

# Inferring Optimal Resource Allocations from Experimental Data

Debopam Bhattacharya <sup>\*</sup>, <sup>†</sup>

Dartmouth College.

First draft: February 15, 2006; this draft: March 14, 2007.

## Abstract

This paper concerns the problem of optimally allocating scarce indivisible resources 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 develops asymptotic statistical inference on the estimated solutions and values which depends on the degree of uniqueness of the population solution. Subsampling based inference is shown to be invalid when near-uniqueness is allowed but conservative inference is possible. This work complements the "set identification" literature by conducting inference on the optimized value of the criterion function and the "treatment choice" literature by allowing aggregate resource constraints. Applicability of these techniques extends beyond linear maximands like the mean to other important policy objectives like outcome quantiles which, though nonlinear, are shown to be quasi-convex in the allocation probabilities. Implementation of these techniques using publicly available routines is discussed.

---

<sup>\*</sup>Address for correspondence: 327, Rockefeller Hall, Dartmouth College, Hanover, NH 03755. e-mail: debopam@dartmouth.edu.

<sup>†</sup>The original version of this paper included an application to a dataset of Dartmouth College undergraduates. But owing to confidentiality requirements, the micro-data cannot be made publicly available. Accordingly, at the suggestion of the editor, this application has been removed to avoid conflict with the data availability policy of the present journal and is being rewritten into a separate paper. The application, as it was reported in the original version, can be viewed at: [www.dartmouth.edu/~debopamb/Papers/allocation\\_application.pdf](http://www.dartmouth.edu/~debopamb/Papers/allocation_application.pdf)

# 1 Introduction

Design of optimal policies is inarguably an important goal of economic analysis. While the econometric methodology for evaluating specific public policies is now well-developed, how the knowledge derived from these evaluations actually translates to the design of optimal policies is a less discussed issue. The present paper develops the econometric methodology for inferring optimal policies, based on one's knowledge of the effect of alternative policies on a target population. Specifically, the paper considers a class of resource allocation problems under aggregate resource constraints. The goal is to infer the allocation(s) which will maximize an overall outcome for the entire population and calculate the corresponding maximum value. The objective function in this constrained optimization problem will typically involve components which are unknown but can be estimated by observing the outcomes of a sample of individuals, drawn from the target population, who were randomly assigned to alternative resource combinations. 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. The present paper develops precisely that methodology.

The formal set-up is as follows. A finite set of discretely valued resources is to be distributed among members of a target population. There are  $m$  possible distinct combinations of the different resources- where each combination may be viewed as one type of treatment. An allocation  $\mathbf{p}$  is an  $m$ -vector whose  $j$ th coordinate specifies the proportion of the population which will be randomly assigned to resource combination  $j$ . It is straightforward to generalize this to the case where the population is characterized by a set of observed covariates and allocations are to be made within each of the covariate categories (see example 2 below). If  $F_j(\cdot)$  denotes the cdf of outcome for individuals receiving the  $j$ th combination, then the cdf of the overall population outcome induced by the allocation  $\mathbf{p}$  is  $F_{\mathbf{p}} = \sum_{j=1}^m p_j F_j(\cdot)$ . Let the set of feasible allocations be denoted by  $\mathcal{P}$  and the social objective functional to be maximized by  $\Lambda(\cdot)$ . The problem of choosing the optimal policy is then given by

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

The set  $\mathcal{P}$  represents the constraints on the allocation probabilities which will include that  $\mathbf{p} \geq 0$  and  $\mathbf{p}'\mathbf{1} = 1$  possibly in addition to other constraints. 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. 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  $\Lambda(\cdot)$  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.

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. estimates  $\hat{F}_j(\cdot)$  are observed and the policymaker solves the problem

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

The primary objective of this paper is to characterize the statistical properties of the solution and optimal values of this sample maximization problem. 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 to the various combinations. If a random sample of  $n$  individuals were randomly assigned to  $m$  combinations  $G_j$ ,  $j = 1, \dots, m$  and  $y_i$  denotes the outcome of the  $i$ th individual, then a consistent estimate of  $F_j$  is given by

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

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

## 1.1 Examples

**Example 1 (roommate allocation under fixed marginals):**<sup>1</sup> 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,

---

<sup>1</sup>An actual application of this example may be viewed at: [www.dartmouth.edu/~debopamb/Papers/allocation\\_application.pdf](http://www.dartmouth.edu/~debopamb/Papers/allocation_application.pdf)

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  $\mathbf{p}'\mathbf{1} = 1$ . 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.

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

s.t.

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

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

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

where

$$\begin{aligned} \hat{\mathbf{g}} &= (\hat{g}_{ww}, \hat{g}_{bb}, \hat{g}_{oo}, \hat{g}_{wo}, \hat{g}_{wb}, \hat{g}_{bo})' \\ \mathbf{p} &= (p_{ww}, p_{bb}, p_{oo}, p_{wo}, p_{wb}, p_{bo})' \\ A &= \begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 & 0 & 1 \end{pmatrix} \\ \boldsymbol{\pi} &= (\pi_w, \pi_b, \pi_o)'. \end{aligned}$$

Alternatively, the planner could consider the problem of maximizing a quantile. Increasing the outcome at lower quantiles could be viewed as an equity enhancing policy. Moreover, as is well-known, means can be greatly affected by only a few outliers but the median is often a more robust measure of the representative outcome. Hence policies based on increasing the median can be viewed as more democratic. Define the  $\tau$ th quantile of the outcome distribution corresponding to allocation  $\mathbf{p}$  as

$$\mu_{\tau,p} = \inf_{\mu} \left( \sum_{i=w,b,o} \sum_{j=w,b,o} p_{ij} F_{ij}(\mu) \geq \tau \right)$$

where  $F_{ij}(\cdot)$  is the CDF of the outcome for individuals in roomtype  $(i,j)$ . Then the population problem for maximizing the  $\tau$ th quantile has the form

$$\max_{\mathbf{p}} \mu_{\tau,p} \text{ s.t. } A\mathbf{p} = \boldsymbol{\pi}, \mathbf{p} \geq \mathbf{0}.$$

