GPA based on categorizing ACA into high (H) and low (L). Within each panel, the first row gives the maximum value and rows 2, 3, 4 give the proportion of each type of room in the maximizing allocation. Column (1) shows results for the mean while columns 2-4 show the 10^{th}, 50^{th} and 90 percentiles. Column 2 for men, for instance, shows that the maximized value of the 10^{th} percentile of GPA is 2.63. This is achieved by an allocation which sets the proportion of HH type rooms to 0.036, HL to 0.963 and LL to 0. Overall, desegregation is seen to maximize lower quantiles and minimize upper quantiles for both men and women.

References

[1] Aitchison, J., and Silvey, SD (1958): Maximum-likelihood estimation of parameters subject to restraints. Annals of Mathematical Statistics, 29, pages 813-828.

[2] Bandiera, O. and I. Rasul (2006): Social networks and technology adoption in northern Mozambique, Economic Journal, Volume: 116 Issue: 514 Pages: 869-902.

[3] Bartholomew, R.J. (1961): A test of homogeneity of means under ordered alternatives, Journal of the Royal Statistical Society: Series B, vol. 23 no. 2, pp. 239-281.

[4] Bertsekas, D. (2005): Dynamic Programming and Optimal Control, Athena Scientific, Belmont, MA.

[5] Bhattacharya, D. and P. Dupas (2008): Nonparametric analysis of efficient treatment assignment under budget constraints, mimeo. Dartmouth College.

[6] Birge, J.R. & F. Louveaux (1997): Introduction to stochastic programming, Springer.

[7] Dehejia, Rajeev H (2005): Program Evaluation as a Decision Problem, Journal of Econometrics, vol. 125, no. 1-2, pp. 141-73.

[8] Duflo, E. (2005): Field Experiments in Development Economics, 2005, mimeo.

[9] Fletcher (2007): Social multipliers in sexual initiation decisions among US high school students, Demography, Volume 44 Issue 2, Pages 373-388.

[10] Fukuda, Komei: cdd routines, http://www.ifor.math.ethz.ch/~fukuda/cdd_home/cdd.html

[11] Graham, B. (2008): Identifying social interactions through conditional variance restrictions, Econometrica, Volume: 76 Issue: 3, Pages: 643-660.

[12] Graham, B., G. Imbens and G. Ridder (2005): Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis, manuscript.

[13] Graham, B., G. Imbens and G. Ridder (2006): Complementarity and the Optimal Allocation of Inputs, manuscript.

[14] Hirano, K. and J. Porter (2005): Asymptotics for Statistical Treatment Rules. Manuscript.

[15] Hoxby, C. (2002): The power of peers: how does the makeup of a classroom influence achievement, Education Next 2 (2) (2002), pp. 56–63.

[16] King, A. and R. Rockafellar (1993): Asymptotic theory for solutions in statistical estimation and stochastic programming, Mathematics of Operations Research, Volume: 18 Issue: 1 Pages: 148-162

[17] Luenberger, D. (1984): Linear and nonlinear programming. Addison Wesley.

[18] Manski, C. (1988): Ordinal utility models of decision making under uncertainty, Theory and Decision, 25, pp. 79-104.

[19] Manski, C. (1995): Identification problems in the social sciences, Cambridge and London: Harvard University Press, 1995.

[20] Manski, C. (2004):Statistical Treatment Rules for Heterogeneous Populations, Econometrica, vol. 72, no. 4, pp. 1221-46.

[21] Manski, C. (2008): Identification for prediction and decision, Harvard University Press.

[22] Manski and Tamer (2002): Inference on Regressions with Interval Data on a Regressor or Outcome, Econometrica, 70, 519-546.

[23] Mas, A. and E. Moretti (2008): Peers at work, forthcoming, American Economic Review.

[24] Rosen, A. (2006): Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities, mimeo.

[25] Rostek, M. (2005): Quantile maximization in decision theory, manuscript.

[26] Roth, A. and M. Sotomayor (1990): Two-sided matching, Cambridge University Press.

[27] Sacerdote, B. (2001): Peer Effects with Random Assignment: Results for Dartmouth Roommates, Quarterly Journal of Economics, vol. 116, no. 2, May 2001, pp. 681-704.

[28] Shapiro, A. (1989): Asymptotic properties of statistical estimators in stochastic programming, Annals of statistics, vol. 17, No. 2, pp. 841-858.

[29] Shapiro, A. (1993): Asymptotic Behavior of Optimal Solutions in Stochastic Programming, Mathematics of Operations Research, Vol. 18, No. 4, pp. 829-845

[30] Shapiro, A. and T. Homem-de Mello (2000): On the rate of convergence of optimal solutions of Monte Carlo approximations of stochastic programs, SIAM journal on optimization, 11, pp. 70-86.

