Inferring Optimal Resource Allocations from Experimental Data

Debopam Bhattacharya,[*][†]

Dartmouth College.

First draft: February 15, 2006; this draft: November 2, 2006.
Comments Welcome

Abstract

This paper concerns the problem of optimally allocating a scarce indivisible resource among a target population based on experimental data for a sample drawn from this population. For a wide class of social welfare functions, the problem can be set up as a mathematical program with estimated components. The paper provides a framework for statistical inference on the estimated optimal solutions and corresponding optimal values. The analysis carefully distinguishes between situations of uniqueness, nonuniqueness and near non-uniqueness of the solutions and shows that in the nearly non-unique case, subsampling gives inconsistent inference. Yet, confidence intervals which are potentially conservative but robust to the degree of uniqueness can be constructed. A key contribution of this paper is to show that applicability of these techniques extends beyond linear maximands like the mean

to other important policy objectives like outcome quantiles which are *nonlinear* in the allocation probabilities. This work differs from the recent "treatment choice" literature as it allows for aggregate resource constraints and focuses on the asymptotic analysis of estimated optimal solutions and values for a fixed decision rule. The methods are illustrated using data from Dartmouth's random assignment of freshmen to dorm rooms, where Sacerdote (2001) detected significant contextual peer effects. Segregation by previous academic achievement and by race are seen to minimize mean enrolment into sororities and maximize mean enrolment into fraternities. Segregation appears to have no effect on mean and median freshman year GPA but increases the higher and decreases the lower percentiles of the GPA distribution.

1   Introduction

Design of optimal policies is an important goal of economic analysis but its implementation in reality is a complicated endeavor. Firstly, the welfare criterion which the optimal policy would seek to maximize is often subjective and depends on political economy considerations. Even after deciding on a welfare criterion, choosing a policy to maximize it can be difficult because it requires knowledge of how individuals will behave under counterfactual situations. As is well-known, observational data cannot be used in such circumstances since they arise out of nonrandom self-selection by individuals, leading to endogeneity bias. The endogeneity problem can be largely overcome when one has data on outcomes arising from random assignment of individuals to situations. Even when data from random assignment are available, the policymaker would still need to solve an optimization problem under constraints imposed by resource limitations. The objective function in this optimization problem will typically involve components which are estimated by observing the behavior of a sample of individuals (who were randomized into treatment) drawn from the target population. In some cases, the constraints may also involve estimated components if some feature of the population distribution is to be preserved by the optimal allocation. Because of sampling variability, the resulting solutions and optimal values would warrant a statistical distribution theory on which to base inference about the optimal policy.

1.1 Resource Allocation Problems

The present paper considers a class of resource allocation problems aimed at maximizing an overall outcome for a target population. It is assumed that the policymaker can access outcome data from a randomized experiment for a random sample drawn from the same population. The set-up is as follows. The target population contains countably finite types of subjects and each policy is defined by the (finite dimensional) vector specifying the share of resource accruing to each type. Assume that there are m types of subjects and an allocation $\mathbf{p}$ is an m-vector. If $F_j(.)$ denotes the cdf of outcome for the jth type, then the cdf of aggregate outcome induced by the allocation $\mathbf{p}$ is $F_{\mathbf{p}} = \sum_{j=1}^{m} p_j F_j(.)$. Let the set of feasible allocations be denoted by $\mathcal{P}$ and the social objective functional to be maximized by $\nu(.)$. The problem of choosing the optimal policy is then given by

$$\max_{\mathbf{p} \in \mathcal{P}} \nu(F_{\mathbf{p}}) \tag{1}$$

The set $\mathcal{P}$ represents the constraints on the allocation probabilities which will include among others that $\mathbf{p} \geq 0$ and $\mathbf{p}'1 = 1$. In this paper, I will restrict attention to the situation where $\mathcal{P}$ is known exactly to the policymaker, as will be the case in many problems of interest.[1] I will also assume that $\mathcal{P}$ is compact and convex (as they are in many, if not most, real-life problems (see examples below) and the functional $\nu(.)$ is continuous. So there will be a unique maximum value of the problem (1) but generally more than one $\mathbf{p}$ can solve the problem.

Clearly, problem (1) is interesting only when different allocations lead to different *aggregate* outcomes, i.e. when $\nu(F_{\mathbf{p}})$ varies with $\mathbf{p}$ at least for $\mathbf{p}$ in some subset of $\mathcal{P}$. When $\nu(.)$ is the mean, for instance, this variation will happen if the attributes defining the type have some degree of complementarity in producing the outcome. In general, one would expect the aggregate outcome to vary with $\mathbf{p}$ if the F_j's are sufficiently different across j.

Social Objectives: In this paper, I will consider two types of objective functionals $\nu(.)$ for the policymaker- (i) mean of the outcome and (ii) a quantile of the outcome distribution. Maximization of mean amounts to maximizing productive efficiency. Maximization of central quantiles such as the median is motivated by the well-known fact that means are often heavily affected by outliers. So policies which maximize the mean can do so by increasing the outcome for a very small number of subjects while the majority has

---

[1] The present author is considering the case of estimated constraints in a different paper,

poor outcomes. Maximizing the median guarantees that at least half the population has a good outcome and median maximization can therefore be viewed as a more democratic criterion. Maximization of the lower quantiles can be viewed as equity enhancing policies and its implementation is analytically very similar to median maximization. The outcome in many of these problems (such as health, test-score etc.) are hard to redistribute after they have been realized which implies that all distributional goals can be targeted only at the allocation stage.

Parameters of interest: Once the social objective function is fixed, one would like to know the maximal allocation- i.e. the shares of the various types which will maximize the objective function. But on many occasions, the maximum value is an object of independent and significant policy interest. This is especially when all technologically feasible allocations are not practically feasible due to political reasons. For example, allocations based on race or gender could be viewed as illegal and therefore banned. Comparing the maximum value corresponding to a politically unconstrained problem with that of the constrained one will then give an estimate of the "efficiency costs" of the political constraint. Furthermore, this "unconstrained" maximum value can serve as a benchmark against which specific allocation rules can be compared.

The Sample Problem: Typically, the F_j's are not observed by the policy-maker. This paper considers scenarios where the F_j's can be estimated consistently based on a sample drawn from the target population, i.e some estimates $\hat{F}_j(.)$ are observed and the policymaker solves the problem

$$\max_{\mathbf{p} \in \mathcal{P}} \nu\left(\hat{F}_{\mathbf{p}}\right).$$

The primary objective of this paper is to characterize the statistical properties of the solution and optimal values of this "sample" maximization problem. As noted in the first paragraph, consistent estimates of the F_j's can be obtained if one has access to outcome data for a sample drawn from the same population who were randomly assigned the attributes (which define the type). If a random sample of n individuals were randomly assigned to m types or groups labeled as G_j, $j = 1, ...m$ and y_i denotes the outcome of the ith individual, then a consistent estimate of F_j is given by

$$\hat{F}_j(a) = \frac{\frac{1}{n}\sum_{i=1}^{n} I\left(i \in G_j\right) I\left(y_i \leq a\right)}{\frac{1}{n}\sum_{i=j}^{n} I\left(i \in G_j\right)} \tag{2}$$

where $I\left(i \in G_j\right)$ is the indicator for individual i belonging to type G_j.

4

1.2 Examples

Example 1 (resource allocation under fixed marginals): Suppose that a college authority wants to improve average (and thus total) 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 into an effect of own background plus an effect of roommate's background. Otherwise, all assignments should yield the same total, and thus average, outcome.

Assume that every dorm room can accommodate two students and the college can assign individuals to dorms based on their race. Assume that race is classified as white, black, others- abbreviated by w, b, o. Denote the expected total score of a dorm room with each of 6 types of couples, denoted by $\mathbf{g} = (g_{ww}, g_{bb}, g_{oo}, g_{wo}, g_{wb}, g_{bo})'$. For instance, g_{bo} is the mean *per person* GPA score across all rooms which have one b-type and one o-type student, i.e. *if one b type and one o type student were randomly picked from the population and assigned to be roommates, then the expected value of the sum of their GPA's is g_{bo}.* Denote the marginal distribution of race for the current class by $\boldsymbol{\pi} = \pi_w, \pi_b, \pi_o$. Then an allocation is a vector $\mathbf{p} = (p_{ww}, p_{bb}, p_{oo}, p_{wo}, p_{wb}, p_{bo})'$, satisfying $p_{ij} \geq 0$ and $\mathbf{p'1} = 1$. Here p_{ij} (which equals p_{ji} by definition) 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 problem.

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

s.t.

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

The first set of linear constraints are the restrictions imposed by the fixed marginal distributions. For example, the first linear constraint simply says that the total number of

students of w type in the dorm rooms (in every ww type room there are two w type students and hence the multiplier 2 appear before p_{ww}) should equal the total number of w type students that year. The first of these quantities is $N/2 \times (2p_{ww} + p_{wo} + p_{wb})$ if there are N students and hence $N/2$ dorm rooms. The second is $N \times \pi_w$. Thus one can view the g's as the preliminary parameters of interest and the solution to the LP problem and the resulting maximum value as functions of g's which constitute the ultimate parameters of interest. Note that the solution (the p_{ij}'s that solve the problem) may not always be unique but the maximum value is, provided the g's are bounded.

In general $\mathbf{g}$ will be unknown and so the above problem is infeasible. Now suppose, a sample was drawn from the same population from which the incoming freshmen are drawn (e.g. the freshmen class in the previous year). Further assume that this "pilot" sample was randomly grouped into rooms and the planner has access to freshman year GPA data for each member of this sample. Then the planner can calculate mean total score for this sample across dorm rooms, say, with one w and one o type to estimate $\hat{g}_{wo}$ which will be a good estimate of the unknown g_{wo} if the sample size is large.

Replacing unknown g's by the sample counterparts, the planner now solves

$$\max_{\mathbf{p}} \hat{\mathbf{g}}'\mathbf{p} \text{ s.t. } A\mathbf{p} = \boldsymbol{\pi}, \ \mathbf{p} \geq \mathbf{0} \tag{3}$$

where

$$\hat{\mathbf{g}} = (\hat{g}_{ww}, \hat{g}_{bb}, \hat{g}_{oo}, \hat{g}_{wo}, \hat{g}_{wb}, \hat{g}_{bo})'$$

$$\mathbf{p} = (p_{ww}, p_{bb}, p_{oo}, p_{wo}, p_{wb}, p_{bo})'$$

$$A = \begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 & 0 & 1 \end{pmatrix}$$

$$\boldsymbol{\pi} = (\pi_w, \pi_b, \pi_o)'.$$

Example 2 (resource allocation under budget constraint): Suppose a policymaker is trying to reduce smoking in a target population. Suppose she is considering two policies- distributing nicotine patches free of charge and (fully) subsidizing counselling sessions. Assume that the target group contains N smokers, each nicotine patch treatment costs the policymaker w_1, each counselling program costs w_2 and the policymaker's budget is W. An allocation here is a 4-dimensional allocation vector $\mathbf{p} = (p_{tT}, p_{tC}, p_{cT}, p_{cC})$ where T (t) and C (c) respectively refer to getting and not getting a patch (counselling) and p_{tC}

is the proportion of smokers who would be counselled but will not get a patch; similarly for p_{tT}, p_{cT}, p_{cC}. The feasible set of allocations $\mathcal{P}$ is defined as the set of $\mathbf{p}$ with $\mathbf{p}'1 = 1$, $\mathbf{p} \geq \mathbf{0}$ and the policymaker's budget constraint

$$w_1 \left(p_{tT} + p_{cT}\right) + w_2 \left(p_{tT} + p_{tC}\right) = W/N.$$

Let the vector of cdf for smoking reduction be denoted by $F\left(.\right) = \left(F_{tT}\left(.\right), F_{tC}\left(.\right), F_{cT}\left(.\right), F_{cC}\left(.\right)\right)$ where $F_{cT}\left(a\right)$ has the interpretation: if an individual was randomly picked from the population and assigned a patch but no counselling, then $F_{cT}\left(a\right)$ is the probability that the number of cigarettes smoked by him fell by a or less. If the objective is to maximize median reduction, then the population problem is given by

$$\max_{p \in \mathcal{P}} \mu_p$$

where μ_p is defined by

$$0.5 = p_{tT} F_{tT}\left(\mu_p\right) + p_{tC} F_{tC}\left(\mu_p\right) + p_{cT} F_{cT}\left(\mu_p\right) + p_{cC} F_{cC}\left(\mu_p\right).^2$$

One could consider a richer problem of conditioning allocations on a set of observed (discrete) covariates- X (such as age, smoking intensity, gender etc.). If X takes values 1,...M, with $\pi_j = \Pr\left(X = j\right)$, then an allocation will be the vector $\{\mathbf{p}_j\}_{j=1,...M}$ where $\mathbf{p}_j = \left(p_{tTj}, p_{tCj}, p_{cTj}, p_{cCj}\right)$, the budget constraint is

$$\sum_{j=1}^{M} \pi_j \left\{w_1 \left(p_{tTj} + p_{cTj}\right) + w_2 \left(p_{tTj} + p_{tCj}\right)\right\} = W/N.$$

The distribution of smoking reduction will now depend both on the treatment and the covariates X and thus the objective function μ_p will be given by

$$0.5 = \sum_{j=1}^{M} \pi_j \left[p_{tT} F_{tT}\left(j, \mu_p\right) + p_{tC} F_{tC}\left(j, \mu_p\right) + p_{cT} F_{cT}\left(j, \mu_p\right) + p_{cC} F_{cC}\left(j, \mu_p\right)\right],$$

---

[2] In this simple example, it should be obvious that in an optimal allocation rule, either $p_{Tt} = 0 = p_{Cc}$ or $p_{Tc} = p_{tC} = 0$. When more policy instruments are added, e.g. use of nicotine gums, such intuitive reasoning is not straightforward and many cases need to be considered which will make the problem intractable. The programming approach outlined below provides an elegant and effective method for such settings especially when the number of possible allocations is large.

where $F_{cT}(j, a)$ has the interpretation: if an individual was randomly picked from the sub-population with X=j and assigned a patch but no counselling, then $F_{cT}(j, a)$ is the probability that the number of cigarettes smoked by him falls by a or less.