The corresponding sample problem will replace  $\mu_{\tau,p}$  by the sample analog

$$\begin{aligned} \hat{\mu}_{\tau,p} &= \inf_{\mu} \left( \sum_{i \in \{w,b,o\}} \sum_{j \in \{w,b,o\}} p_{ij} \hat{F}_{ij}(\mu) \geq \tau \right) \text{ where} \\ \hat{F}_{ij}(\mu) &= \frac{\sum_{k=1}^n \mathbf{1}(y_k \leq \mu) \mathbf{1}(k \in (i,j))}{\sum_{k=1}^n \mathbf{1}(k \in (i,j))}. \end{aligned}$$

**Example 2 (treatment combination under budget constraint):** Suppose a policy-maker 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  $\mathbf{p}'\mathbf{1} = 1$ ,  $\mathbf{p} \geq \mathbf{0}$  and the policymaker's budget constraint

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

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

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

where  $\mu_p$  is defined by

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

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  $1, \dots, M$ , with  $\pi_j = \Pr(X = j)$ , then an allocation will be the vector  $\{\mathbf{p}_j\}_{j=1, \dots, M}$  where  $\mathbf{p}_j = (p_{tTj}, p_{tCj}, p_{cTj}, p_{cCj})$ , the budget constraint is

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

---

<sup>2</sup>In this simple example, it should be obvious that in an optimal allocation, 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 here provides an elegant and effective method for such settings especially when the number of possible allocations is large.

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

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

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.  $F(\cdot)$ . 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  $F(\cdot)$  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.

**Example 3 (One-to-one matching between asymmetric agents):**<sup>3</sup> Suppose a dating agency receives profiles of  $N$  men and  $N$  women clients with observed discrete characteristics whose supports are indexed by  $j = 1, \dots, J$  for men and  $k = 1, \dots, K$  for women. Let the corresponding marginal probabilities be denoted by  $\alpha_j$  and  $\beta_k$  respectively. Assume that the probability that a match between a randomly chosen  $j$  type man and a randomly chosen  $k$ -type woman leads to a formal date is given by  $g_{jk}$ . The agency chooses what fraction of matches will be of the  $(j, k)$  type and then randomly matches within the  $(j, k)$  types for each  $j$  and  $k$ . Suppose that the agency's objective is to maximize the expected number of dates. Then the population problem is

$$\begin{aligned} \max_{\{p_{jk}\}} & \sum_{j=1}^J \sum_{k=1}^K p_{jk} g_{jk} \text{ s.t.} \\ \sum_{j=1}^J p_{jk} &= \beta_k \text{ for } k = 1, \dots, K, \quad \sum_{k=1}^K p_{jk} = \alpha_j \text{ for } j = 1, \dots, J, \quad p_{jk} \geq 0 \text{ for all } j, k. \end{aligned}$$

---

<sup>3</sup>This example differs from example 1 in that two different types of agents are to be matched here unlike roommates who are symmetric.

To estimate the  $g$ 's consistently, one can utilize results of speed-dating experiments where couples were randomly matched, such as the ones conducted by Fisman et al (2006).

**Example 4 (Many-to-one matching):** Suppose a firm hires  $2N$  new workers and wants to assign them to  $N$  supervisors during the initial training period. To equalize workload, each supervisor is assigned exactly two trainees and each combination of one supervisor and two trainees will be called a team. Assume that the firm wants to choose optimal allocation based on gender and thus there are 2 types of trainees- call them m and f with marginal probabilities  $\alpha_m$  and  $\alpha_f$ - and two types of supervisors-call them M and F with marginal probabilities  $\beta_M$  and  $\beta_F$ . Observe that each supervisor can have one of three types of trainee combinations (m,m), (m,f), (f,f). These can be denoted by the index  $k = 1, 2, 3$  respectively. Now an allocation is a vector  $p = (p_{jk})$ ,  $j \in \{M, F\}$  and  $k = 1, 2, 3$  where, for instance  $p_{M2}$  denotes the fraction of teams with a male supervisor and one male and one female trainee. The corresponding expected productivity of that team (measured in terms of e.g. revenue generated by that team) is denoted by  $g_{jk}$ . Now the firm's problem of maximizing mean productivity can be stated as

$$\begin{aligned} \max_{\{p_{jk}\}} \quad & \sum_{j \in \{M, F\}} \sum_{k=1}^3 p_{jk} g_{jk} \text{ s.t.} \\ \sum_{j \in \{M, F\}} (2 p_{j1} + p_{j2}) &= 2\alpha_m, \quad \sum_{j \in \{M, F\}} (2 p_{j3} + p_{j2}) = 2\alpha_f \\ \sum_{k=1}^3 p_{jk} &= \beta_j \text{ for } j \in \{M, F\}, \quad p_{jk} \geq 0 \text{ for all } j, k. \end{aligned}$$

Just as before,  $g_{jk}$  can be consistently estimated if productivities are observed corresponding to a random allocation of trainees to supervisors. This example shows that many-to-one matching problems can be analyzed similarly by redefining "types".

**Remark 1** *An implicit assumption in all the matching examples above is that agents cannot reassign themselves after the initial allocation and before realization of the outcome- i.e. there is perfect compliance. So agents' own preferences over characteristics of partners are irrelevant here for designing the match and the planner is fully paternalistic. This makes these problems distinct from the two-sided matching literature (c.f. Roth and Sotomayor (1990)) which designs matching algorithms based on the agents' preferences, that are stable and Pareto optimal.*

## 1.2 Social Objectives

In this paper, I will consider two types of objective functionals  $\Lambda(\cdot)$  for the policymaker- (i) mean outcome and (ii) a quantile of the outcome distribution (when individual outcome is nonbinary). Maximization of mean amounts to maximizing productive efficiency. Maximization of central quantiles such as the median guarantees that allocation is not determined by influential outliers and can therefore be viewed as a more democratic criterion. Maximization of the lower quantiles can be interpreted as equity enhancing policies. 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 distributional goals can be targeted only at the allocation stage.

## 1.3 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 in many situations, the maximum value is an object of independent and significant policy interest. This is especially true when all technologically feasible allocations are not practically feasible due to political reasons. For example, allocations based on race or state-of-birth could be viewed as unfair 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 any specific allocation rule can be compared.

In example 1 above, if race-based allocation is not permitted and the planner allocates students randomly (since no other covariate is observed in this example), then the corresponding expected value is given by  $v_r = g'p_r$  where  $p_r = (\pi_w^2, \pi_b^2, \pi_o^2, 2\pi_w\pi_o, 2\pi_w\pi_b, 2\pi_o\pi_b)$ . The efficiency cost of non-discrimination is therefore  $(v - v_r)$  where  $v = g'p^*(g, \boldsymbol{\pi})$  and  $p^*(g, \boldsymbol{\pi})$  solves the LP problem of example 1.