[31] Silvapulle, M. and P. K. Sen (2005): Constrained statistical inference, Wiley.

[32] Winston, G. and Zimmerman, D. J. (2003): in (C. Hoxby ed) Peer Effects in Higher Education. College Decisions: How Students Actually Make Them and How They Could, University of Chicago Press & National Bureau of Economic Research, Chicago.

[33] Wolak, F. (1987): Testing inequality constraints in linear econometric models, Journal of econometrics, volume 41, issue 2, pp. 205-35.

8 Appendix:

8.1 Proofs

Proof of Proposition 2 (consistency of estimated maximum). Let $\hat{\mathbf{z}} = \arg\max_{\mathbf{z}\in S} \hat{\boldsymbol{\gamma}}'\mathbf{z}$. Then

$$\hat{v} - v = (\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\mathbf{z}} - \sqrt{n}\,(v - \boldsymbol{\gamma}'\hat{\mathbf{z}}) \times 1/\sqrt{n}.$$

Now, $\Pr(\hat{\mathbf{z}} \in \Theta_0) \to 1$ (from proposition 1), implying $\sqrt{n}\,(v - \boldsymbol{\gamma}'\hat{\mathbf{z}}) = O_p(1)$; also $\hat{\mathbf{z}}$ is bounded almost surely (since $\mathcal{P}$ is compact) and $(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = o_p(1)$. So the RHS of the previous display is of the form $o_p(1) - O_p(1) \times o(1) = o_p(1)$. ■

Proof of Proposition 3. Observe that

$$\begin{aligned}
& \sqrt{n}\,(\hat{v} - v) \\
= \; & \sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\mathbf{z}} + \sqrt{n}\,(\boldsymbol{\gamma}'\hat{\mathbf{z}} - v) \\
= \; & \left[\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\mathbf{z}} + \sqrt{n}\,(\boldsymbol{\gamma}'\hat{\mathbf{z}} - v)\right] I(\hat{\mathbf{z}} \in \Theta_0) + \left[\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\mathbf{z}} + \sqrt{n}\,(\boldsymbol{\gamma}'\hat{\mathbf{z}} - v)\right] I(\hat{\mathbf{z}} \notin \Theta_0) \\
= \; & \left[\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\mathbf{z}} + \sqrt{n}\,(\boldsymbol{\gamma}'\hat{\mathbf{z}} - v)\right] X_n + \sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\mathbf{z}} I(\hat{\mathbf{z}} \notin \Theta_0) + \sqrt{n}\,(\boldsymbol{\gamma}'\hat{\mathbf{z}} - v)\, I(\hat{\mathbf{z}} \notin \Theta_0)
\end{aligned}$$

where $X_n = I(\hat{\mathbf{z}} \in \Theta_0) \to 1$ in first mean and therefore in probability. Now, consider the second term on the RHS of the last display. Since $\hat{\mathbf{z}} \in \mathcal{P}$, compact and $\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma}) = O_p(1)$, we have that the second term is $o_p(1)$. Next, consider the third term. Observe that $(\boldsymbol{\gamma}'\hat{\mathbf{z}} - v)$ is bounded by assumption and use corollary 1 with $a_n = \sqrt{n}$ to conclude that the 3rd term converges to 0 in probability. This establishes (4).

Next, recall that $\sqrt{n}\,(v - \boldsymbol{\gamma}'\mathbf{z}_j) \to 0$ for $1 \le j \le J$. Then ignoring $o_p(1)$ terms and denoting the asymptotic limit for $\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})$ by w, one gets

$$\begin{aligned}
& \sqrt{n}\,(\hat{v} - v) \\
= \; & \left[\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \hat{\mathbf{z}} - \sqrt{n}\,(v - \boldsymbol{\gamma}'\hat{\mathbf{z}})\right] I(\hat{\mathbf{z}} \in \Theta_0) \\
= \; & \sum_{j=1}^{J} \left[\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \mathbf{z}_j\right] I\left(\hat{\boldsymbol{\gamma}}'(\mathbf{z}_j - \mathbf{z}_1) > 0 ..., \hat{\boldsymbol{\gamma}}'(\mathbf{z}_j - \mathbf{z}_J) > 0\right) \\
= \; & \sum_{j=1}^{J} \left(\begin{array}{c} \left[\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})' \mathbf{z}_j\right] \\ \times I\left(\sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\mathbf{z}_j - \mathbf{z}_1) > \sqrt{n}\boldsymbol{\gamma}'(\mathbf{z}_1 - \mathbf{z}_j), ..., \sqrt{n}\,(\hat{\boldsymbol{\gamma}} - \boldsymbol{\gamma})'(\mathbf{z}_j - \mathbf{z}_J) > \sqrt{n}\boldsymbol{\gamma}'(\mathbf{z}_J - \mathbf{z}_j)\right) \end{array} \right).
\end{aligned}$$