Typically the policymaker will not know the distributions of reductions, i.e. $F(.)$. If she decides to infer them from observational data, she will likely get biased estimates, since individuals who have a stronger desire to quit are more likely to both buy patches and take counselling sessions. If however, she has access to outcome data from a randomized experiment where patches and/or counselling were randomly assigned (made available) to smokers, she can obtain unbiased estimates of $F(.)$ and base her analysis on them. The experiment has to be designed so that the subjects are allowed to exert behavioral responses (e.g. compliance) during the experiment as they would in a nonexperimental setting. The outcome, therefore, will be net of those behavioral responses.

1.3 Discussions

In this subsection, I discuss three separate issues which qualify the validity of my approach.

First, the analysis in this paper assumes that the model for outcome as a function of the allocation is unaffected by whether allocations are made through general randomization (as with the pilot sample) or by randomization within types (as will be done by the planner). This assumption can fail to hold if, in example 1 for instance, students are more antagonistic to roommates who are different from them if they know that this allocation was, at least partly, a result of conscious choice of the planner.[3]

Second, observe that I use estimates 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. In example 1, mean GPA in the OO subgroup of the pilot sample (e.g. year 1 entrants) is given by $E_P(Y|OO)$ where $E_P(.)$ denotes the distribution in the pilot sample. Clearly, interaction between the two roommates' other covariates such as academic aptitude (e.g. SAT score), noise tolerance etc. (jointly called X) also affect mean GPA, so that

$$E_P(Y|OO) = \int E_P(Y|OO, X) \, dF_P(X|OO)$$

---

[3]In example 1, it is also assumed that nature of interactions are unaffected by the aggregate proportions of H, M and L types, which can fail to hold if for instance, high types are more accommodating of low types if low types are a small minority but more antagonistic if high types are a minority. I am grateful to Prof David Green of UBC for pointing out the second caveat.

where $F_P(X|OO)$ denotes the distribution of X for the OO individuals in year 1. It is reasonable to assume that

$$E_P(Y|OO, X) = E_{Tar}(Y|OO, X)$$

where E_{Tar} denotes expectation w.r.t. the distributions in the targeted subgroup (e.g. year 2 entrants). However, unless the distribution of covariates $F_{Tar}(X|OO) = F_P(X|OO)$, in general we will not have that $E_P(Y|OO) = E_{Tar}(Y|OO)$. This seems to suggest that one should use as many covariates as possible to define the types to make the corresponding estimates more robust to heterogeneity between the target and the pilot. However, as one increases the number of covariates on which to match, the data requirement becomes more severe. Suppose one were to match students in example 1 on both race and whether they were athletes in high school. Now there will be 6 types of individuals and so 21 types of rooms. If one adds to this mothers' educational status (e.g. high school, college, graduate degree), there will be 171 types of rooms and so on. This is will likely lead to the familiar "small-cell" problem, e.g. there will be very few rooms, perhaps none, with two black non-athletes both with HS type mothers. This will make it hard, if not impossible, to calculate optimal allocations. This trade-off is no different in spirit from the issue of how many and which covariates (and their interactions) to include in an usual regression equation.

A third point raised by previous readers of this paper is that it might be preferable to write down an explicit model of outcomes as a function of the matches, estimate it using data from randomization and then perform the optimization by using those estimates as inputs. When covariates are discrete and all interactions are allowed for, this approach is no different from the approach in the present paper, since the coefficients on the covariates and their interactions in the alternative model-based approach are one-to-one transformations of subgroup level means used in the present approach. Certainly, the "model-based" approach can accommodate prior information about the nature of interactions into the analysis, e.g. that some interactions have no or little effect. Such restrictions can potentially improve the precision of the remaining estimates and consequently improve the precision of the optimal allocation estimates. However, there are at least three potential problems with this approach. First, prior information of this kind is typically case-specific and imposing them on the problem at hand can be viewed as ad-hoc and overtly restrictive. Of course, one could test such assumptions based on the pilot survey and then estimate a frugal model if one concludes the absence of interactions.

9

But then the first step estimation error will affect the precision of the second round estimates. So there is unlikely to be any ultimate gain in precision. Secondly, a model based approach will not solve the small-cell (or empty-cell) problem because small cells imply that coefficients on the corresponding interaction will be estimated imprecisely.[4] Third, unlike the mean maximization case, a model-based approach for quantile maximization is difficult in this setting. If one writes a model for the conditional quantile of the outcome as a function of the allocations (as in usual quantile regressions), one cannot in general express the maximand- i.e. the overall population quantile in terms of the parameters of this conditional quantile model, e.g. $med\left(y\right) \neq \int med\left(y|x\right) dF\left(x\right)$. It is not clear how one may use a "model-based" analysis for maximizing a quantile.

When covariates are continuous, a frugal (perhaps parametric) model rather than a completely unrestricted one can improve precision (of mean-maximizing allocations) significantly. However, the main difficulty in the continuous covariate case is that 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 determining the outcome.

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. But, in consequence, the methods proposed here are relevant only when one is confident that the distribution of omitted covariates, conditional on the observed ones, are identical in the target and pilot surveys- which will hold, in particular, when the two surveys are drawn from the same population.

1.4   Plan of the paper

Section 2 summarizes the contributions of the paper in relation to the existing literature. Section 3 and 4 focus on the relatively easier case of mean maximization. Section 3 discusses the issue of uniqueness of optimal solution(s) while section 4 discusses consistency and asymptotic distribution theory for both the maximal allocation and the maximum value in both the point and set-identified case and proposes a test of uniqueness. Section 5 considers the analytically more interesting case of maximizing a quantile and considers identification and inference. Section 6 discusses the application to a real dataset con-

---

[4]There is no reason to believe that interactions whose effects are hard to estimate are unimportant ones and can be left out of the equation.

cerning room allocation of Dartmouth College undergraduates and section 7 concludes. Longer proofs are collected in the appendix.

2 Related Literature and Contributions

The main contribution of the present paper is to use insights from mathematical programming theory to analyze identification and asymptotic inference in constrained resource allocation problems. It is shown here that when the objective of resource allocation is to maximize mean outcome, the optimal allocation problem reduces to a standard linear program (LP) with an estimated objective function. Consequently, the fundamental theorem of LP reduces the parameter space to the countably finite collection of extreme points of the constraint set. I show that the set of estimated solutions converges to the set of true solutions arbitrarily fast so that sampling error in estimating the optimal solution(s) has no effect asymptotically on the distribution of the estimated maximum value. This latter distribution depends qualitatively on whether the solution set is unique, non-unique or nearly non-unique. Since it is typically unknown as to which of these is the case, an exact asymptotic confidence interval (CI) for the maximum value cannot be constructed here. I further show that subsampling based confidence intervals are generally *inconsistent* in the problematic case of near-uniqueness. Two types of one-sided CI for the maximum value are constructed which are robust to the degree of nonuniqueness of the solution. One type is computationally simpler but longer while the other is computationally more involved but shorter in length.

A further contribution of this paper is to show that this programming theory based analysis can be extended to the problem of maximizing any quantile of the resulting outcome distribution even though the quantile objective function is nonlinear in the allocation probabilities. Quantile maximization in itself is a nonstandard exercise in that it cannot be interpreted as an M (or Z)-estimation problem, unlike e.g. quantile regressions. To the best of my knowledge, analysis of identification and estimation in quantile maximization problems has not been attempted before[5] and the present paper addresses these issues in the context of resource allocation, where quantile maximization is potentially an important policy objective. The paper also describes what are the general class of policy

---

[5]Manski (1988) and Rostek (2005) have previously considered decision theoretic interpretations of quantile maximization but they do not address identification or inference from a statistical perspective.

11

maximands to which these ideas can be applied.

The work presented here was originally inspired by input-matching problems first discussed in Graham, Imbens and Ridder (2005) (henceforth, GIR). They considered a setting like example 1 where two inputs with fixed marginal distributions were available and were potentially complementary in the production of an output. GIR compared a fixed set of matching rules between them (like positive and negative assortative matching) to see which one yields the higher expected pay-off. In that paper, they do not consider optimal allocations. Having a method for calculating the optima, as outlined in the present paper, also enables one to compute the efficiency of a given allocation rule by comparing it to the maximum attainable. In contrast to GIR (2005), however, this paper is concerned exclusively with discrete types,[6] which enables one to use results from the finite-dimensional programming literature.

In independent work, GIR (2006) consider mean-based optimality from a decision theoretic standpoint when covariates are binary. In contrast, the present paper focuses on one specific albeit natural plug-in type decision rule for a general discrete valued covariate and, perhaps more importantly, considers the maximization of both means and quantiles. GIR take identification of the population optima as given while the present paper investigates point and set-valued identification and conducts inference which is robust to the degree of uniqueness of the population solution. In fact, when the inputs are binary as in GIR (2006), it can be shown that either the population solution is unique or else all feasible allocations are optimal. For general discrete valued covariates, as discussed in the present paper, one can have nontrivial set-identified situations. This is further elaborated below in section 3. The method of analysis in the present paper is also fundamentally different from GIR (2006) in that it uses a unified programming theory based approach to identification and estimation of the optimal allocation. But unlike GIR (2006), this paper does not investigate decision theoretic issues associated with the problem. Thus the two papers are complementary to each other in terms of both objectives and methodology.

The work presented here is also related to the more recent treatment choice literature, c.f. Manski (2004), Dahejia (2003) and Hirano and Porter (2005)- in that it concerns de-

---

[6]The nonstochastic version of mean based optimization in the continuous case can be related to the Monge-Kantorovich mass transportation problem. The author of the present paper is developing a continuous analog of matching problems.

signing optimal policies based on results of a random experiment. But this paper differs substantively from the above papers in at least three ways. First, it analyzes a *constrained* decision-making problem which makes the analysis applicable to a different set of situations where a large number of individuals have to be allocated simultaneously and not everyone can be assigned what is the "first best" for them, unlike the treatment choice situations analyzed in the above papers. Presence of the constraints also makes the problem analytically different. Secondly, this paper is concerned with asymptotic properties of *a specific* but natural sample-based decision rule for a given choice of covariates rather than a finite sample-based analysis of finding the optimal rule.[7] And thirdly, this paper develops and analyzes the problem of quantile maximization which is not considered in the treatment choice literature at all. Since outcomes like survival after a surgery or test-scores after an intervention cannot be redistributed among agents, every distributional goal has to be met during the process that generates the outcome. This reinforces the importance of quantile maximization as a policy objective in such situations.

Optimization under uncertainty has been a recurrent theme of research in macroeconomics, operations research and applied mathematics. The operations research literature has mainly focussed on algorithms for actually solving stochastic optimization problems. Birge (1998) provides a textbook treatment of this line of research. The classical optimal control literature in economics has considered the role of uncertainty in representative agent optimization problems and discussed the implications of such decision-making for the movement of macroeconomic variables. Stokey and Lucas (1994) provides a textbook treatment of this approach. Neither literature, to my knowledge, is concerned with statistical inference on the estimated solution and value functions.

There is a moderately sized literature in statistics and econometrics (henceforth SOL) which focuses on the statistical properties of solutions and values of certain stochastic optimization problems. Some of this was developed in the context of testing inequality constraints (c.f. Bartholomew (1957), Wolak (1987) and references therein) while some other works deal with such problems in somewhat greater generality (c.f. Shapiro (1989) and references therein). The present paper differs from that literature in three fundamental ways. First, the objective functions, i.e. maximands/minimands considered in that literature are simple population averages and include neither ratios of averages

---

[7]Manski (2004) also restricts attention to "conditional empirical success" (CES) rules but considers the problem of choosing the set of covariates on which CES is conditioned as the decision problem.

(c.f. equation (2) above) nor more complicated functionals like quantiles which cannot be expressed as population means.[8] Second, SOL does not carefully consider situations of nonuniqueness versus *near*-nonuniqueness as this paper will do. That way, the present paper will contribute to the growing econometrics literature on set-valued identification and inference, pioneered by Manski (1995). Near-nonuniqueness refers to the situation where at some points in the feasible set, the value of the objective function differs from the optimal value by a term of order $O\left(1/\sqrt{n}\right)$ (see section 3 for more details). Near-nonuniqueness, which is more likely than the somewhat nongeneric situation of strict nonuniqueness, is shown to complicate inference. It invalidates asymptotic approximations which assume that the solution is either unique or (strictly) nonunique and also renders subsampling-based inference inconsistent. The approach developed here shows how to make asymptotic inference robust to whether the solution is unique, nonunique and nearly non-unique. Third, the effective parameter space here is discrete which makes the magnitude of the difference between the estimated and the true solution(s) arbitrarily small in probability, asymptotically.[9] This result seems nonstandard and very different from some exponential rates derived in the SOL (c.f. Shapiro and de Mello (2000)). The SOL results show that the probability of the estimated solution differing from the true solution(s) converges to zero exponentially fast, using large-deviations results which require thin tail assumptions (or bounded supports) for the outcome. The arbitrarily fast rate derived here concerns the (probabilistic) order of magnitude of the difference between the estimated solution and the true solutions and apply under bounded second moments alone. It turns out that the latter results are sufficient to establish a distribution theory for the estimated maximum value.