Note that both the optimal allocation(s)  $p^*$  and the corresponding maximum value  $v^*$  in the population problem can be viewed as functionals of the underlying parameters  $F = (F_1(\cdot), \dots, F_m(\cdot))$ , i.e.  $v^* = v(F) \equiv \max_{\mathbf{p} \in \mathcal{P}} \Lambda\left(\sum_j p_j F_j(\cdot)\right)$ . The approach followed in this paper is to use consistent estimates  $\hat{F}$  of  $F$  to estimate  $v$  by  $\hat{v} = v(\hat{F})$ , show that such a  $\hat{v}$  is consistent and has an asymptotic distribution which can be simulated. Whether

it is possible to improve upon the estimator  $v(\hat{F})$  in terms of lower variance remains an open question. The difficulty in answering this question arises from the complicated nature of the functional  $v(\cdot)$  which is defined through the constrained optimization problem.

One could also investigate whether the simple and intuitive plug-in type rule described above is optimal in some decision theoretic sense. Such analysis, though relevant, is outside the scope of the present paper. See Manski (2004) and Hirano and Porter (2006) for a decision theoretic analysis of the unconstrained treatment choice case.

## 1.4 Qualifications

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. In example 1, it is also assumed that nature of interactions are unaffected by the aggregate proportions of  $w$ ,  $b$  and  $o$  types, which can fail to hold if for instance,  $w$  types are more accommodating of  $b$  types if  $b$  types are a small minority but more antagonistic if  $w$  types are a minority.

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  $E_P(\cdot)$  denotes the distribution in the pilot sample. Clearly, interaction between the two roommates' other covariates such as academic aptitude, noise tolerance etc. (jointly called  $X$ ) also affect mean GPA, so that

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

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. Even when heterogeneity is not a concern, conditioning on more covariates is likely to yield a larger maximum value. However, as one increases the number of covariates on which to match, the data requirement becomes more severe. Suppose one were to match students in example 1 on race (W,B,O), whether they were athletes in high school and mothers' educational status (e.g. high school, college, graduate degree). Now, there will be  $M=3*3*2=18$  types of students and  $M*(M+1)/2=171$  types of rooms. This is likely to 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 concerns an alternative strategy where one writes down an explicit model of outcomes as a function of the matches, estimates it using data from randomization and then performs the optimization by using those estimates as inputs. When covariates are discrete and *all interactions are allowed for*, subgroup level means used in the present approach are linear combinations of coefficients on the covariates and their interactions in the alternative model-based approach. For instance, corresponding to example 1, consider the model

$$GPA = \beta_0 + \beta_1 H + \beta_2 h + \beta_3 H * h + \beta_4 (1 - H) * (1 - h) + u$$

where  $GPA$  denotes an individual's GPA,  $H$  is a dummy which equals 1 if her own SAT was high and 0 otherwise and  $h$  is a dummy for whether her roommate's SAT was high and  $u$  is an error unrelated to  $H$  and  $h$  due to the randomization. Then  $g_{HH} = 2(\beta_0 + \beta_1 + \beta_2 + \beta_3)$ ,  $g_{HL} = 2\beta_0 + \beta_1$  and  $g_{LL} = 2(\beta_0 + \beta_4)$  and it does not matter whether one first estimates the  $\beta$ 's or estimates the  $g$ 's directly since the ultimate objective function is a linear combination of the  $g$ 's by definition. 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. 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. 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. 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. 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.  $med(y) \neq \int med(y|x) dF(x)$ . It is not clear how one may use a model-based analysis for maximizing a quantile.<sup>4</sup>

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.

## 1.5 Plan of the paper

Section 2 summarizes the contributions of the paper in relation to the existing literature. Section 3 and 4 focus on 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. Section 5 considers the analytically more interesting case of maximizing a quantile. Section 6 discusses ways of implementing the procedures and section 7 concludes. Longer proofs are collected in the appendix.

---

<sup>4</sup>With continuous covariates, the main technical difficulty 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.

## 2 Related Literature and Contributions

The main contribution of the present paper is to use insights from mathematical programming theory to analyze identification<sup>5</sup> and asymptotic statistical 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. A key finding is that 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, a uniformly valid asymptotic confidence interval (CI) for the maximum value cannot be constructed here. I further show that subsampling based confidence intervals are generally *inconsistent* when near-uniqueness is allowed for. Conservative one-sided CI for the maximum value are constructed which are robust to the degree of nonuniqueness of the solution. More generally, these methods concern inference on the maximized value of the criterion function in a set identified situation and thus can be viewed as an extension of that literature but specialized to a discrete parameter space.

The next key contribution of the present 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 is analytically a nonstandard exercise in that it cannot be interpreted as an M (or Z)-estimation problem, unlike e.g. quantile regressions. Apparently, analysis of identification and estimation in quantile maximization problems has not been attempted before<sup>6</sup> and the present paper addresses these issues in the context of resource allocation, where quantile maximization is potentially an important policy objective especially when the outcome cannot be ex-post redistributed among

---

<sup>5</sup> "Identification" in this paper refers to finding the solution(s) to the population problem. Identification of the objective function as a "causal" relationship is assumed here through the availability of randomized experimental data.

<sup>6</sup>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.

agents. The paper also describes what are the general class of policy maximands, beyond means and quantiles, 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,<sup>7</sup> 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 (i) focuses on one specific albeit natural plug-in type decision rule for a general discrete valued covariate, (ii) considers the maximization of both means and quantiles, (iii) investigates point and set-valued identification and (iv) conducts inference which is robust to the degree of uniqueness of the population solution. When the inputs are binary as in GIR (2006), 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 broadly related to the recent treatment choice literature, c.f. Manski (2004), Dahejia (2003) and Hirano and Porter (2005)- in that it concerns designing optimal policies based on results of a random experiment. But this

---

<sup>7</sup>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.