Ignoring $o_p(1)$ terms, one gets

$$\sqrt{n}\left(\hat{v}-v\right)=\sum_{j=1}^{J}\left[w'\mathbf{z}_j\right]\times I\left\{\cap_{l\neq j,l=1,\ldots J}\left\{w'\left(\mathbf{z}_j-\mathbf{z}_l\right)>0\right\}\right\}$$

$$=\sum_{j=1}^{J}\left[w'\mathbf{z}_j\right]\times I\left\{\cap_{l\neq j,l=1,\ldots J}\left\{w'\mathbf{z}_j>w'\mathbf{z}_l\right\}\right\}=\max_{1\leq j\leq J}\left\{w'\mathbf{z}_j\right\}.$$

■

Proof of Proposition 4. If $\max_{\mathbf{z}\in\hat{\Theta}_n}\left\{\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}\right\}\neq\max_{\mathbf{z}\in\Theta_0}\left\{\sqrt{n}\left(\hat{\boldsymbol{\gamma}}-\boldsymbol{\gamma}\right)'\mathbf{z}\right\}$ for any finite n, there will exist either a $\mathbf{z}\in\hat{\Theta}_n$ with $\mathbf{z}\notin\Theta_0$ or a $\mathbf{z}\notin\hat{\Theta}_n$ with $\mathbf{z}\in\Theta_0$ and the probability of both these events tend to 0 by proposition 1. ■

Proof of Proposition 5. Denote the medians corresponding to allocations $\mathbf{p},\mathbf{q},\mathbf{r}$ by $\mu_{\mathbf{p}},\mu_{\mathbf{q}}$ and $\mu_{\mathbf{r}}$, respectively and WLOG assume that $\mu_{\mathbf{p}}>\mu_{\mathbf{q}}$.

First, suppose $\mu_{\mathbf{r}}>\mu_{\mathbf{p}}$. Then

$$\begin{aligned}\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{p}}\right) &= \lambda\sum_{k=1}^{m}p_kF_k\left(\mu_{\mathbf{p}}\right)+\left(1-\lambda\right)\sum_{k=1}^{m}q_kF_k\left(\mu_{\mathbf{p}}\right)\\ &\geq \lambda\sum_{k=1}^{m}p_kF_k\left(\mu_{\mathbf{p}}\right)+\left(1-\lambda\right)\sum_{k=1}^{m}q_kF_k\left(\mu_{\mathbf{q}}\right)\\ &= 0.5=\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{r}}\right).\end{aligned}$$

If the weak inequality is an equality, then $\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{p}}\right)=0.5=\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{r}}\right)$, with $\mu_{\mathbf{r}}>\mu_{\mathbf{p}}$ contradicting the definition of $\mu_{\mathbf{r}}$. So we must have that

$$\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{p}}\right)>\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{r}}\right). \quad (9)$$

Now, since $\mu_r>\mu_{\mathbf{p}}$, one has that for each k, $F_k\left(\mu_{\mathbf{r}}\right)\geq F_k\left(\mu_{\mathbf{p}}\right)$ implying $\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{p}}\right)\leq\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{r}}\right)$- which contradicts (9). This implies that $\mu_{\mathbf{r}}\leq\mu_{\mathbf{p}}$.

Next,

$$\begin{aligned}\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{q}}\right) &= \lambda\sum_{k=1}^{m}p_kF_k\left(\mu_{\mathbf{q}}\right)+\left(1-\lambda\right)\sum_{k=1}^{m}q_kF_k\left(\mu_{\mathbf{q}}\right)\\ &= \lambda\times\sum_{k=1}^{m}p_kF_k\left(\mu_{\mathbf{q}}\right)+\left(1-\lambda\right)\times 0.5\\ &< \lambda\times 0.5+\left(1-\lambda\right)\times 0.5\\ &= 0.5=\sum_{k=1}^{m}r_kF_k\left(\mu_{\mathbf{r}}\right),\end{aligned}$$

implying that $\mu_{\mathbf{q}} < \mu_{\mathbf{r}}$ since each $F_k(\cdot)$ is non-decreasing. The inequality follows from the definition of the medians. Indeed, if $\sum_{k=1}^{m} p_k F_k\left(\mu_{\mathbf{q}}\right) = 0.5$ this contradicts that $\mu_{\mathbf{p}} > \mu_{\mathbf{q}}$ according to the definition of $\mu_{\mathbf{p}}$ and the condition that $\mu_{\mathbf{p}} > \mu_{\mathbf{q}}$. Thus we must have $\mu_{\mathbf{p}} \geq \mu_{\mathbf{r}} > \mu_{\mathbf{q}}$. ■

8.2 A caveat to the application

A complication in the Dartmouth assignment process is that room assignment was random, conditional on three self-reported binary covariates– smoking habit, cleanliness and music playing– collectively called Z– and gender (of these, only gender is used as a conditioning covariate in the paper). For instance, no smoker was partnered with a non-smoker. An ideal analysis therefore would proceed by conditioning on both gender and Z, which leads to severe small cell-problems for most values of Z.