For the sake of completeness, I derive a large deviations result for the median case below. This derivation is also very different from the large deviations results of the SOL not only because medians are not averages (which is what SOL is really about) but also

---

[8] Note that quantile maximization is a very different problem from quantile regression, the latter having an objective function which is a population mean.

[9] The reason why such discreteness does not happen in e.g. Wolak (1987) is as follows. Wolak (1987) considers minimizing a strictly convex objective function on a compact and convex parameter space P. This is the same as maximizing a strictly **concave** objective function on P. The maximization problem in the present paper maximizes a quasi-**convex** functional on P; note that means are trivially quasi-convex, that quantiles are also quasi-convex will be shown in section 5. The fact that the maxima can occur only at the extreme points of P is a consequence of the quasi-convexity. **Concave** functionals (e.g. negative of the OLS criterion function as in Wolak (1987)) are not quasi-convex in general.

because the estimates here will involve random denominators. Further, it is shown that in the case of the median, one does not need to restrict the outcome distribution to have very thin tails (as is required by SOL for applying Cramer's theorem); it is sufficient that first moments are finite.

3 Mean-maximization: Set-up

3.1 LP formulation

I will first outline the necessary LP results in the context of mean maximization and will return to the quantile case in section 6.

Let the vector of conditional mean of GPA of the m types (obtained from a random assignment of the entire population) be denoted by $g = (g_j)_{j=1,..m}$. Let the constraints be given by $\mathcal{P} = \{\mathbf{p} : A\mathbf{p} = \pi, \mathbf{p} \geq 0\}$, where A is a known $M \times m$ matrix. Then the planner's problem, if she knew g, is referred to as the "population" problem and is given by

$$\max_{\mathbf{p}=(p_j)_{j=1,..m}} \mathbf{p}'g \ \ \text{s.t.} \ \ A\mathbf{p} = \pi, \mathbf{p} \geq 0.$$

Typically, the π's will be known and g's will not be known. A random sample is assumed to have been drawn from the same population from which the target comes 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 $\hat{g} = (\hat{g}_j)$, $j = 1,..m$. The planner solves the problem in the previous display with g replaced by $\hat{g}$, which I will call the "sample problem". Whether this is the "optimal" action by the planner in the decision theoretic sense is an interesting and relevant question but is outside the scope of the present paper, which focusses on the asymptotic properties of a specific but "natural" action. For the asymptotic analysis of the decision theoretic optima in the treatment choice case see Hirano and Porter (2005). I note in passing that in my formulation, the constraint set $\mathcal{P}$ for the sample problem is identical to that of the population problem.

Notice that the constraint set $\mathcal{P}$ is a convex polytope with extreme points corresponding to the set of basic solutions (e.g. Luenberger (1984) page 19). Let $S = \{z_1, ...z_{|S|}\}$ denote the set of basic feasible solutions.[10] Then the fundamental theorem of LP implies

---

[10] Clearly $|S| \leq \binom{m}{M}$.

that

$$v = \max \left\{ z_1' g, ... z_{|S|}' g \right\} \text{ and } \hat{v} = \max \left\{ z_1' \hat{g}, ... z_{|S|}' \hat{g} \right\}.$$

Moreover, the compact and convex constraint set $\mathcal{P}$ equals $conv(S)$ where $conv(A)$ is the convex hull of set A (e.g. Luenberger (1984), page 470, theorem 3 and Grunbaum (1967) pages 14 and 36). This would imply that for the purpose of identification of the optimal allocation (and, as will be shown below, asymptotic inference), one can simply concentrate on the countably finite set S rather than the much larger constraint set $\mathcal{P}$.

3.2 Uniqueness

Since the $\hat{g}$'s will be asymptotically normally distributed, the sample solution will be unique at least for large enough n no matter whether the population solution is unique or not.[11] For the population solution, there are three possible scenarios: (i) the "degenerate" case: $z_1' g = z_2' g = ... = z_{|S|}' g$, (ii) the "intermediate" case: $\exists\ z_1 \neq z_2 \neq z_3 \in \mathcal{P}$ such that $z_1, z_2 \in \{z^* \in S : z^{*\prime} g = \max_{z \in S} (g'z) \equiv v\}$ and $g'z_1 = g'z_2 = v > g'z_3$ and (iii) the "unique" case: $g\prime z_1 > g'z_j$ 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 $g'z_1 = g'z_2 = ...g'z_J$ and $g'z_j < v$ 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).[12]

By way of illustration I now describe some scenarios where one can have situations (i), (ii) and (ii) using example 1 of the introduction. Case (i) will arise if for instance $g_{wb} = (g_{ww} + g_{bb})/2$, $g_{wo} = (g_{ww} + g_{oo})/2$ and $g_{bo} = (g_{bb} + g_{oo})/2$. Case (ii) will arise if e.g. $g_{ww} = g_{wb} = g_{bb} = a > b = g_{wo} = g_{bo} = g_{oo}$. 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

---

[11] Otherwise there will exist $z_1 \neq z_2$ st $\hat{g}'(z_1 - z_2) = 0$ which has 0 probability if $\hat{g}$ is continuously distributed.

[12] Given that $\mathcal{P} = conv(S)$ and since $g'p$ is a linear function of p, case (i) above is equivalent to the entire set $\mathcal{P}$ being optimal. Case (ii) is equivalent to there existing vectors $p \neq q \neq r \in \mathcal{P}$ such that $p, q \in \{p^* : p^{*\prime}g = \max_{p \in \mathcal{P}}(g'p)\}$ and $g'p = g'q > g'r$. And case (iii) is equivalent to z_1 being the only solution in $\mathcal{P}$.

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. $g_{wb} < \min\{g_{ww}, g_{bb}\}$, $g_{wo} < \min\{g_{ww}, g_{oo}\}$ and $g_{bo} < \{g_{oo}, g_{bb}\}$.

Remark 1 *It turns out that when $M = 2$, one can have either situation (i) or situation (iii) but not situation (ii). To see this, denote the two categories as w and o with all notations analogous to the example above. Then the LP problem is $\max\{g_{oo}p_{oo} + g_{ow}p_{ow} + g_{ww}p_{ww}\}$ s.t. $2p_{oo} + p_{wo} = 2\pi_o$ and $2p_{ww} + p_{wo} = 2\pi_w$ and $p_{oo}, p_{wo}, p_{ww} \geq 0$. Solving out, the objective function equals $g_{oo}\pi_o + g_{ww}\pi_w - p_{wo}(g_{ww} + g_{oo} - 2g_{wo})/2$. So when $g_{ww} + g_{oo} - 2g_{wo}$ is either strictly positive or strictly negative, there is a unique solution and if it is zero, all allocations are optimal. In the input matching problem considered in GIR (2006) with two binary covariates, either the population problem will have a unique solution or all allocations will be optimal. In other words, a nontrivial set-identified situation, of the type discussed in section 3 of this paper, will not arise there.*

3.3 Near-nonuniqueness

For the purpose of constructing confidence intervals for the maximum value (which will be done in the next section), it is important to consider situations where some extreme points give values that are within $c/\sqrt{n}$ of the maximum, where c is a fixed constant and n is the sample size. Such near-uniqueness can arise from an almost exact linear relationship between some of the g's, e.g. in example 1, if for some positive constants c_1, c_2, c_3 we have $g_{ob} = (g_{oo} + g_{bb})/2 + c_1/\sqrt{n}$, $g_{ow} = (g_{oo} + g_{ww})/2 + c_2/\sqrt{n}$ and $g_{bw} = (g_{bb} + g_{ww})/2 + c_3/\sqrt{n}$. The idea is that if there are points where the objective is very close to the maximum, the set of estimated optimal solutions will asymptotically include them and a confidence interval for the maximum value will need to take this fact into account. I will refer to this situation as the "nearly-nonunique" scenario. Near-nonuniqueness is probably a more realistic scenario than strict non-uniqueness in that it corresponds to existence of nearly exact relationships between the g's, rather than an exactly linear one (which may seem somewhat of a nongeneric knife-edge situation).

It will be shown below that near-nonuniqueness invalidates asymptotic approximations which assume that the solution is either unique or (strictly) nonunique and also renders subsampling-based inference inconsistent. The approach to asymptotic inference developed here will be robust to whether the solution is unique, nonunique and nearly

non-unique.

3.4 The asymptotic behavior of $\hat{g}$:

Rewrite

$$\hat{g}_j = \frac{\frac{1}{n}\sum_{i=1}^{n} D_{ij}y_i}{\frac{1}{n}\sum_{i=1}^{n} D_{ij}} \equiv \frac{\bar{y}_j}{\bar{d}_j} \text{ and } g_j = \frac{E\left(\bar{y}_j\right)}{E\left(\bar{d}_j\right)} = \frac{\mu_j}{\delta_j}$$

where D_{ij} is a dummy which equals 1 if the ith sampled individual, following random allocation, is in group type G_j, $i = 1,...n$ and $j = 1,...m$. 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.

For the rest of this paper, I will assume that $\delta_j >> \bar{m}$ for all j. Then $\hat{g}$ is consistent for g and $\sqrt{n}\left(\hat{g} - g\right)$ converges in distribution to a normal $N\left(0,\Sigma\right)$ which follows from classical weak laws and CLT's under standard conditions, since $\hat{g}$'s are ratios of means.

Covariance matrix: It is interesting to note that asymptotically, $\hat{g}_j$ and $\hat{g}_k$ will be uncorrelated for $j \neq k$ even if both their numerators and both their denominators are correlated. Asymptotically,

$$
\begin{aligned}
\sqrt{n}\left(\hat{g}_j - g_j\right) &= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(D_{ij}y_i - \mu_j\right)}{\delta_j} - \frac{\mu_j}{\delta_j^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\left(D_{ij} - \delta_j\right) + o_p\left(1\right) \\
&= \frac{\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_{ij}y_i}{\delta_j} - \frac{\mu_j}{\delta_j^2}\frac{1}{\sqrt{n}}\sum_{i=1}^{n} D_{ij} + o_p\left(1\right) \\
&= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\frac{D_{ij}}{\delta_j}\left(y_i - \frac{\mu_j}{\delta_j}\right) + o_p\left(1\right).
\end{aligned}
$$

So the asymptotic covariance is given by,

$$
E\left[\frac{D_{ij}}{\delta_j}\left(y_i - \frac{\mu_j}{\delta_j}\right)\frac{D_{ik}}{\delta_k}\left(y_i - \frac{\mu_k}{\delta_k}\right)\right] - E\left[\frac{D_{ij}}{\delta_j}\left(y_i - \frac{\mu_j}{\delta_j}\right)\right]E\left[\frac{D_{ik}}{\delta_k}\left(y_i - \frac{\mu_k}{\delta_k}\right)\right] = 0
$$

since $D_{ij}D_{ik} \equiv 0$ and $E\left(D_{ij}y_i - \mu_j\right) = 0$. This implies that Σ will be diagonal and so estimated very precisely. A partial intuitive explanation for this is that the different components of $\hat{g}$ pertain to non-overlapping subpopulations (although this does not explain why the correlations in the denominators do not matter).

4 Mean Maximization: Asymptotic Properties

4.1 Solution(s)

In accordance with the discussion above on uniqueness and near-uniqueness, let $v = \sup_{z \in S} g'z$ and define the sets

$$
\begin{aligned}
\Theta_0 &= \{z \in S : g'z = v\} \\
\Theta &= \left\{ \{z_1, ... z_J\} \in S : \max_{1 \leq j \leq J} \lim_{n \to \infty} \sqrt{n} \left(v - g'z_j\right) < \infty \right\}
\end{aligned}
$$

and $\lim_{n \to \infty} \sqrt{n} \left(v - g'z_j\right) \to \infty$ for $j > J$ where $1 \leq J \leq |S|$. That is, Θ is the set of extreme points where the objective function is within an $O\left(1/\sqrt{n}\right)$-neighborhood of the maximum and the set of maximizers $\Theta_0 \sqsubseteq \Theta$. The following proposition states and establishes consistency of the sample solution in the Hausdorff sense. This proposition can be viewed as an extension of Manski and Tamer's (2002) result on consistency for set identified parameters to the conceptually new "near-unique" situation but specialized to a discrete parameter space. This extension to the near-nonunique case has non-trivial implications for the distribution of the estimated maximum value. In particular, possibility of near-uniqueness will prevent us from constructing an exact asymptotic confidence interval for the true maximum value, which will be shown below. To the best of my knowledge, results of this spirit do not exist in the SOL because it does not carefully consider issues of uniqueness.

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

$$
\Theta_n = \left\{ z^* \in S : \hat{g}'z^* \geq \max_{z \in S} \hat{g}'z - c_n \right\},
$$

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

$$
A = \left\{ z \in S : \lim_{n \to \infty} \sqrt{n}\left(v - g'z\right) \to \infty, \ \lim_{n \to \infty} \frac{v - g'z}{c_n} < \infty \right\},
$$

then (iii) $\Pr\left(\Theta_n \subset \Theta \cup A\right) \to 1$.