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.<sup>8</sup> 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 solving stochastic optimization problems. Birge et al (1998) provides a textbook treatment of this line of research. The classic 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 discusses statistical properties of solutions and values of certain stochastic optimization problems. The focus there is on inference in presence of inequality constraints on the parameters (c.f. Bartholomew (1957), Wolak (1987), 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 SOL are simple population averages and include neither ratios of averages (c.f. equation (2) above) nor more complicated functionals like quantiles which cannot be expressed as population means.<sup>9</sup>

---

<sup>8</sup>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.

<sup>9</sup>Note that quantile maximization is a very different problem from quantile regression, the latter having an objective function which is a population mean.

Second, SOL does not consider situations of nonuniqueness or *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 small term of order  $O(1/\sqrt{n})$  (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 for the type of maximands considered here turns out to be discrete which makes the magnitude of the difference between the estimated and the true solution(s) arbitrarily small in probability, asymptotically.<sup>10</sup> This result seems nonstandard and implies that sampling error in the estimation of the solutions has no effect on the distribution of the estimated maximum value. Note that this result is very different from large deviations (LD) based derivations in the SOL which require additional thin-tail assumptions but are unnecessary for my purpose (see section 3 for details). For completeness though, I will derive an LD result for the median case. This derivation is very different from the LD results of the SOL not only because medians are not averages (which is what SOL is really about) but also because the estimates here will involve random denominators. Further, one does not need to restrict the outcome distribution to have very thin tails; it is sufficient here that densities are positive around the median.

---

<sup>10</sup>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.

### 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 5.

Let the vector of conditional mean outcome for the  $m$  combinations (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

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

Typically, the  $\pi$ '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.

Notice that the constraint set  $\mathcal{P}$ , which is the same for the sample and population problems, is a convex polytope with extreme points corresponding to the set of basic solutions (e.g. Luenberger (1984) page 19). Let  $S = \{z_1, \dots, z_{|S|}\}$  denote the set of basic feasible solutions with  $|S| \leq \binom{m}{M}$ . Then the fundamental theorem of LP implies that

$$v = \max \{z'_1 g, \dots, z'_{|S|} g\} \quad \text{and} \quad \hat{v} = \max \{z'_1 \hat{g}, \dots, z'_{|S|} \hat{g}\}.$$

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

#### 3.2 Uniqueness

For the population solution, there are three possible scenarios: (i) the "degenerate" case:  $z'_1 g = z'_2 g = \dots = z'_{|S|} g$ , (ii) the intermediate case:  $\exists z_1 \neq z_2 \neq z_3 \in \mathcal{P}$  such that

$z_1, z_2 \in \{z^* \in S : z^*g = \max_{z \in S} (g'z) \equiv v\}$  and  $g'z_1 = g'z_2 = v > g'z_3$  and (iii) the unique case:  $g'z_1 > g'z_j$  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  $g'z_1 = g'z_2 = \dots g'z_J$  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). Given that  $\mathcal{P} = \text{conv}(S)$  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  $p, q \in \{p^* : p^*g = \max_{p \in \mathcal{P}} (g'p)\}$  and  $g'p = g'q > g'r$ . And case (iii) is equivalent to  $z_1$  being the only solution in  $\mathcal{P}$ .

By way of illustration I 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 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  $g_{bo} < \{g_{oo}, g_{bb}\}$ .

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 here, 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.<sup>11</sup> 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  $g_{bw} = (g_{bb} + g_{ww})/2 + c_3/\sqrt{n}$ . This would happen if, for instance,  $g_{jk} = \gamma_j + \delta_k + a/\sqrt{n}$ , i.e. the production function is "nearly" additively separable. Now, 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 Sample solution

Since the  $\hat{g}$ 's will be asymptotically normally distributed, the sample solution will likely be unique at least for large enough  $n$  no matter whether the population solution is unique or not. Otherwise there will exist  $z_1 \neq z_2$  s.t.  $\hat{g}'(z_1 - z_2) = 0$  which has 0 probability if  $\hat{g}$  is continuously distributed. If sample size is not large enough that  $\hat{g}$  is continuously distributed and sample solution is nonunique, one can pick any one of the sample solutions and call it  $\hat{z}$ .

---

<sup>11</sup>Such local parametrization in terms of the sample size is a common method of investigation in statistics (e.g. analysis of local power of hypothesis tests) and in econometrics (e.g. weak instrument asymptotics).

### 3.5 The asymptotic behavior of $\hat{g}$ :

Rewrite

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

where  $D_{ij}$  is a dummy which equals 1 if the  $i$ th sampled individual, following random allocation, is in group type  $G_j$ ,  $i = 1, \dots, n$  and  $j = 1, \dots, m$ . 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  $\delta_j \gg \bar{m}$  for all  $j$ . Then  $\hat{g}$  is consistent for  $g$  and  $\sqrt{n}(\hat{g} - g)$  converges in distribution to a normal  $N(0, \Sigma)$  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,

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

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  $E(D_{ij} y_i - \mu_j) = 0$ . This implies that  $\Sigma$  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

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

and  $\lim_{n \rightarrow \infty} \sqrt{n} (v - g'z_j) \rightarrow \infty$  for  $j > J$  where  $1 \leq J \leq |S|$ . That is,  $\Theta$  is the set of extreme points where the objective function is within an  $O(1/\sqrt{n})$ -neighborhood of the maximum. The set of maximizers  $\Theta_0 \subseteq \Theta$ . 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 a uniformly valid 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.

**Proposition 1** (*consistency of solution*) Assume that  $\sqrt{n}(\hat{g} - g) \xrightarrow{d} w \equiv N(0, \Sigma)$ . Then (i)  $\Pr(\hat{z} \in \Theta)$  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  $\sqrt{n}c_n \rightarrow \infty$ . Then (ii)  $\Pr(\Theta \subset \Theta_n) \rightarrow 1$ . Moreover, if  $c_n \rightarrow 0$  and

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

then (iii)  $\Pr(\Theta_n \subset \Theta \cup A) \rightarrow 1$ .