Ignoring Z can still lead to optimal allocations if different combinations of Z between the 2 roommates do not lead to different effects of a given combination of conditioning covariates on the aggregate outcome for the room, i.e., if a separability condition holds. To see this, consider for simplicity the variable smoking and let S denote a smoker, N denote a non-smoker and a room denoted by SN (resp., SS and NN) if it has one smoker and one nonsmoker. Let Y denote outcome for the room. Further let $E(\cdot)$ denote expectation w.r.t. the distribution induced by unconstrained allocation and $E^*(\cdot)$ the one due to constrained allocation which does not allow SN type rooms. Suppose one conditions the allocation on a binary variable taking values H and L. Then

$$\begin{aligned} E(Y|HH) &= E(Y|HH,SS)\Pr(SS|HH) + E(Y|HH,SN)\Pr(SN|HH) \\ &\quad + E(Y|HH,NN)\Pr(NN|HH) \\ &\neq E^*(Y|HH,SS)\Pr{}^*(SS|HH) + E^*(Y|HH,NN)\Pr{}^*(NN|HH) \\ &= E^*(Y|HH). \end{aligned}$$

If in the experimental data, there is no SN type room, then $E(Y|HH,SN)$ and therefore $E(Y|HH)$ cannot be consistently estimated. Now assume that for $a = HH, HL, LL$ and $b = SS, SN, NN$ we have

$$E(Y|a,b) = \gamma_1(a) + \gamma_2(b).$$

Then for allocation $\mathbf{p} = (p_{HH}, p_{HL}, p_{LL})$, the unconstrained expected value of the objective function is

$$\begin{aligned}
& p_{HH}E\left(Y|HH\right)+p_{HL}E\left(Y|HL\right)+p_{LL}E\left(Y|LL\right) \\
= \; & p_{HH}\gamma_1\left(HH\right)+p_{HL}\gamma_1\left(HL\right)+p_{LL}\gamma_1\left(LL\right) \\
& +\left[\gamma_2\left(SS\right)\Pr\left(SS\right)+\gamma_2\left(SN\right)\Pr\left(SN\right)+\gamma_2\left(NN\right)\Pr\left(NN\right)\right] \qquad (10)
\end{aligned}$$

and the term within [.] does not depend on $\mathbf{p}$ since the probabilities $\Pr\left(SN\right), \Pr\left(SS\right)$ and $\Pr\left(NN\right)$ depend only on the population proportion of smokers. Similarly, the value of the constrained maximum will be

$$\begin{aligned}
& p_{HH}E^*\left(Y|HH\right)+p_{HL}E^*\left(Y|HL\right)+p_{LL}E^*\left(Y|LL\right) \\
= \; & p_{HH}\gamma_1\left(HH\right)+p_{HL}\gamma_1\left(HL\right)+p_{LL}\gamma_1\left(LL\right) \\
& +\left[\gamma_2\left(SS\right)\Pr{}^*\left(SS\right)+\gamma_2\left(NN\right)\Pr{}^*\left(NN\right)\right] . \qquad (11)
\end{aligned}$$

Comparing (10) and (11), one can see that the two objective functions differ by a term which does not depend on the allocation, implying that we can consistently estimate the solution.

Under nonseparability, this argument is no longer valid. To address this problem in the application, the analysis is first performed by taking gender into account but ignoring Z and then condition on both gender and Z. However, in all cases where Z was conditioned on, degeneracy could not be rejected and those results are not reported here. Since smokers constitute about 1% of the total sample, when performing the analysis ignoring the effect of Z, all rooms (a total of 3 rooms out of about 400) which contained smokers were dropped. Therefore the above analysis essentially applies to non-smokers and ignores the potentially confounding effects of cleanliness and music-listening.

When the outcome is joining a fraternity and allocation is conditioned on ACA or race, it seems safe to assume that different combinations of Z do not have different effects on the outcome, for a given combination of the conditining covariates. So the analysis which ignores Z will likely give us consistent estimates of both the optimal allocation and value. But when the outcome is GPA and the allocation is conditioned on ACA or race, it is possible that some combinations of Z will have large negative effects on the outcome, e.g., cohabitation of an untidy music-listener and a tidy music-hater could severely reduce total GPA of the room, for any given combinations of ACA. This will lead to inconsistent estimates of optimal allocations and values unless separability holds. The results presented above should be interpreted with this caveat in mind.