    Proof. Appendix ∎

The idea here is that by fixing a threshold c_n, it is guaranteed that the whole set Θ_0 is covered but in so doing, one has to "let in" some extra z's. These extra z's, collected in the set $\Theta \cup A$, are the ones where the objective function differs from v by a term which is of order smaller than either $n^{-1/2}$ (i.e. the z's in $\Theta \backslash \Theta_0$) or larger than $n^{-1/2}$ but smaller than c_n (the z's in A). If for all $j = 1, ,,, |S|$, one has that either $\lim_{n \to \infty} \sqrt{n} (v - g'z_j) = k$ with k either 0 or ∞, then $\Theta_0 = \Theta = \Theta \cup A$ is the conventional set of maximizers and one gets $\Pr(\Theta_n = \Theta) \to 1$, which is a stronger version of the conclusion $d(\Theta_n, \Theta) \xrightarrow{P} 0$ in Manski and Tamer (2002), where $d(.,.)$ is the Hausdorff distance. This stronger version holds because the parameter space is finite. When k can take finite non-0 values, I get the result that $\Pr(\Theta \subset \Theta_n) \to 1$. However, because near non-uniqueness is allowed for, I do not get that $\Pr(\Theta_n \subset \Theta) \to 1$.[13]

CI for solution set: In relation to the recent literature on set-identified parameters, the situation in the present paper is evidently somewhat unusual. Because of the discreteness of the parameter space, the estimated set contains the true set with probability approaching 1, i.e . $\Pr(\Theta \subset \Theta_n) \to 1$. Therefore, the estimated set itself behaves like a confidence set for the true set of maximizers (and near-maximizers) with coverage probability arbitrarily close to 1 for large enough n. This shows that one need not separately consider building asymptotic confidence regions for the identified set.

Rate of convergence: The proof of proposition 1 showed that the sample maxima $\hat{z}$ is included in the set Θ 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.

---

[13]Since c_n can be chosen larger than but arbitrarily close to $n^{-1/2}$, the set A can be made arbitrarily small. Here the set S is countably finite and so there exist choices of c_n such that A will in fact be empty. If A is empty, then one gets back the result that $\Pr(\Theta_n = \Theta) \to 1$. It will turn out nonetheless that the set A, even when nonempty, does not affect the asymptotic distribution of the estimated maximum value, but the entire set Θ and not just the set Θ_0 will affect that distribution. In other words, the set of extreme points where the true objective function is further than $n^{-1/2}$ from the true maximum will not affect the asymptotic distribution of the estimated maximum value. However, in order to simulate this distribution consistently using the set Θ_n, I will require the assumption of an empty A but I will not need that $\Theta = \Theta_0$. In other words, for consistent simulation, I can allow for near-nonuniqueness upto $n^{-1/2}$ but not beyond.

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

Proof.

$$\Pr\{a_n 1(\hat{z} \notin \Theta) \neq 0\} = \Pr[1(\hat{z} \notin \Theta) \neq 0] = \Pr[1(\hat{z} \notin \Theta) = 1] = \Pr[\hat{z} \notin \Theta] \to 0.$$

■

This result is somewhat nonstandard and very different from the exponential rates derived in SOL (c.f. Shapiro and de Mello (2000)), alluded to in the introduction. The SOL results concern the speed at which the probability of the estimated solution differing from the true solution(s) converges to zero, based on large deviations principles, which require thin tail assumptions (or bounded supports) for the outcome. The arbitrarily fast rate derived here concerns the (probabilistic) order of magnitude of the difference between the estimated solution and the true solutions. For instance if there is a unique solution z_0, then the SOL results say that $\Pr(\hat{z} \neq z_0) \to 0$ at the rate of $e^{-\rho n}$ for some $\rho > 0$ while what is shown in the corollary above is that $\Pr(a_n(\hat{z} - z_0) = 0) \to 0$ for any $a_n \to \infty$. It turns out that the latter type of result, i.e corollary 1, is sufficient to derive an asymptotic distribution theory for $\hat{v}$. Furthermore, the corollary applies under bounded second moments alone which is much weaker than thin tail requirements of large deviations results.[14]

4.2 Maximum value

The maximum value of the optimization program is often a parameter of significant interest, as explained in the introduction. Also recall that the maximum value is point-identified in this setting unlike the maxima. In this subsection, I derive a distributional theory for the estimated maximum value which will be used to construct a confidence interval for the population maximum. I will show 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

---

[14]For the sake of completeness, however, I derive a large deviations result for the median case in section 5 below. This derivation is nonstandard not only because medians are not averages but also because the estimates here will involve random denominators. Further, these large deviations result does not require a thin tail for the outcome and bounded second moments suffice, unlike the case for means. This is a result of the fact that the averages involved here are those of indicator functions which are bounded.

be nonstandard in that it will depend critically on uniqueness and near-nonuniqueness of the solutions. Two different approaches to the distribution theory will be taken. In one, we will always get a conservative one-sided confidence interval irrespective of the uniqueness regime. In the other approach, we can get an one sided CI which will have the exact (asymptotic) coverage probability for a certain "degree of uniqueness" and a conservative CI for other degrees. The second type of CI is more appealing because it is shorter but harder to compute.

Proposition 2 *(consistency of estimated maximum) If $\sqrt{n}\left(\hat{g}-g\right)=O_{p}\left(1\right)$, and $\hat{v}=\max_{p\in\mathcal{P}}\hat{g}'p\equiv\max_{z\in S}\hat{g}'z$, then*

$$p\lim\hat{v}=v$$

where $v=\max_{p\in\mathcal{P}}g'p$.

Proof. Let $\hat{z}=\arg\max_{z\in S}\hat{g}'z$. Then

$$\hat{v}-v=\left(\hat{g}-g\right)'\hat{z}-\sqrt{n}\left(v-g'\hat{z}\right)\times1/\sqrt{n}.$$

Now, $\Pr\left(\hat{z}\in\Theta\right)\rightarrow1$ (see proof of previous proposition), implying $\sqrt{n}\left(v-g'\hat{z}\right)=O_{p}\left(1\right)$; also $\hat{z}$ is bounded almost surely (since $\mathcal{P}$ is compact) and $p\lim\left(\hat{g}-g\right)=0$. So the RHS of the previous display is of the form $o_{p}\left(1\right)-O_{p}\left(1\right)\times o(1)=o_{p}\left(1\right)$. ∎

A distribution theory for the estimated maximum value is now presented. The key feature of this distribution is that it is not affected by sampling error in the estimation of the sample maximum $\hat{z}$.

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

$$\sqrt{n}\left(\hat{v}-v\right)=\left[\sqrt{n}\left(\hat{g}-g\right)'\hat{z}-\sqrt{n}\left(v-g'\hat{z}\right)\right]1\left(\hat{z}\in\Theta\right)+o_{p}\left(1\right),\tag{4}$$

with $\Pr\left(\hat{z}\in\Theta\right)\rightarrow1$. If $\sqrt{n}\left(v-g'z_{j}\right)\rightarrow k_{j}\geq0$ for $1\leq j\leq J$, then

$$\sqrt{n}\left(\hat{v}-v\right)=\max_{1\leq j\leq J}\left\{w'z_{j}-k_{j}\right\}+o_{p}\left(1\right).\tag{5}$$

Proof. Appendix ∎

Observe that in (4), terms involving $1\left(\hat{z}\notin\Theta\right)$ are at most $o_{p}\left(1\right)$ and hence the sampling error in the estimation of $\hat{z}$ 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 $\Pr\left(a_{n}1\left(\hat{z}\notin\Theta\right)\neq0\right)\rightarrow0$ for any sequence a_{n}.

Corollary 2

$$\sqrt{n}\left(\hat{v}-v\right) \leq \max_{z\in\Theta}\left[\sqrt{n}\left(\hat{g}-g\right)'z\right] + o_p\left(1\right).$$

Proof. Since $g'z \leq v$ for all $z \in \mathcal{P} \supset \Theta$, we have that $\sqrt{n}\left(v - g'\hat{z}\right)$ is both bounded and nonnegative for every value of n, whence the corollary follows. ∎

Corollary 3 *If all z in Θ satisfy $\sqrt{n}\left(g'z - v\right) \to 0$, then*

$$\sqrt{n}\left(\hat{v}-v\right) \xrightarrow{d} \max_{z\in\Theta}\left[w'z\right].$$

In particular, if $|\Theta| = 1$, then $\sqrt{n}\left(\hat{v}-v\right) = \sqrt{n}\left(\hat{g}-g\right)'z_0 + o_p\left(1\right)$ whose distribution can be simulated by drawing w from $N\left(0, \hat{\Sigma}\right)$ and calculating $w'\hat{z}$.

Proof. If all z in Θ satisfy $\sqrt{n}\left(g'z - v\right) \to 0$, then

$$
\begin{aligned}
&\sqrt{n}\left(\hat{v}-v\right)\\
=\ & \sqrt{n}\left(\max_{z\in\Theta}\hat{g}'z - v\right) + o_p\left(1\right)\\
=\ & \sqrt{n}\max_{z\in\Theta}\left(\hat{g}'z - g'z\right) + o_p\left(1\right)\\
=\ & \max_{z\in\Theta}\left[\sqrt{n}\left(\hat{g}-g\right)'z\right] + o_p\left(1\right).
\end{aligned}
$$

∎

As can be seen from (4), the asymptotic distribution of $\sqrt{n}\left(\hat{v}-v\right)$ depends on $\sqrt{n}\left(v - g'\hat{z}\right)$, whose limit as $n \to \infty$ is generally unknown. This prevents one from computing the exact asymptotic distribution of $\sqrt{n}\left(\hat{v}-v\right)$ and therefore, it does not seem possible to calculate an asymptotic CI for v with the exact coverage probability.

This gives rise to two alternative ways of constructing somewhat conservative CI's. The first is based on the first corollary, i.e. using the asymptotic distribution of $\max_{z\in\Theta}\left[\sqrt{n}\left(\hat{g}-g\right)'z\right]$, which will provide at best a conservative confidence interval with coverage probability at least $1-\alpha$. Moreover, such a confidence interval will have to be necessarily one-sided, i.e. the CI are of the form $[\hat{v} - a/\sqrt{n}, \infty)$ where

$$1 - \alpha = \Pr\left(\max_{z\in\Theta}\left[\sqrt{n}\left(\hat{g}-g\right)'z\right] < a\right) \leq \Pr\left(\sqrt{n}\left(\hat{v}-v\right) < a\right).$$

The second method, which is described now, is computationally more intensive but is likely to provide CI's with smaller length and at least the nominal coverage probability.

Return to (4), let κ denote the vector $(k_1, ... k_J)$ and recall that

$$
\begin{aligned}
& \sqrt{n}\, (\hat{v} - v) \\
= & \sum_{j=1}^{J} [w'z_j - k_j]\, 1\, (w'\,(z_j - z_1) > (k_j - k_1)\, , ..., w'\,(z_j - z_J) > (k_j - k_J)) + o_p\,(1) \\
= & \max_{1 \leq j \leq J} \{w'z_j - k_j\} + o_p\,(1)\,.
\end{aligned}
$$

Since the k's are nonnegative, all quantiles of the distribution of $\sqrt{n}\,(\hat{v} - v)$ will be bounded above for all possible values of κ (if fewer than J of the k's increase to ∞, then the corresponding z's will not appear in the asymptotic distribution; if all the k's increase to ∞, the maximum will not be attained there since the quantile will tend to $\rightarrow -\infty$ then. The upper bound on the upper $1 - \alpha$ quantile, denoted by $q_{1-\alpha} = \sup_\kappa Q_{1-\alpha}\,(\max_{1 \leq j \leq J} \{w'z_j - k_j\})$, where $Q_{1-\alpha}\,(.)$ denotes the upper $(1 - \alpha)$ quantile, is therefore well-defined. Unfortunately, a corresponding lower bound for the lower αth quantile (which would give us an equal tailed CI) seems difficult to compute since the lower quantile is relatively flat in the k's when k's are relatively distant from zero. Neither in simulations, nor in the real application has the author been able to make the minimization algorithm converge. Yet, the upper bound on the upper $(1 - \alpha)$ quantile of the distribution of $\sqrt{n}\,(\hat{v} - v)$ is smaller than the one based on corollary 2 and therefore leads to a larger lower bound for a 1-sided CI for v.

On the other hand, constructing a symmetric two-tailed CI is problematic. Notice that by the continuous mapping theorem

$$
\left| \sqrt{n}\,(\hat{v} - v) \right| \xrightarrow{d} \left| \max_{1 \leq j \leq J} \{w'z_j - k_j\} \right|.
$$

But if all k's increase to ∞, the upper quantile of this distribution does not stay bounded. This makes it hard to simulate a 2-sided CI unless the value of k_j is known for at least one specific j.

To summarize, the computationally harder CI uses bounds which dominate only the relevant quantiles rather than using quantiles for a uniformly dominating distribution. Therefore, these CI are less conservative (shorter) for the same coverage probability, compared to the former computationally easier CI based on corollary 2. In neither case, can one get a 2-sided interval.

Simulating the CI: First consider the CI based on corollary 2. Notice that the distribution of $\max_{z \in \Theta} \left[\sqrt{n}\,(\hat{g} - g)'\, z \right]$ is not pivotal. However, since one can estimate

24

the asymptotic covariance matrix of $\hat{g}$, i.e. Σ, one can simulate the distribution of $\sqrt{n}\,(\hat{g} - g)$. Under the assumption that the set A defined in proposition 1 is empty (so that $\Pr(\Theta = \Theta_n) \to 1$), one can simply replace the unknown Θ everywhere in proposition 2 and its corollaries by the estimated set Θ_n. The assumption of empty A guarantees that asymptotically all extreme points z_j which enter Θ_n lead to a well-defined limiting distribution for $\{w'z_j - k_j\}$.[15] In particular, the lower limit of the CI can be calculated by drawing observations w from a $N\left(0, \hat{\Sigma}\right)$ and calculating $\max_{z \in \Theta_n}\{w'z\}$. Repeating this process many times, one can calculate the upper α quantile of this simulated distribution which gives a.