**Proof.** Appendix ■

The idea here is that by fixing a threshold  $c_n$ , it is guaranteed that the whole set  $\Theta_0$  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  $\Theta \setminus \Theta_0$ ) or larger than  $n^{-1/2}$  but smaller than  $c_n$  (the  $z$ 's in  $A$ ). If for all  $j = 1, \dots, |S|$ , one has that either  $\lim_{n \rightarrow \infty} \sqrt{n}(v - g'z_j) = k_j$  with  $k_j$  either 0 or  $\infty$ , then  $\Theta_0 = \Theta = \Theta \cup A$  is the conventional set of maximizers and one gets  $\Pr(\Theta_n = \Theta) \rightarrow 1$ , which is a stronger version of the conclusion  $d(\Theta_n, \Theta) \xrightarrow{P} 0$  in Manski and Tamer (2002), where  $d(\cdot, \cdot)$  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  $\Pr(\Theta \subset \Theta_n) \rightarrow 1$ . However, because near non-uniqueness is allowed for, I do not get that  $\Pr(\Theta_n \subset \Theta) \rightarrow 1$ . It will be shown below that the set  $A$ , even when nonempty, does not affect the asymptotic distribution of the estimated maximum value.

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

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

**Proof.**

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

■

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 LD 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  $\Pr(\hat{z} \neq z_0) \rightarrow 0$  at the rate of  $e^{-\rho n}$  for some  $\rho > 0$  while what is shown in the corollary above is that  $\Pr(a_n(\hat{z} - z_0) = 0) \rightarrow 0$  for any  $a_n \rightarrow \infty$ . 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 LD results.

I do not pursue construction of confidence sets for the set of solutions here since the methods for these are now well-known. Moreover, here one gets  $\Pr(\Theta_n = \Theta) \rightarrow 1$  and not just that  $\Pr(\Theta_n \supseteq \Theta) \rightarrow 1$  and so constructing a separate confidence set for  $\Theta$  with coverage probability less than 1 seems to be a somewhat unnecessary exercise.

## 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 be nonstandard in that it will depend on uniqueness and near-nonuniqueness of the solutions.

**Proposition 2** (*consistency of estimated maximum*) *If  $\sqrt{n}(\hat{g} - g) = O_p(1)$ , 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.** Appendix. ■

A distribution theory for the estimated maximum value is now presented. The two key features of this distribution are that (i) it is not affected by sampling error in the estimation of the sample maximum  $\hat{z}$  and (ii) it *may* involve nuisance parameters depending on the "degree of uniqueness" of the population solution.

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

$$\sqrt{n}(\hat{v} - v) = [\sqrt{n}(\hat{g} - g)' \hat{z} - \sqrt{n}(v - g' \hat{z})] 1(\hat{z} \in \Theta) + o_p(1), \quad (4)$$

with  $\Pr(\hat{z} \in \Theta) \rightarrow 1$ . If  $\sqrt{n}(v - g'z_j) \rightarrow k_j \geq 0$  for  $1 \leq j \leq J$ , then

$$\sqrt{n}(\hat{v} - v) = \max_{1 \leq j \leq J} \{w'z_j - k_j\} + o_p(1). \quad (5)$$

**Proof.** Appendix. ■

Observe that in (4), terms involving  $1(\hat{z} \notin \Theta)$  are at most  $o_p(1)$  and hence the sampling error in the estimation of  $\hat{z}$  does not affect the distribution of  $\sqrt{n}(\hat{v} - v)$ . As can be seen in the proof, this is a consequence of a discrete parameter space which ensures that  $\Pr(\sqrt{n}1(\hat{z} \notin \Theta) \neq 0) \rightarrow 0$ . The  $k_j$ 's are the nuisance parameters which will arise in the nearly nonunique case.

**Corollary 2**

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

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

**Corollary 3** *If all  $z$  in  $\Theta$  satisfy  $\sqrt{n}(g'z - v) \rightarrow 0$ , then*

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

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

**Proof.** If all  $z$  in  $\Theta$  satisfy  $\sqrt{n}(g'z - v) \rightarrow 0$ , then

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

■

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

But one can construct conservative CI's, based on the first corollary, i.e. using the asymptotic distribution of  $\max_{z \in \Theta} [\sqrt{n}(\hat{g} - g)'z]$ , 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

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

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

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

But if all  $k$ 's increase to  $\infty$ , 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  $j$ .

**Simulating the CI:** Notice that the distribution of  $\max_{z \in \Theta} [\sqrt{n} (\hat{g} - g)' z]$  is not pivotal. However, since one can estimate the asymptotic covariance matrix of  $\hat{g}$ , i.e.  $\Sigma$ , 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  $\Pr (\Theta = \Theta_n) \rightarrow 1$ ), one can simply replace the unknown  $\Theta$  everywhere in proposition 2 and its corollaries by the estimated set  $\Theta_n$ . The assumption of empty  $A$  guarantees that asymptotically all extreme points  $z_j$  which enter  $\Theta_n$  lead to a well-defined limiting distribution for  $\{w' z_j - k_j\}$ . It does not seem possible to relax this assumption if the distribution of  $\hat{v}$  is to be simulated consistently. In particular, the lower limit of the CI can be calculated by drawing observations  $w$  from a  $N(0, \hat{\Sigma})$  and calculating  $\max_{z \in \Theta_n} \{w' z\}$ . Repeating this process many times, one can calculate the upper  $\alpha$  quantile of this simulated distribution which gives  $a$ .

### 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  $\mu_1$  and  $\mu_2$  be two real-valued unknown parameters (in the mean example above, assume that  $\Theta = S = \{z_1, z_2\}$  and identify  $\mu_j$  with  $g' z_j$ ,  $j = 1, 2$ ). Assume that we have usual asymptotically normal estimate  $\hat{\mu}_j$  of  $\mu_j$ , i.e.  $\sqrt{n} (\hat{\mu}_j - \mu_j) \xrightarrow{d} N(0, j)$ ,  $j = 1, 2$  and let  $\sqrt{n} (\hat{\mu}_1 - \mu_1) - \sqrt{n} (\hat{\mu}_2 - \mu_2) \xrightarrow{d} N(0, 3)$  (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. Denote subsampling size by  $b$  and the subsample based estimate of  $\mu$  by  $\hat{\mu}^b$  with the usual requirement that  $b \rightarrow \infty$  and  $b/n \rightarrow 0$  as  $n \rightarrow \infty$ . Then subsampling is consistent only when  $\sqrt{n}\sigma_n \rightarrow \infty$  and  $\sqrt{b}\sigma_n \rightarrow 0$  which are precisely the two cases when the asymptotic distribution of  $\sqrt{n}(\hat{\mu} - \mu)$  is fully known. In the problematic situation of  $\sqrt{n}\sigma_n \rightarrow k < \infty$ , or even when  $\sqrt{n}\sigma_n \rightarrow \infty$  but  $\sqrt{b}\sigma_n \rightarrow k_1 < \infty$ , subsampling leads to invalid CI's. This fact is shown in details in the appendix.

This example adds to some recently discovered cases where subsampling fails to provide asymptotic confidence intervals which are uniformly valid (c.f. Andrews and Guggenberger (2006)). In the above example, one can think of  $\gamma = \mu_1 - \mu_2$  as a nuisance parameter and the asymptotic distribution of interest is discontinuous at  $\gamma = 0$ . Subsampling fails precisely when  $\gamma$  is local to 0.

## 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.<sup>12</sup> If this weighting function is  $u(\cdot)$ , e.g. a concave "utility", define  $g_j = E(u(y) | j)$ , i.e. the expected value of  $u(outcome)$  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. In the mean case for example 1, the sufficient condition will hold if mean outcome of an individual conditional on own and roommate's characteristics is additively separable between own and roommate's characteristics. This

---

<sup>12</sup>I am grateful to Vadim Mermer for raising this issue.

makes the allocation problem vacuous and so one should first check that this is not the case.

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  $T_n = \sqrt{n} (\max_{z \in S} \hat{g}'z - \min_{z \in S} \hat{g}'z)$  will be distributed as  $U - V$  where  $U, V$  are defined by  $\sqrt{n} \max_{z \in S} ((\hat{g} - g)'z) \xrightarrow{d} U$  and  $\sqrt{n} \min_{z \in S} ((\hat{g} - g)'z) \xrightarrow{d} V$ . 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  $N(0, \hat{\Sigma})$ . It is straightforward that this test will have nontrivial power against  $1/\sqrt{n}$  alternatives and will be consistent against fixed alternatives. Observe that implementing this test does not require calculating the extreme points in the mean case. One can simply solve the original LP problems to get  $T_n$  and then replace  $\hat{g}$  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 to be taken in an application would be to first test for degeneracy and upon rejection, construct the appropriate confidence intervals. The overall confidence interval is, therefore, given by

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

where  $c$  is the critical value used in the test of degeneracy, described above,  $CI_{nd} = (c_l, c_u)$  is the CI constructed upon rejection of degeneracy.  $CI_d$  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

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

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  $\alpha'$  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 combinations be  $m$  as before. Denote by  $F = (F_1, \dots, F_m)$  the vector of CDF's of the outcome for the  $m$  types of combinations. Let the constraint set  $\{p : Ap = \pi\}$  be denoted by  $\mathcal{P}$ , as before. Let  $\mu_p$  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

$$\max_{p \in \mathcal{P}} \mu_p \tag{6}$$

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  $\mu_p$  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 geometric 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 allocation  $r$  lies "between" the medians corresponding to the allocations  $p$  and  $q$ . This idea will be key to the rest of this section. The idea is illustrated in Figure 1 where  $F(p)$ ,  $F(q)$  and  $F(r)$  are the cdf's corresponding to allocations  $p$ ,  $q$  and  $r = \lambda p + (1 - \lambda) q$ , respectively. A formal statement and its proof now follow.

**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 (6).*

**Proof.** Appendix ■

It follows from the above proposition that for quantile maximization, just like in the mean case, one can focus on the extreme points of the constraint set which can be calculated a priori. 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  $\Lambda(\cdot)$  of the CDF which is quasi-convex, i.e. satisfies

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

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

$$\Theta = \left\{ \{z_1, \dots, z_J\} \in S : \max_{1 \leq j \leq J} \lim_{n \rightarrow \infty} \sqrt{n} (v - \mu_{z_j}) < \infty \right\} \quad (7)$$

and  $\lim_{n \rightarrow \infty} \sqrt{n} (v - \mu_{z_j}) \rightarrow \infty$  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  $\mu_z$  which equals the median (or any other quantile) corresponding to the cdf  $z'F$  where  $F(\cdot) = (F_j(\cdot))_{j=1, \dots, m}$ .

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

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

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

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

## 5.1 A large deviations result

I now derive a LD type result for the median in the present setting. This lemma will imply that the probability that the estimated median-maximizing solution belongs to  $\Theta$  goes to 1 exponentially fast. 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  $\mu_p$ . Then for some  $c, \rho > 0$ ,*

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

**Proof.** Appendix ■

**Corollary 4** *Under the hypotheses of Lemma 1, for some  $c, \rho > 0$ , we have*

$$\Pr(\hat{z} \notin \Theta) \leq ce^{-n\rho}.$$

## 6 Implementation

A natural approach to implementing these procedures is to first test for degeneracy as described in section 4 above. This test, as already pointed out, does not require the computation of the extreme points and is therefore easy to implement. If degeneracy cannot be rejected, then there is nothing left to do. If it is rejected, then one can go on to implement the maximization. In that sense, one may attach a 'pretest' interpretation to the confidence intervals as described in section 4.

Standard routines appear in all packages for solving linear programming problems. The computation of quantile maximization requires one to compute the extreme points of the constraint set. To do this, one can use the "cdd" routines written by Komei Fukuda and coauthors which implements the double description algorithm of Motzkin et al. (1952). Whether there exists more efficient algorithms for computing quantile maximizers is an interesting question, perhaps worth pursuing as a separate project.

When constructing confidence intervals based on the discussion in sections 3 and 4, one may use a sequence  $c_n = k \ln(n) / \sqrt{n}$  for a range of values of  $k$ - e.g. ones that are at most  $\ln(\ln n)$ .

An actual application of the type described in example 1 above which uses the above steps was reported in the original version of the paper. This application now appears as a separate document (please see footnote on page 1) and may be downloaded from the website: [www.dartmouth.edu/~debopamb/Papers/allocation\\_application](http://www.dartmouth.edu/~debopamb/Papers/allocation_application).

The data in the application pertains to about 800 Dartmouth College undergraduates who were randomly assigned to double rooms in dormitories. These data, which were

previously analyzed in Sacerdote (2001) to detect contextual peer effects, present the ideal setting in which to apply the methods of the present paper. On applying the methods, it was found that segregating Dartmouth students by race or previous academic achievements led to minimization of the probability of joining sororities by female students and maximization of probability of joining fraternities by male students. The most likely explanation for this is 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 appears to dominate. Reallocation based on race or prior academic credentials appeared to have no effect on mean (freshman year) GPA. However, full segregation by prior academic standing led to maximization of the 90th percentile and minimization of the 10th percentile of the GPA distribution for both men and women. A detailed analysis of this application appears at the location cited above and presents an integrated application of the methods proposed here.