The second type of CI requires a potentially high dimensional optimization w.r.t. κ and can be computationally demanding if $|\Theta_n|$ is large. Notice that this CI is based on the relation that (up to $o_p(1)$ terms)

$$
\begin{aligned}
&\sqrt{n}\,(\hat{v} - v) \\
=\ &\sum_{j=1}^{J}\left[\sqrt{n}\,(\hat{g} - g)'\,z_j - \sqrt{n}\,(v - g'z_j)\right] \\
&\times 1\left(\sqrt{n}\,(\hat{g} - g)'\,(z_j - z_1) > \sqrt{n}g'\,(z_1 - z_j),\,...,\,\sqrt{n}\,(\hat{g} - g)'\,(z_j - z_J) > \sqrt{n}g'\,(z_J - z_j)\right)
\end{aligned}
$$

For each choice of $\kappa = \left(k_1,...k_{|\Theta_n|}\right)$, go through the following steps:

- (1) Draw w from a $N\left(0, \hat{\Sigma}\right)$

- (2) Calculate the vector $r_j = \left[w'\,(z_j - z_1),\,...w'\,(z_j - z_J)\right]$

- (3) Calculate

$$
v_j = 1\left[r_j \geq \left(k_j - k_1,\,....,\,k_j - k_{j-1},\,k_j - k_{j+1},\,...k_{|\Theta_n|} - k_J\right)\right].
$$

- (3) Calculate over $j = 1,...\,|\Theta_n|$, the sum

$$
U = \sum_{j=1}^{|\Theta_n|}\{[w'z_j] - k_j\}\,v_j
$$

- (4) Repeat steps 1-3 many times and each time compute the corresponding U.

---

[15]It does not seem possible to relax this assumption if the distribution of $\hat{v}$ is to be simulated consistently.

- (5) Calculate the upper quantile $Q_{1-\alpha}$ for the distribution of U generated by step 4.

- (6) Choose κ to maximize $Q_{1-\alpha}(\kappa)$, call this maximum value $q_{1-\alpha}$ and calculate CI by $[\hat{v} - q_{1-\alpha}/\sqrt{n}, \infty)$.

4.3   Inconsistency of Subsampling

It may seem natural to apply a resampling technique to construct a confidence interval for v in this setting. However, the subsampling distribution will not generally be consistent for the true asymptotic distribution in the problematic situation of near-nonuniqueness. To see this, consider the following simplified version of the original problem.

Let μ_1 and μ_2 be two real-valued unknown parameters (in the mean example above, assume that $\Theta = S = \{z_1, z_2\}$ and identify μ_j with $g'z_j$, $j = 1, 2$). Assume that we have usual asymptotically normal estimate $\hat{\mu}_j$ of μ_j, i.e. $\sqrt{n}\left(\hat{\mu}_j - \mu_j\right) \xrightarrow{d} N(0, j)$, $j = 1, 2$ and let $\sqrt{n}\left(\hat{\mu}_1 - \mu_1\right) - \sqrt{n}\left(\hat{\mu}_2 - \mu_2\right) \xrightarrow{d} N(0, 3)$ (e.g. $\hat{\mu}_1$ and $\hat{\mu}_2$ are independent). The problem is to construct a confidence interval for $\mu = \max\{\mu_1, \mu_2\}$, based on $\hat{\mu} = \max\{\hat{\mu}_1, \hat{\mu}_2\}$.

Assume that $(\mu_1 - \mu_2) = \sigma_n > 0$ (the case of $\sigma_n < 0$) is analogous. Then

$$
\begin{aligned}
&\sqrt{n}\left(\hat{\mu} - \mu\right) \\
={} &\sqrt{n}\left(\hat{\mu}_1 - \mu_1\right) 1\left(\hat{\mu}_1 - \hat{\mu}_2 > 0\right) + \sqrt{n}\left(\hat{\mu}_2 - \mu_1\right) 1\left(\hat{\mu}_2 - \hat{\mu}_1 > 0\right) \\
={} &\sqrt{n}\left(\hat{\mu}_1 - \mu_1\right) 1\left(\hat{\mu}_1 - \hat{\mu}_2 - \mu_1 + \mu_2 > -\sigma_n\right) \\
&+\sqrt{n}\left(\hat{\mu}_2 - \mu_1\right) 1\left(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1 > \sigma_n\right) \\
={} &\sqrt{n}\left(\hat{\mu}_1 - \mu_1\right) 1\left\{\sqrt{n}\left(\hat{\mu}_1 - \hat{\mu}_2 - \mu_1 + \mu_2\right) > -\sqrt{n}\sigma_n\right\} \\
&+\sqrt{n}\left(\hat{\mu}_2 - \mu_1\right) 1\left\{\sqrt{n}\left(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1\right) > \sqrt{n}\sigma_n\right\} \\
={} &\sqrt{n}\left(\hat{\mu}_1 - \mu_1\right) 1\left\{\sqrt{n}\left(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1\right) < \sqrt{n}\sigma_n\right\} \\
&+\sqrt{n}\left(\hat{\mu}_2 - \mu_2\right) 1\left\{\sqrt{n}\left(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1\right) > \sqrt{n}\sigma_n\right\} \\
&-\sqrt{n}\sigma_n 1\left\{\sqrt{n}\left(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1\right) > \sqrt{n}\sigma_n\right\}.
\end{aligned}
\tag{6}
$$

So

- if $\sqrt{n}\sigma_n \to \infty$, then

$$
\sqrt{n}\left(\hat{\mu} - \mu\right) = \sqrt{n}\left(\hat{\mu}_1 - \mu_1\right) + o_p(1) \xrightarrow{d} N(0, 1),
\tag{7}
$$

(note that

$$
\sqrt{n}\sigma_n\left[1 - \Phi\left(\sqrt{n}\sigma_n\right)\right] = \sqrt{n}\sigma_n\left[\Phi\left(-\sqrt{n}\sigma_n\right)\right] \to 0),
$$

26

so that the third term in (6) tends to 0 in expectation.

- if $\sqrt{n}\sigma_n \to k < \infty$, then

$$\sqrt{n}\left(\hat{\mu} - \mu\right) = \max\left\{\sqrt{n}\left(\hat{\mu}_2 - \mu_2\right) - k, \sqrt{n}\left(\hat{\mu}_1 - \mu_1\right)\right\} + o_p(1). \qquad (8)$$

Since we will not know the value of k or whether $\sqrt{n}\sigma_n \to \infty$, it is not possible to construct an asymptotic CI for μ which has exactly a pre-specified coverage probability under every value of k. Then the question is: will a subsampling based CI have this property? The answer is no. To see that, denote subsampling size by b and the subsample based estimate of μ by $\hat{\mu}^b$ with the usual requirement that $b \to \infty$ and $b/n \to 0$ as $n \to \infty$. As usual,

$$\sqrt{b}\left(\hat{\mu}^b - \hat{\mu}\right) = \sqrt{b}\left(\hat{\mu}^b - \mu\right) - \frac{\sqrt{b}}{\sqrt{n}}\sqrt{n}\left(\hat{\mu} - \mu\right) = \sqrt{b}\left(\hat{\mu}^b - \mu\right) + o_p(1).$$

So consider the distribution of $\sqrt{b}\left(\hat{\mu}^b - \mu\right)$.

$$
\begin{aligned}
&\sqrt{b}\left(\hat{\mu}^b - \mu\right) \\
={} &\sqrt{b}\left(\hat{\mu}_1^b - \mu_1\right)\mathbf{1}\left(\hat{\mu}_1^b - \hat{\mu}_2^b > 0\right) + \sqrt{b}\left(\hat{\mu}_2^b - \mu_1\right)\mathbf{1}\left(\hat{\mu}_2^b - \hat{\mu}_1^b > 0\right) \\
={} &\sqrt{b}\left(\hat{\mu}_1^b - \mu_1\right)\mathbf{1}\left(\hat{\mu}_1^b - \hat{\mu}_2^b - \mu_1 + \mu_2 > -\sigma_n\right) \\
&+ \sqrt{b}\left(\hat{\mu}_2^b - \mu_1\right)\mathbf{1}\left(\hat{\mu}_2^b - \hat{\mu}_1^b - \mu_2 + \mu_1 > \sigma_n\right) \\
={} &\sqrt{b}\left(\hat{\mu}_1^b - \mu_1\right)\mathbf{1}\left\{\sqrt{b}\left(\hat{\mu}_1^b - \hat{\mu}_2^b - \mu_1 + \mu_2\right) > -\sqrt{b}\sigma_n\right\} \\
&+ \sqrt{b}\left(\hat{\mu}_2^b - \mu_2\right)\mathbf{1}\left\{\sqrt{b}\left(\hat{\mu}_2^b - \hat{\mu}_1^b - \mu_2 + \mu_1\right) > \sqrt{b}\sigma_n\right\} \\
&- \sqrt{b}\sigma_n\mathbf{1}\left\{\sqrt{b}\left(\hat{\mu}_2^b - \hat{\mu}_1^b - \mu_2 + \mu_1\right) > \sqrt{b}\sigma_n\right\}.
\end{aligned}
$$

- When $\sqrt{n}\sigma_n \to \infty$,

  - if $\sqrt{b}\sigma_n \to \infty$, then $\sqrt{b}\left(\hat{\mu}^b - \mu\right) = \sqrt{b}\left(\hat{\mu}_1^b - \mu_1\right) + o_p(1) \overset{d}{\to} N(0,1)$ and
  - if $\sqrt{b}\sigma_n \to k_1 < \infty$, then

$$
\begin{aligned}
\sqrt{b}\left(\hat{\mu}^b - \mu\right) &= \max\left\{\sqrt{b}\left(\hat{\mu}_2^b - \mu_2\right) - k_1, \sqrt{b}\left(\hat{\mu}_1^b - \mu_1\right)\right\} + o_p(1) \\
&\overset{d}{\nrightarrow} N(0,1). \qquad (9)
\end{aligned}
$$

- When $\sqrt{n}\sigma_n \to k < \infty$, then $\sqrt{b}\sigma_n \to 0$ and so

$$\sqrt{b}\left(\hat{\mu}^b - \mu\right) = \max\left\{\sqrt{b}\left(\hat{\mu}_1^b - \mu_1\right), \sqrt{b}\left(\hat{\mu}_2^b - \mu_2\right)\right\} + o_p(1). \qquad (10)$$

27

Comparing (7) with (9) and (8) with (10) and recalling the fact that $\sqrt{b}\left(\hat{\mu}_j^b - \mu_j\right)$ and $\sqrt{n}\left(\hat{\mu}_j - \mu_j\right)$ converge to identical limits in distribution for $j = 1, 2$, one can see that $\sqrt{b}\left(\hat{\mu}^b - \mu\right)$ and $\sqrt{n}\left(\hat{\mu} - \mu\right)$ have the same asymptotic distribution either when $\sqrt{n}\sigma_n \to \infty$ and $\sqrt{b}\sigma_n \to \infty$ or else when $\sqrt{n}\sigma_n \to 0$ (which implies $\sqrt{b}\sigma_n \to 0$). But $\sqrt{n}\sigma_n \to \infty$ and $\sqrt{n}\sigma_n \to 0$ are precisely the two cases when the asymptotic distribution of $\sqrt{n}\left(\hat{\mu} - \mu\right)$ is fully known. In the problematic situation of $\sqrt{n}\sigma_n \to k < \infty$, or even when $\sqrt{n}\sigma_n \to \infty$ but $\sqrt{b}\sigma_n \to k_1 < \infty$, subsampling leads to inconsistent CI's.

4.4 Extensions

Assortative matching: Solution to the LP problem provides asymptotically a higher (not lower) 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 confidence intervals for relative efficiency of alternative allocations compared to the optimal one.

Risk aversion: The analysis for the means can be extended without any substantive changes to the situation where the planner puts different weights on different values of the outcome.[16] If this weighting function is $u\left(.\right)$, e.g. a concave "utility", define $g_j = E\left(u\left(y\right)|j\right)$, i.e. the expected value of $u\left(outcome\right)$ across subjects of type j and the overall mean utility- the objective function- as $\sum_j p_j g_j$, which is still linear in the p's. Therefore, the entire analysis presented above, which rests on linearity of the objective function and constraints in p, will carry through.

Degeneracy: A situation which can arise in allocation problems is that all *feasible* allocations yield the same aggregate outcome. A sufficient condition for this is that all allocations yield the same aggregate outcome. In the mean case for the room assignment example of the introduction, the sufficient condition will hold if, for instance, mean outcome of an individual conditional on own and roommate's characteristics is additively separable between own and roommate's characteristics. This makes the allocation problem vacuous and so one might be better-off by first checking that this is not the case.[17]

---

[16] I am grateful to Vadim Mermer for raising this issue.

[17] However, inference conditional on the conclusion of this test is subject to pre-testing considerations which are outside the scope of the present paper.

A test of degeneracy can be formulated as follows. Note that the null of degeneracy is equivalent to $\max_{z \in S} g'z = \min_{z \in S} g'z$. Under the null of degeneracy, the test statistic $T_n = \sqrt{n} \left(\max_{z \in S} \hat{g}'z - \min_{z \in S} \hat{g}'z \right)$ will be distributed as $U - V$ where U, V are defined by $\sqrt{n} \max_{z \in S} \left((\hat{g} - g)'z \right) \xrightarrow{d} U$ and $\sqrt{n} \min_{z \in S} \left((\hat{g} - g)'z \right) \xrightarrow{d} V$. Therefore the asymptotic null distribution of T_n can be simulated by the distribution of $\max_{z \in S} w'z - \min_{z \in S} w'z$ where w is a draw from a $N \left(0, \hat{\Sigma} \right)$. 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 $\hat{g}$ by w in the LP problem to get $\max_{z \in S} w'z$ and $\min_{z \in S} w'z$ which are the same as $\max_{z \in \mathcal{P}} w'z$ and $\min_{z \in \mathcal{P}} w'z$ 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 methods do not seem to be extendable to 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: The approach taken in the application is that I first test for degeneracy and upon rejection, construct the appropriate confidence intervals. The overall confidence interval is, therefore, given by

$$\hat{I} = 1 \left(T_n \geq c \right) CI_{nd} + 1 \left(T_n < c \right) CI_d,$$

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 and CI_d is the CI corresponding to the uninteresting case when degeneracy cannot be rejected. Consequently, for any $p \in \mathcal{P}$, one can use normal CI because

$$\sqrt{n} \left(\hat{v} - v \right) = \sqrt{n} \left(\hat{g} - g \right)' p \xrightarrow{d} N \left(0, p'\Sigma p \right).$$

We denote any such CI by (d_l, d_u). The following proposition describes the probability that this pre-test $\hat{I}$ covers the maximum value v.

Proposition 4 *Under the assumptions of propositions 3, 6 and 7, for all g,*

$$\Pr \left(v \in \hat{I} \right) \geq 1 - \alpha - \alpha',$$

where α' is the size of the test of degeneracy which rejects the null of degeneracy when $T_n > c$.

5   Quantiles

In this section, I present an important extension to this problem, viz. finding an allocation that maximizes a certain quantile rather than the mean of the (continuously distributed) outcome. Typically, policy-makers are concerned about distributional equity to some extent and maximizing the lower quantiles of the distribution is then the relevant optimization exercise. It is also interesting to compare the allocations that lead to maximizing upper and lower quantiles with those that maximize the mean, 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. Further, from a purely statistical standpoint, quantiles are robust to outliers unlike means, so that quantile based decisions can be viewed as more "democratic". For simplicity of notation, I concentrate on the median but the methods presented here apply to any fixed quantile. From a technical point of view, maximizing the median, unlike maximizing the mean, is not a standard M-estimation problem which makes the problem analytically different from e.g. calculating median regressions.

In the general case, let the covariate assume M possible values and let number of possible rooms be $m = M(M+1)/2$ as before. Denote by $F = (F_1, ...F_m)$ the vector of CDF's of the outcome for the m types of rooms. Let the constraint set $\{p : Ap = \pi\}$ be denoted by $\mathcal{P}$, as before. Let μ_p denote the population median corresponding to weighting $p \in \mathcal{P}$, i.e.

$$\mu_p = \inf\left(\mu : 0.5 \leq \sum_{j=1}^{m} p_j F_j(\mu)\right).$$

Then the population maximization problem can be stated as

$$\max_{p \in \mathcal{P}} \mu_p \tag{11}$$

Examining uniqueness of the population problem and finding asymptotic properties of the sample based estimate may first seem highly complicated due to the nonlinearity of μ_p in $\{p_j\}$. I will argue below that several of the key insights for the mean case actually

carry over to the median and the analysis is quite similar to the mean case although this is hardly obvious to start with.

The first proposition will establish that the "extreme point" idea, which is key to the mean analysis above, extends to the median. The "geometrical" idea involved is that if $r = \lambda p + (1 - \lambda) q \in \mathcal{P}$ with $\lambda \in (0, 1)$, then the median corresponding to the weighting r lies "between" the medians corresponding to the allocations p and q. This idea will be key to the rest of this section.

Proposition 5 *Let $p, q \in \mathcal{P}$. Let $r = \lambda p + (1 - \lambda) q \in \mathcal{P}$ with $\lambda \in (0, 1)$. Then r cannot be the unique solution to the population problem (11).*

Proof. Appendix ∎

It follows from the above proposition that just like in the mean case, one can focus on the extreme points of the constraint set which can be calculated a priori. To my knowledge, this insight is new in regards to the programming literature as well. 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 $\nu(.)$ of the CDF which satisfies

$$\nu\left(F_{\lambda p + (1-\lambda)q}\right) \leq \max\left\{\nu\left(F_p\right), \nu\left(F_q\right)\right\}$$

for every $\lambda \in [0, 1]$ and $F_p(.) \equiv \sum_{j=1}^{m} p_j F_j(.)$. It is not clear at this point if there is an alternative and more familiar characterization of the set of all such functionals $\nu(.)$. Means and quantiles seem to be 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

$$\Theta = \left\{\{z_1, ... z_J\} \in S : \max_{1 \leq j \leq J} \lim_{n \to \infty} \sqrt{n}\left(v - \mu_{z_j}\right) < \infty\right\}$$

and $\lim_{n \to \infty} \sqrt{n}\left(v - \mu_{z_j}\right) \to \infty$ for $j > J$ where $1 \leq J \leq |S|$,

$$\Theta_n = \left\{z^* \in S : \hat{\mu}_{z^*} \geq \max_{z \in S} \hat{\mu}_z - c_n\right\},$$

and replacing $g'z$ by μ_z which equals the median (or any other quantile) corresponding to the cdf $z'\mathbf{F}$ where $F\left(.\right)=\left(F_j\left(.\right)\right)_{j=1,...m}$.

The only step left to be specified is then the asymptotic distribution of $\sqrt{n}\left(\hat{\mu}_p - \mu_p\right)$ for any fixed p. To this end, define

$$\alpha_{pj} = E\left(D_{ij}1\left(Y_i \leq \mu_p\right)\right), \ f_j\left(.\right) = \frac{\partial}{\partial x}F_j\left(.\right),$$

assuming of course that these quantities exist where evaluated. It then follows from standard Taylor expansions and results on asymptotic normality of medians that

$$\sqrt{n}\left(\hat{\mu}_p - \mu_p\right) \ = \ \frac{1}{\sqrt{n}}\sum_{i=1}^{n}v_i + o_p\left(1\right) \ \text{where}$$

$$v_i \ = \ -\left(\sum_{j=1}^{m}p_j f_j\left(\mu_p\right)\right)^{-1}\sum_{j=1}^{m}p_j\left(\frac{D_{ij}1\left(Y_i \leq \mu_p\right) - \alpha_{pj}}{\delta_j} - \frac{\alpha_{pj}}{\delta_j^2}\left(D_{ij} - \delta_j\right)\right)$$

$$= \ -\left(\sum_{j=1}^{m}p_j f_j\left(\mu_p\right)\right)^{-1}\sum_{j=1}^{m}p_j D_{ij}\left(\frac{1\left(Y_i \leq \mu_p\right)}{\delta_j} - \frac{\alpha_{pj}}{\delta_j^2}\right).$$

5.1 A large deviations result

I now derive a large deviations type result for the median in the present setting. This result is analogous to the exponential rates derived for the mean in the stochastic optimization literature in statistics but its derivation is more complicated both because (estimated) medians are not simple sample averages (which is what the stochastic optimization literature concerns) and in allocation problems, they involve random denominators. Furthermore, in the case of the median, one does not need to restrict the outcome distribution to have very thin tails: it is sufficient that first moments are finite.

Lemma 1 *Let $p, q \in S$ be two allocations with corresponding medians satisfying $\mu_p > \mu_q$. Let $\hat{\mu}_p$, $\hat{\mu}_q$ denote their estimates. Assume that for every allocation p, F_p admits a density around its median μ_p. Then for some $c, \rho > 0$,*

$$\Pr\left(\hat{\mu}_p < \hat{\mu}_q\right) \leq ce^{-n\rho}.$$

Proof. Appendix ∎

The above lemma implies that the probability that the estimated median-maximizing solution belongs to Θ goes to 1 exponentially fast.

6 Application

I now apply these methods to calculate optimal allocation of freshmen to dorm rooms based on observed outcomes of a random assignment process at Dartmouth College for the graduating classes of 1997 and 1998. These data were previously analyzed by Bruce Sacerdote (2001) who provides more detailed descriptions of the various variables in the dataset.

The two outcomes I will consider are (i) freshman year GPA and (ii) eventual enrolment into a fraternity or sorority. The two separate covariates that I will use to design the optimal matching rule will be (i) an academic index, called ACA in this paper henceforth, which is a weighted sum of the student's high school GPA, high school class rank and SAT score and (ii) race. I will use the observed marginal distribution of covariates in the sample as a benchmark marginal distribution. But I emphasize that I still adhere to the theoretical set up which assumes that this marginal distribution is known exactly to the planner and an optimal matching rule maps each fixed set of marginals to an optimal joint distribution. The analysis will be done separately for men and women and will be restricted to individuals who were assigned to double rooms. Table 0 contains summary statistics for variables used. For all other details about the background and the assignment process, please see Sacerdote (2001).

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. So, 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 the effect of Z is separable from that of ACA or race and in particular, if Z has no effect on the outcome, conditional on race or ACA. (Notice that the issue here is whether different combinations of Z between the 2 roommates has different effects on aggregate Y for the room and not whether individual values of Z affect individual Y). If Z affects Y conditional on X, the estimated maximized value will still be biased unless an exact linear relationship exists between the means conditioned on different values of Z. These facts are shown in the appendix under the heading "Additive separability".

To address these problems, I first perform the analysis taking gender into account but ignoring Z and then condition on both gender and Z within those types of Z for which

the number of rooms is the largest. It turns out that for race (classified as minority and nonminority), exactly one type of Z (both roommates non-smoker, clean, non-music-listeners) had 2 or more rooms of each category for both men and women. For X=ACA and 2 categories, I chose the two largest blocks for men and for women (which turn out to be the same two blocks as in the race case). 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, I dropped all rooms (a total of 3 rooms out of about 400) which contained smokers. Therefore this analysis essentially applies to non-smokers and ignores the potentially confounding effects of cleanliness and music-listening.

When Y is joining a fraternity and X is ACA or race, it seems safe to assume that different combinations of Z do not have different effects on Y, for a given combination of X. So the analysis which ignores Z will likely give us consistent estimates of both the optimal allocation and value. But when Y is GPA and X is ACA or race, it is possible that some combinations of Z will have large negative effects on Y, 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 below should be interpreted with this caveat in mind.

First consider the case where the policy covariate is ACA. This variable assumes values between 151 and 231 in the data. I impose discreteness by dividing the sample into several ranges of ACA and show results 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 1 separately for men and women when the outcome of interest is mean freshman year GPA and in table 2A, 2B for the outcome "joining a fraternity/sorority in junior year".

When constructing confidence intervals based on the discussion in sections 3 and 4, I use a sequence $c_n = k \ln(n)/\sqrt{n}$ for three values of k- 0.01, 0.1 and 1. Since for my applications, 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 imply values of c_n equal to 0.003, 0.03 and 0.3. To compute the extreme points where necessary, I have used the "cdd" routines written by Komei Fukuda and coauthors which implements the double description algorithm of Motzkin et al. (1952).

Column 1 in table 1 reports the p-value for a test of degeneracy described in section 4.4. As is seen in table 1, degeneracy cannot be rejected anywhere.

In table 2A, the same exercise is repeated for the outcome "joining a Greek organization". Degeneracy is rejected at 5% or 10% level of significance for small (2) number of categories but not for the larger ones. This is likely 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 the last 2 columns. Column (6) gives a 2-sided CI which is from the simple normal distribution when $|\Theta_n| = 1$ and column (7) reports the conservative one-sided CI based on corollary 2, followed by the computationally intensive CI which is less conservative. The conservative versus non-conservative one sided CI are hardly different when number of categories is small but for women and larger number of categories, the nonconservative ones are indeed shorter. Note that in all tables 95% is to be interpreted as a "pretest" one, as discussed in section 4.4.

Table 2B describes the nature of the maximizing allocations corresponding to the cases where degeneracy was rejected in table 2A. Columns 2 and 3 of table 2B 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 2A, 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 form 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. The most likely explanation for this 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 her". For men, peers like oneself reinforce one's tendencies and this effect dominates.

Obviously, by imposing discreteness on ACA, one is getting a smaller maximum or higher minimum than if a finer partitioning was allowed (and, in the extreme case, if the continuous problem could be solved). On the other hand, a finer partition (i.e. more categories) implies that within-category means or cdf's will be estimated less precisely given limited amounts of data 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 and the estimated maximum (recall also the second point in section 1.1 above). Since sample sizes from experimental studies will typically be small, I believe that conditioning policy on fewer categories is a better strategy since that leads to good estimates of what is sought. Conditioning on more categories seeks a higher goal which is harder to attain, given data limitations and hence reallocation policies may erroneously seem futile. Indeed, as the number of ACA categories is increased to 4 from 2 in table 2A, tests of degeneracy fail to be rejected due to imprecise estimation[18].

Similar exercises with race are reported in tables 3 and 4A-B. The results here are sensitive to the definition of race. I 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 3 shows that when Y is GPA, degeneracy cannot be rejected. Corresponding results are reported in table 4A when the outcome is joining a Greek house. There seems to be significant nondegeneracy 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, there does not seem to be any evidence of non-degeneracy in (race-induced) peer effects for either GPA or joining fraternities.[19]

---

[18]Notice that for 2 categories we concluded nondegeneracy and for 4 categories, we must have a larger maximum and smaller minimum than with 2 categories. So the only way we can reject degeneracy for 2 but not for 4 is that standard errors are too high for the 4-category case.

[19]Other categorizations such as "non-blacks" and "blacks" caused problems in estimation since the number of rooms with two black men is one- which makes it impossible to estimate the requisite variance.

Finally, in table 5, I investigate the nature of maximizing allocations when the outcomes of interest are quantiles of the first year GPA distribution. I consider the case of 2 categories- H and L- 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. Analogously 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 5 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.

Thus, a policy-maker who wants to reduce inequality and discourage Greek enrolment, faces a policy dilemma in the case of women. This is because segregation enhances 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.