## 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 and implies an arbitrarily fast rate of convergence of estimated solution(s) to the true solution(s). It is further shown that asymptotic properties of the estimated maximum value depend qualitatively 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. An important 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 (c.f. Duflo (2005)), 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.

This paper does not analyze allocation problems from a decision theoretic standpoint and does not consider continuous covariates both of which are reserved for future research.

## References

- [1] Andrews, D & P. Guggenberger (2006): The limit of finite sample size and a problem with subsampling, mimeo.
- [2] 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.
- [3] Birge, J.R. & F. Louveaux (1997): *Introduction to stochastic programming*, Springer.
- [4] Dehejia, Rajeev H (2005): Program Evaluation as a Decision Problem, *Journal of Econometrics*, vol. 125, no. 1-2, pp. 141-73.
- [5] Duflo, E. (2005): *Field Experiments in Development Economics*, 2005, mimeo.
- [6] Fisman, R., Iyengar, S., Kamenica, E. and I. Simonson (2006): Gender difference in mate selection: theory and evidence, *Quarterly Journal of Economics*, vol. 121, no. 2, pp. 673-97.
- [7] Fukuda, Komei: cdd routines, [http://www.ifor.math.ethz.ch/~fukuda/cdd\\_home/cdd.html](http://www.ifor.math.ethz.ch/~fukuda/cdd_home/cdd.html)
- [8] Graham, B., G. Imbens and G. Ridder (2005): Complementarity and Aggregate Implications of Assortative Matching: A Nonparametric Analysis, manuscript.
- [9] Graham, B., G. Imbens and G. Ridder (2006): Complementarity and the Optimal Allocation of Inputs, manuscript.

- [10] Grunbaum, B. (1967): Convex polytopes, John Wiley, New York.
- [11] Hirano, K. and J. Porter (2005): Asymptotics for Statistical Treatment Rules. Manuscript.
- [12] Hollander, F. (2000): Large Deviations, American Mathematical Society.
- [13] Luenberger, D. (1984): Linear and nonlinear programming. Addison Wesley.
- [14] Manski, C. (1988): Ordinal utility models of decision making under uncertainty, Theory and Decision, 25, pp. 79-104.
- [15] Manski, C. (1995): Identification problems in the social sciences, Cambridge and London: Harvard University Press, 1995.
- [16] Manski, C. (2004): Statistical Treatment Rules for Heterogeneous Populations, Econometrica, vol. 72, no. 4, pp. 1221-46.
- [17] Manski and Tamer (2002): Inference on Regressions with Interval Data on a Regressor or Outcome, Econometrica, 70, 519-546.
- [18] 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.
- [19] Rostek, M. (2005): Quantile maximization in decision theory, manuscript.
- [20] 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.
- [21] Roth, A. and M. Sotomayor (1990): Two-sided matching, Cambridge University Press.
- [22] Shapiro, A. (1989): Asymptotic properties of statistical estimators in stochastic programming, Annals of statistics, vol. 17, No. 2, pp. 841-858.
- [23] 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.

[24] 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:

**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

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

But for  $n$  large enough,  $\sqrt{n}(v - g'z_j) \rightarrow \infty$  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.

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

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

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

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

$\infty$ . Then

$$\begin{aligned}
\Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n) &= \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \in \Theta) + \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \notin \Theta) \\
&\stackrel{(1)}{=} \Pr(\hat{g}'z - g'z > \hat{g}'\hat{z} - g'\hat{z} + g'\hat{z} - g'z - c_n, \hat{z} \in \Theta) + o(1) \\
&= \Pr((\hat{g} - g)'(z - \hat{z}) > g'(\hat{z} - z) - c_n, \hat{z} \in \Theta) + o(1) \\
&= \Pr\left(\frac{(\hat{g} - g)'(z - \hat{z})}{c_n} > \frac{(v - g'z)}{c_n} - \frac{\sqrt{n}(v - g'\hat{z})}{\sqrt{nc_n}} - 1, \hat{z} \in \Theta\right) + o(1) \tag{8}
\end{aligned}$$

Equality (1) uses the fact that  $\Pr(\hat{z} \notin \Theta)$  converges to 0. Now,  $z \notin \Theta$  and  $z \notin A$  implies that  $\frac{(v - g'z)}{c_n} \rightarrow \infty$ .  $\hat{z} \in \Theta$  plus  $\sqrt{nc_n} \rightarrow \infty$  imply that  $\frac{\sqrt{n}(v - g'\hat{z})}{\sqrt{nc_n}} \rightarrow 0$ . Since  $\Pr(\hat{z} \in \Theta) \rightarrow 1$  and  $\frac{(\hat{g} - g)'(z - \hat{z})}{c_n} = \frac{\sqrt{n}(\hat{g} - g)'(z - \hat{z})}{\sqrt{nc_n}} = O_p(1) \times O_p\left(\frac{1}{\sqrt{nc_n}}\right) = o_p(1)$ , we get that the probability in (8) goes to 0. This shows that  $z \notin \Theta$  and  $z \notin A$  imply that  $\Pr(z \in \Theta_n) \rightarrow 0$ . Thus we get that  $\Pr(\Theta_n \subset \Theta \cup A) \rightarrow 1$ . ■

**Proposition 2** (consistency of estimated maximum):

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

$$\hat{v} - v = (\hat{g} - g)' \hat{z} - \sqrt{n}(v - g'\hat{z}) \times 1/\sqrt{n}.$$

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

**Proposition 3:**

**Proof.** Observe that

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

where  $X_n = 1(\hat{z} \in \Theta) \rightarrow 1$  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  $\sqrt{n}(v - g'z_j) \rightarrow k_j \geq 0$  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

$$\begin{aligned}
& \sqrt{n}(\hat{v} - v) \\
&= [\sqrt{n}(\hat{g} - g)' \hat{z} - \sqrt{n}(v - g' \hat{z})] \mathbf{1}(\hat{z} \in \Theta) \\
&= \sum_{j=1}^J [\sqrt{n}(\hat{g} - g)' z_j - \sqrt{n}(v - g' z_j)] \mathbf{1}(\hat{g}'(z_j - z_1) > 0, \dots, \hat{g}'(z_j - z_J) > 0) \\
&= \sum_{j=1}^J [\sqrt{n}(\hat{g} - g)' z_j - \sqrt{n}(v - g' z_j)] \\
&\quad \times \mathbf{1}(\sqrt{n}(\hat{g} - g)'(z_j - z_1) > \sqrt{n}g'(z_1 - z_j), \dots, \sqrt{n}(\hat{g} - g)'(z_j - z_J) > \sqrt{n}g'(z_J - z_j)).
\end{aligned}$$