7   Conclusion

This paper draws on insights from the mathematical programming literature to study identification and estimation of optimal allocations and values in resource allocation problems. It is shown that when a planner is interested in maximizing mean outcome by choosing an allocation, the fundamental theorem of LP reduces the relevant parameter space to a countably finite set. This simplifies computation of the optimal solution to evaluating the objective function at a finite number of known points. It is shown that asymptotic properties of the estimated maximum value depend on uniqueness of the population solution. In particular, when the solution is nearly nonunique, I show that asymptotic approximations which assume either strict uniqueness or strict nonuniqueness are invalid and subsampling leads to inconsistent inference. A method of inference is consequently developed which is robust to the extent of uniqueness. A further contribution of the paper is to extend the analysis to maximizing quantiles of the outcome distribution and showing that almost all

the insights from mean-maximization carry over to the quantile-maximization case even if the quantile objective function, unlike the mean problem, is nonlinear in the allocation probabilities.

Randomized field experiments, which are becoming increasingly popular among applied microeconomists, 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 structural functions which could, in turn, be utilized to compute optimal allocations. So the applicability of methods developed in this paper are not restricted exclusively to randomized settings.

When these methods are applied to the randomized setting of room allocation for Dartmouth freshmen, the analysis reveals several interesting conclusions. First, segregation by previous background seems 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. This likely results from the average woman's tendency to find close companionship among other women with whom they can identify. 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. This highlights the importance of considering the effects of allocation rules on quantiles of the distribution if equity is to be an important goal of policy-making.

This paper does not analyze allocation problems from a decision theoretic standpoint as is being independently investigated by GIR (2006) in the context of input matching when the relevant covariate is binary. Nor does the present paper consider continuous covariates which is reserved for future research.

References

[1] 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.

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

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

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

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

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

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

[8] Grunbaum, B. (1967): Convex polytopes, John Wiley, New York.

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

[10] Hollander, F. (2000): Large Deviations, American Mathematical Society.

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

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

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

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

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

[16] T. S. Motzkin, H. Raifa, G. L. Thompson, and R. M. Thrall (1953): The double description method. In H. W. Kuhn and A. W. Tucker, editors, Contributions to the Theory of Games – Volume II, number 28, pages 51–73. Princeton University Press, Princeton.

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

[18] 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.

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

[20] 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.

[21] 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

Proposition 1:

$\quad$ **Proof.** Let $\hat{z} = \arg\max_{z \in S} \hat{g}'z$. First, I show that $\Pr(\hat{z} \in \Theta)$ converges to 1. Observe that

$$
\begin{aligned}
\Pr(\hat{z} \notin \Theta) \ &\leq \ \Pr\left[\cup_{j=J+1}^{|S|} (\hat{g}'z_j \geq \hat{g}'z_1)\right] \leq \sum_{j=J+1}^{|S|} \Pr(\hat{g}'z_j \geq \hat{g}'z_1) \\
&= \ \sum_{j=J+1}^{|S|} \Pr\left((\hat{g}-g)'(z_j - z_1) \geq g'(z_1 - z_j)\right) \\
&= \ \sum_{j=J+1}^{|S|} \Pr\left[\sqrt{n}(\hat{g}-g)'(z_j - z_1) \geq \sqrt{n}(v - g'z_j) - \sqrt{n}(v - g'z_1)\right]
\end{aligned}
$$

But for n large enough, $\sqrt{n}(v - g'z_j) \to \infty$ for every $j \geq J+1$, $\sqrt{n}(v - g'z_1)$ remains bounded and $\sqrt{n}(\hat{g}-g)'(z_j - z_1) = O_p(1)$. Therefore the probability in the last display goes to 0.

$\quad$ Now, assume that $z \in \Theta$. Then

$$
\begin{aligned}
\Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n) \ &= \ \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \in \Theta) + \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \notin \Theta) \\
&= \ \Pr(\hat{g}'z - g'z > \hat{g}'\hat{z} - g'\hat{z} + g'(\hat{z} - z) - c_n, \hat{z} \in \Theta) + o(1) \\
&= \ \Pr\left((\hat{g}-g)'(z - \hat{z}) > -c_n + g'(\hat{z}-z), \hat{z} \in \Theta\right) + o(1) \\
&= \ \Pr\left(\sqrt{n}(\hat{g}-g)'(z - \hat{z}) > -\sqrt{n}c_n + \sqrt{n}g'(\hat{z}-z), \hat{z} \in \Theta\right) + o(1).
\end{aligned}
$$

40

But for $z, \hat{z} \in \Theta$, we have $\lim_{n\to\infty} \sqrt{n} g'(\hat{z} - z) = \lim_{n\to\infty} \sqrt{n}[v - g'z] - \lim_{n\to\infty} \sqrt{n}(v - g'\hat{z}) < \infty$. Now, $\Pr(\hat{z} \in \Theta) \to 1$, $\sqrt{n}(\hat{g} - g) = O_p(1)$ and $(z - \hat{z}) = O_p(1)$ since $z, \hat{z} \in \mathcal{P}$ which is compact. Also, $-\sqrt{n} c_n \to -\infty$ as $n \to \infty$. So the above probability tends to 1. This implies that $\Pr(z \in \Theta_n) \to 1$ and so we have shown that $\Pr(\Theta \subset \Theta_n) \to 1$.

Conversely, suppose $z \notin \Theta$ and $z \notin A$. Then $\lim_{n\to\infty} \sqrt{n}(v - g'z) \to \infty$ and $\lim_{n\to\infty} \frac{v - g'z}{c_n} = \infty$. Then

$$
\begin{aligned}
\Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n) &= \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \in \Theta) + \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \notin \Theta) \\
&\overset{(1)}{=} \Pr(\hat{g}'z - g'z > \hat{g}'\hat{z} - g'\hat{z} + g'\hat{z} - g'z - c_n, \hat{z} \in \Theta) + o(1) \\
&= \Pr((\hat{g} - g)'(z - \hat{z}) > g'(\hat{z} - z) - c_n, \hat{z} \in \Theta) + o(1) \\
&= \Pr\left(\frac{(\hat{g} - g)'(z - \hat{z})}{c_n} > \frac{(v - g'z)}{c_n} - \frac{\sqrt{n}(v - g'\hat{z})}{\sqrt{n}c_n} - 1, \hat{z} \in \Theta\right) + o(12)
\end{aligned}
$$

Equality (1) uses the fact that $\Pr(\hat{z} \notin \Theta)$ converges to 0. Now, $z \notin \Theta$ and $z \notin A$ implies that $\frac{(v - g'z)}{c_n} \to \infty$. $\hat{z} \in \Theta$ plus $\sqrt{n}c_n \to \infty$ imply that $\frac{\sqrt{n}(v - g'\hat{z})}{\sqrt{n}c_n} \to 0$. Since $\Pr(\hat{z} \in \Theta) \to 1$ and $\frac{(\hat{g} - g)'(z - \hat{z})}{c_n} = \frac{\sqrt{n}(\hat{g} - g)'(z - \hat{z})}{\sqrt{n}c_n} = O_p(1) \times O_p\left(\frac{1}{\sqrt{n}c_n}\right) = o_p(1)$, we get that the probability in (12) goes to 0. This shows that $z \notin \Theta$ and $z \notin A$ imply that $\Pr(z \in \Theta_n) \to 0$. Thus we get that $\Pr(\Theta_n \subset \Theta \cup A) \to 1$. $\blacksquare$

Proposition 3:

Proof. Observe that

$$
\begin{aligned}
\sqrt{n}(\hat{v} - v) &= \sqrt{n}(\hat{g} - g)'\hat{z} + \sqrt{n}(g'\hat{z} - v) \\
&= \left[\sqrt{n}(\hat{g} - g)'\hat{z} + \sqrt{n}(g'\hat{z} - v)\right] 1(\hat{z} \in \Theta) \\
&\quad + \left[\sqrt{n}(\hat{g} - g)'\hat{z} + \sqrt{n}(g'\hat{z} - v)\right] 1(\hat{z} \notin \Theta) \\
&= \left[\sqrt{n}(\hat{g} - g)'\hat{z} + \sqrt{n}(g'\hat{z} - v)\right] X_n \\
&\quad + \sqrt{n}(\hat{g} - g)'\hat{z} 1(\hat{z} \notin \Theta) \\
&\quad + \sqrt{n}(g'\hat{z} - v) 1(\hat{z} \notin \Theta)
\end{aligned}
$$

where $X_n = 1(\hat{z} \in \Theta) \to 1$ in first mean and therefore in probability. Now, consider the second term on the RHS of the last display. Since $\hat{z} \in \mathcal{P}$, compact and $\sqrt{n}(\hat{g} - g) = O_p(1)$, we have that the second term is $o_p(1)$. Next, consider the third term. Observe that $(g'\hat{z} - v)$ 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 $\sqrt{n}(v - g'z_j) \to k_j \geq 0$ for $1 \leq j \leq J$. Then ignoring $o_p(1)$ terms

and denoting the asymptotic limit for $\sqrt{n}\,(\hat{g} - g)$ by w, one gets

$$
\sqrt{n}\,(\hat{v} - v)
$$

$$
= \left[\sqrt{n}\,(\hat{g} - g)'\,\hat{z} - \sqrt{n}\,(v - g'\hat{z}) \right] \mathbf{1}\,(\hat{z} \in \Theta)
$$

$$
= \sum_{j=1}^{J} \left[\sqrt{n}\,(\hat{g} - g)'\,z_j - \sqrt{n}\,(v - g'z_j) \right] \mathbf{1}\,(\hat{g}'\,(z_j - z_1) > 0 ..., \hat{g}'\,(z_j - z_J) > 0)
$$

$$
= \sum_{j=1}^{J} \left[\sqrt{n}\,(\hat{g} - g)'\,z_j - \sqrt{n}\,(v - g'z_j) \right]
$$

$$
\times \mathbf{1}\,\left(\sqrt{n}\,(\hat{g} - g)'\,(z_j - z_1) > \sqrt{n}g'\,(z_1 - z_j), ..., \sqrt{n}\,(\hat{g} - g)'\,(z_j - z_J) > \sqrt{n}g'\,(z_J - z_j) \right).
$$

Ignoring $o_p\,(1)$ terms, one gets

$$
\sqrt{n}\,(\hat{v} - v)
$$

$$
= \sum_{j=1}^{J} [w'z_j - k_j]\,\mathbf{1}\,\{\cap_{l \neq j, l=1,...J}\,\{w'\,(z_j - z_l) > k_j - k_l\}\}
$$

$$
= \sum_{j=1}^{J} [w'z_j - k_j]\,\mathbf{1}\,\{\cap_{l \neq j, l=1,...J}\,\{w'z_j - k_j > w'z_l - k_l\}\}
$$

$$
= \max_{1 \leq j \leq J}\,\{w'z_j - k_j\}.
$$

∎

Proposition 4:

Proof. For values of g for which the solution is nondegenerate (denoted by "ND"), we have

$$
\Pr\left(v \in \hat{I}|ND\right)
$$

$$
= \Pr\left(T_n \geq c,\, c_l \leq \sqrt{n}\,(\hat{g}'\hat{p} - v) \leq c_u|ND\right)
$$

$$
+ \Pr\left(T_n < c,\, d_l \leq \sqrt{n}\,(\hat{g}'\hat{p} - v) \leq d_u|ND\right). \tag{13}
$$

The second term is dominated by $\Pr\,(T_n < c|ND)$ which converges to 0 as $n \to \infty$ since the test based on T_n is consistent. This implies that the second term converges to 0. Observe that for any two events A_1, A_2, $\Pr\,(A_1 \cup A_2) \leq 1$ implying

$$
\Pr\,(A_1 \cap A_2) \geq \Pr\,(A_1) + \Pr\,(A_2) - 1. \tag{14}
$$

So the first term in (13) is at least

$$
\Pr\,(T_n \geq c|ND) - 1 + \Pr\left(c_l \leq \sqrt{n}\,(\hat{g}'\hat{p} - v) \leq c_u|ND\right)
$$

which converges to $(1 - \alpha)$ as $n \to \infty$ since the test based on T_n is consistent. So

$$\Pr\left(v \in \hat{I}|ND\right) \geq 1 - \alpha \tag{15}$$

and so $\hat{I}$ is at worst too conservative.

In the degenerate case (D), we have

$$
\begin{aligned}
&\Pr\left(v \in \hat{I}|D\right)\\
={}&\Pr\left(T_n \geq c,\ c_l \leq \sqrt{n}\left(\hat{g}'\hat{p} - v\right) \leq c_u|D\right)\\
&+\Pr\left(T_n < c,\ \hat{v} - \frac{d_H}{\sqrt{n}} \leq v \leq \hat{v} - \frac{d_L}{\sqrt{n}}|D\right).
\end{aligned}
$$

Using (14), the second probability is at least

$$
\begin{aligned}
&\Pr\left(T_n < c|D\right) + \Pr\left(\hat{v} - \frac{d_H}{\sqrt{n}} \leq v \leq \hat{v} - \frac{d_L}{\sqrt{n}}|D\right) - 1\\
&\underset{n \to \infty}{\to} (1 - \alpha') + (1 - \alpha) - 1 = 1 - \alpha - \alpha',
\end{aligned}
$$

where α' is the size of the test of the null $g \in \mathcal{R}(A)$ using T_n. Therefore,

$$\Pr\left(v \in \hat{I}|D\right) \geq 1 - \alpha - \alpha'. \tag{16}$$

From (15) and (16), the conclusion follows. ∎

Proposition 5:

Proof. WLOG assume that $\mu_p > \mu_q$.

Suppose $\mu_r > \mu_p$. Then

$$
\begin{aligned}
\sum_{j=1}^{m} r_j F_j\left(\mu_p\right) &= \lambda \sum_{j=1}^{m} p_j F_j\left(\mu_p\right) + (1 - \lambda) \sum_{j=1}^{m} q_j F_j\left(\mu_p\right)\\
&\overset{(2)}{\geq} \lambda \sum_{j=1}^{m} p_j F_j\left(\mu_p\right) + (1 - \lambda) \sum_{j=1}^{m} q_j F_j\left(\mu_q\right)\\
&= 0.5 = \sum_{j=1}^{m} r_j F_j\left(\mu_r\right).
\end{aligned}
$$

If (2) is an equality, then $\sum_{j=1}^{m} r_j F_j\left(\mu_p\right) = 0.5 = \sum_{j=1}^{m} r_j F_j\left(\mu_r\right)$, with $\mu_r > \mu_p$ contradicting the definition of μ_r. So we must have that

$$\sum_{j=1}^{m} r_j F_j\left(\mu_p\right) > \sum_{j=1}^{m} r_j F_j\left(\mu_r\right). \tag{17}$$

Now, since $\mu_r > \mu_p$, one has that for each j, $F_j(\mu_r) \geq F_j(\mu_p)$ implying $\sum_{j=1}^{m} r_j F_j(\mu_p) \leq \sum_{j=1}^{m} r_j F_j(\mu_r)$- which contradicts (17). This implies that $\mu_r \leq \mu_p$.

Next,

$$
\begin{aligned}
\sum_{j=1}^{m} r_j F_j(\mu_q) &= \lambda \sum_{j=1}^{m} p_j F_j(\mu_q) + (1-\lambda) \sum_{j=1}^{m} q_j F_j(\mu_q) \\
&= \lambda \times \sum_{j=1}^{m} p_j F_j(\mu_q) + (1-\lambda) \times 0.5 \\
&\overset{(1)}{<} \lambda \times 0.5 + (1-\lambda) \times 0.5 \\
&= 0.5 = \sum_{j=1}^{m} r_j F_j(\mu_r),
\end{aligned}
$$

implying that $\mu_q < \mu_r$ since each $F_j(.)$ is non-decreasing. The inequality $\overset{(1)}{<}$ follows from the definition of the medians. Indeed, if $\sum_{j=1}^{m} p_j F_j(\mu_q) = 0.5$ this contradicts that $\mu_p > \mu_q$ according to the definition of μ_p and the condition that $\mu_p > \mu_q$. The two displays above jointly show that we must have $\mu_p \geq \mu_r > \mu_q$. ∎

Lemma 1:

Proof. Notice that

$$
\begin{aligned}
\Pr(\hat{\mu}_p < \hat{\mu}_q) &= \Pr(\mu_q < \hat{\mu}_p < \hat{\mu}_q < \mu_p) + \Pr(\mu_q < \hat{\mu}_p < \mu_p < \hat{\mu}_q) \\
&\quad + \Pr(\hat{\mu}_p < \mu_q < \hat{\mu}_q < \mu_p) + \Pr(\hat{\mu}_p < \hat{\mu}_q < \mu_q < \mu_p). \quad (18)
\end{aligned}
$$

The first of these probabilities can be further decomposed into

$$
\begin{aligned}
&\Pr\left(\mu_q < \frac{1}{2}(\mu_p + \mu_q) < \hat{\mu}_p < \hat{\mu}_q < \mu_p\right) + \Pr\left(\mu_q < \hat{\mu}_p < \frac{1}{2}(\mu_p + \mu_q) < \hat{\mu}_q < \mu_p\right) \\
&+ \Pr\left(\mu_q < \hat{\mu}_p < \hat{\mu}_q < \frac{1}{2}(\mu_p + \mu_q) < \mu_p\right) \\
\leq\; &2\Pr\left(\mu_q < \frac{1}{2}(\mu_p + \mu_q) < \hat{\mu}_q\right) + \Pr\left(\hat{\mu}_p < \frac{1}{2}(\mu_p + \mu_q) < \mu_p\right) \\
\leq\; &2\Pr\left(\frac{1}{2}(\mu_p - \mu_q) < \hat{\mu}_q - \mu_q\right) + \Pr\left(\hat{\mu}_p - \mu_p < -\frac{1}{2}(\mu_p - \mu_q)\right).
\end{aligned}
$$

Rewriting $\frac{1}{2}(\mu_p - \mu_q)$ as $\delta > 0$, (18) is dominated by

$$
2\Pr(\delta < \hat{\mu}_q - \mu_q) + \Pr(\hat{\mu}_p - \mu_p < -\delta) + \Pr(2\delta < \hat{\mu}_q - \mu_q) + 2\Pr(\hat{\mu}_p - \mu_p < -2\delta).
$$

Now, notice that

$$
\Pr\left(\hat{\mu}_p - \mu_p < -\delta\right) = \Pr\left(\frac{1}{2} < \hat{F}_p\left(\mu_p - \delta\right)\right)
$$

$$
= \Pr\left(\frac{1}{2} - F\left(\mu_p - \delta\right) < \hat{F}_p\left(\mu_p - \delta\right) - F\left(\mu_p - \delta\right)\right)
$$

$$
= \Pr\left(\frac{1}{2} - F\left(\mu_p - \delta\right) < \sum_{j=1}^{m} p_j\left(\frac{\frac{1}{n}\sum_{i=1}^{n} D_{ij}1\left(y_i \le \mu_p - \delta\right)}{\bar{d}_j} - F_j\left(\mu_p - \delta\right)\right)\right)
$$

$$
\le \Pr\left(\cup_{j=1}^{m}\left\{\frac{\frac{1}{n}\sum_{i=1}^{n} D_{ij}1\left(y_i \le \mu_p - \delta\right)}{\bar{d}_j} - F_j\left(\mu_p - \delta\right) > \frac{\frac{1}{2} - F\left(\mu_p - \delta\right)}{m}\right\}\right)
$$

$$
\le \sum_{j=1}^{m}\Pr\left(\frac{1}{\bar{d}_j}\frac{1}{n}\sum_{i=1}^{n}\left\{D_{ij}1\left(y_i \le \mu_p - \delta\right) - \bar{d}_j F_j\left(\mu_p - \delta\right)\right\} > \frac{\frac{1}{2} - F\left(\mu_p - \delta\right)}{m}\right)
$$

$$
\le \sum_{j=1}^{m}\Pr\left(\frac{1}{n}\sum_{i=1}^{n}\left\{D_{ij}1\left(y_i \le \mu_p - \delta\right) - \bar{d}_j F_j\left(\mu_p - \delta\right)\right\} > \frac{\bar{m}}{m}\left(\frac{1}{2} - F\left(\mu_p - \delta\right)\right)\right)
$$

$$
+ \sum_{j=1}^{m}\Pr\left(\bar{d}_j - \delta_j < \bar{m} - \delta_j\right).
$$

Since $\delta > 0$, we have that $\bar{m}\left(\frac{1}{2} - F_p\left(\mu_p - \delta\right)\right) > 0$ since $F_p\left(.\right)$ admits a density in a neighborhood of the median; also $\bar{m} - \delta_j < 0$ for all j. Cramer's theorem implies that both probabilities go to zero exponentially fast since $E\left\{D_{ij}1\left(y_i \le \mu_p - \delta\right) - \bar{d}_j F_j\left(\mu_p - \delta\right)\right\} = E\left\{D_{ij}1\left(y_i \le \mu_p - \delta\right)\right\} - \delta_j F_j\left(\mu_p - \delta\right) = 0$. A similar argument works for the other terms in (18). ∎

Additive separability and ignoring Z: 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 $E\left(.\right)$ denote expectation w.r.t. the distribution induced by unconstrained allocation and $E^*\left(.\right)$ the one due to constrained allocation which does not allow SN type rooms. Then

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

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

$$
\begin{aligned}
E(Y|HH) &= \gamma_1(HH) + \gamma_2(SS)\Pr(SS|HH) + \gamma_2(SN)\Pr(SN|HH) + \gamma_2(NN)\Pr(NN|HH) \\
E(Y|HL) &= \gamma_1(HL) + \gamma_2(SS)\Pr(SS|HL) + \gamma_2(SN)\Pr(SN|HL) + \gamma_2(NN)\Pr(NN|HL) \\
E(Y|LL) &= \gamma_1(LL) + \gamma_2(SS)\Pr(SS|LL) + \gamma_2(SN)\Pr(SN|LL) + \gamma_2(NN)\Pr(NN|LL).
\end{aligned}
$$

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

$$
\begin{aligned}
&p_{HH}E(Y|HH) + p_{HL}E(Y|HL) + p_{LL}E(Y|LL) \\
={}& p_{HH}\gamma_1(HH) + p_{HL}\gamma_1(HL) + p_{LL}\gamma_1(LL) \\
&+ [\gamma_2(SS)\Pr(SS) + \gamma_2(SN)\Pr(SN) + \gamma_2(NN)\Pr(NN)] \qquad (19)
\end{aligned}
$$

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

$$
\begin{aligned}
&p_{HH}E^*(Y|HH) + p_{HL}E^*(Y|HL) + p_{LL}E^*(Y|LL) \\
={}& p_{HH}\gamma_1(HH) + p_{HL}\gamma_1(HL) + p_{LL}\gamma_1(LL) \\
&+ [\gamma_2(SS)\Pr{}^*(SS) + \gamma_2(NN)\Pr{}^*(NN)]. \qquad (20)
\end{aligned}
$$

Comparing (19) and (20), one can see that the two objective functions differ by a term which does not depend on the allocation. If the aggregate probability of smoking is ϕ_s then the second term in (19) is $[\phi_s^2\gamma_2(SS) + 2\gamma_2\phi_s(1 - \phi_s)\gamma_2(SN) + (1 - \phi_s)^2\gamma_2(NN)]$ while the second term in (20) is $[\phi_s\gamma_2(SS) + (1 - \phi_s)\gamma_2(NN)]$ which are equal if and only if

$$\gamma_2(SN) = \frac{1}{2}(\gamma_2(SS) + \gamma_2(NN)).$$

Table 0: Sample Characterestics

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 1: Degeneracy, Y=GPA, X=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 2A: Y=Joining Frat, X=ACA

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
# Categories	p-value	Sample Max	# vertices	Tol	Size of Theta_n	2-sided CI	Lower bound of 1-sided CI Con, Noncon
Men (mean=0.53)							
2	0.07*	0.571	2	0.003	1	0.568, 0.574	
				0.03	1	0.568, 0.575	
				0.3	2		0.5681, 0.5682
3	0.68	0.56	8	.	.		
4	0.215	0.626	40				
Women (mean=0.47)							
2	0.053*	0.513	2	0.003	1	0.509, 0.515	
				0.03	1	0.509, 0.515	
				0.3	2		0.510, 0.511
3	0.029*	0.54	8	0.003	1	0.534, 0.543	
				0.03	2		0. 5351, 0.5354
				0.3	8		0.5347, 0.5353
4	0.35	0.525	41				

Notes: This table shows the results of maximizing 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 maximum value of the probability of joining a frat/sorority. Column 3 contains the number of vertices of the convex polytope that is the feasible set of allocations. Column 4 contains the tolerance level used in the definition of the estimated set of maxima (denoted by theta_n in the text) and column 5 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 2nd type of computationally intensive CI described in the text. Column 7 contains the lower bounds corresponding to the conservative (based on corollary 1) and non-conservative one-sided CI. When the p-value is large (i.e. degeneracy cannot be rejected), no further analysis is pursued.

Table 2B: Y=Joining frat, X=ACA

(0) # Categories	(1) p-value	(2) HH and LL: Max	(3) 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 situations from table 2A where degeneracy was rejected. Column 1 reproduces the p-values from table 2A. 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 3A: X =Race: Non-minorities (L), minorities (H), Y=GPA

(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 trying to maximize and minimize freshman year GPA by allocating students based on race. The race of each student is classified as minority (Black, Hispanic, Native Indian) and non-minority (all others). 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.

4A: X =Race: Non-minorities (L), minorities (H), Y=Joining Frat

(0)	(1)	(2)	(3)	(4)	(5)	(6)	(7)
	# of vertices	p-value	Sample Max	Tol	Size of Theta_n	2-sided CI	Upper Bound of 1-sided CI Cons, Non-cons
Men (mean=0.53)	2	0.06	0.55	0.003	1	0.5505, 0.5557	.
				0.03	1	0.5505, 0.5557	.
				0.5	2		0.5503, 0.5505
Women (mean=0.47)	2	0.03	0.48	0.003	1	0.476, 0.482	.
				0.03	1	0.476, 0.482	.
				0.3	2		0.477, 0.477

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 tolerance level used in the definition of the estimated set of maxima (denoted by theta_n in the text) and column 5 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 2nd type of computationally intensive CI described in the text. Column 7 contains the lower bounds corresponding to the conservative (based on corollary 1) and non-conservative one-sided CI.

Table 4B: X=Race: Non-minorities (L), minorities (H), Y=Joining Frat

(0)	(1)	(2)	(3)
	p-value	HL: Max	HL: Min
Men (mean=0.53)	0.06	0.0	0.23
Women (mean=0.47)	0.03	0.22	0.0

Notes: This table shows the nature of allocations which maximize and minimize the probabilities of joining a frat/sorority corresponding to the situations from table 4A. Column 1 reproduces the p-values from table 4A. Column 2 (3) 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 2, first entry shows for example that for the probability-maximizing allocation, no room should be mixed. This table suggests that segregation leads to maximization of fraternity joining for men and minimization of sorority joining for women.

Table 5: Maximal allocations: Y=GPA, X=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 10th, 50th and 90 percentiles. Column 2 for men, for instance, shows that the maximized value of the 10th 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.