Ignoring  $o_p(1)$  terms, one gets

$$\begin{aligned}
& \sqrt{n}(\hat{v} - v) \\
&= \sum_{j=1}^J [w' z_j - k_j] \mathbf{1}\{\cap_{l \neq j, l=1, \dots, J} \{w'(z_j - z_l) > k_j - k_l\}\} \\
&= \sum_{j=1}^J [w' z_j - k_j] \mathbf{1}\{\cap_{l \neq j, l=1, \dots, J} \{w' z_j - k_j > w' z_l - k_l\}\} \\
&= \max_{1 \leq j \leq J} \{w' z_j - k_j\}.
\end{aligned}$$

■

### Inconsistency of subsampling:

**Proof.** Note that

$$\begin{aligned}
& \sqrt{n}(\hat{\mu} - \mu) \\
&= \sqrt{n}(\hat{\mu}_1 - \mu_1) \mathbf{1}(\hat{\mu}_1 - \hat{\mu}_2 > 0) + \sqrt{n}(\hat{\mu}_2 - \mu_1) \mathbf{1}(\hat{\mu}_2 - \hat{\mu}_1 > 0) \\
&= \sqrt{n}(\hat{\mu}_1 - \mu_1) \mathbf{1}\{\sqrt{n}(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1) < \sqrt{n}\sigma_n\} \\
&\quad + \sqrt{n}(\hat{\mu}_2 - \mu_2) \mathbf{1}\{\sqrt{n}(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1) > \sqrt{n}\sigma_n\} \\
&\quad - \sqrt{n}\sigma_n \mathbf{1}\{\sqrt{n}(\hat{\mu}_2 - \hat{\mu}_1 - \mu_2 + \mu_1) > \sqrt{n}\sigma_n\}.
\end{aligned} \tag{9}$$

So

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

$$\sqrt{n}(\hat{\mu} - \mu) = \sqrt{n}(\hat{\mu}_1 - \mu_1) + o_p(1) \xrightarrow{d} N(0, 1), \tag{10}$$

(note that

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

so that the third term in (9) tends to 0 in expectation.

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

$$\sqrt{n}(\hat{\mu} - \mu) = \max\{\sqrt{n}(\hat{\mu}_2 - \mu_2) - k, \sqrt{n}(\hat{\mu}_1 - \mu_1)\} + o_p(1). \quad (11)$$

Denote subsampling size by  $b$  and the subsample based estimate of  $\mu$  by  $\hat{\mu}^b$  with the usual requirement that  $b \rightarrow \infty$  and  $b/n \rightarrow 0$  as  $n \rightarrow \infty$ . As usual,

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

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

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

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

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

$$\sqrt{b}(\hat{\mu}^b - \mu) = \max\left\{\sqrt{b}(\hat{\mu}_2^b - \mu_2) - k_1, \sqrt{b}(\hat{\mu}_1^b - \mu_1)\right\} + o_p(1) \xrightarrow{d} N(0, 1) \quad (12)$$

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

$$\sqrt{b}(\hat{\mu}^b - \mu) = \max\left\{\sqrt{b}(\hat{\mu}_1^b - \mu_1), \sqrt{b}(\hat{\mu}_2^b - \mu_2)\right\} + o_p(1). \quad (13)$$

Comparing (10) with (12) and (11) with (13) and recalling the fact that  $\sqrt{b}(\hat{\mu}_j^b - \mu_j)$  and  $\sqrt{n}(\hat{\mu}_j - \mu_j)$  converge to identical limits in distribution for  $j = 1, 2$ , one can see that  $\sqrt{b}(\hat{\mu}^b - \mu)$  and  $\sqrt{n}(\hat{\mu} - \mu)$  have the same asymptotic distribution either when  $\sqrt{n}\sigma_n \rightarrow \infty$  and  $\sqrt{b}\sigma_n \rightarrow \infty$  or else when  $\sqrt{n}\sigma_n \rightarrow 0$  (which implies  $\sqrt{b}\sigma_n \rightarrow 0$ ). But  $\sqrt{n}\sigma_n \rightarrow \infty$  and  $\sqrt{b}\sigma_n \rightarrow 0$  are precisely the two cases when the asymptotic distribution of  $\sqrt{n}(\hat{\mu} - \mu)$  is fully known. In the problematic situation of  $\sqrt{n}\sigma_n \rightarrow k < \infty$ , or even when  $\sqrt{n}\sigma_n \rightarrow \infty$  but  $\sqrt{b}\sigma_n \rightarrow k_1 < \infty$ , subsampling leads to inconsistent CI's. ■

**Proposition 4:**

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

$$\begin{aligned} & \Pr(v \in \hat{I}|ND) \\ = & \Pr(T_n \geq c, c_l \leq \sqrt{n}(\hat{g}'\hat{p} - v) \leq c_u|ND) \\ & + \Pr(T_n < c, d_l \leq \sqrt{n}(\hat{g}'\hat{p} - v) \leq d_u|ND). \end{aligned} \quad (14)$$

The second term is dominated by  $\Pr(T_n < c|ND)$  which converges to 0 as  $n \rightarrow \infty$  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

$$\Pr(A_1 \cap A_2) \geq \Pr(A_1) + \Pr(A_2) - 1. \quad (15)$$

So the first term in (14) is at least

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

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

$$\Pr(v \in \hat{I}|ND) \geq 1 - \alpha \quad (16)$$

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

In the degenerate case (D), we have

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

Using (15), the second probability is at least

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

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

$$\Pr(v \in \hat{I}|D) \geq 1 - \alpha - \alpha'. \quad (17)$$

From (16) and (17), the conclusion follows. ■

**Proposition 5:**

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

Suppose  $\mu_r > \mu_p$ . Then

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

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

$$\sum_{j=1}^m r_j F_j(\mu_p) > \sum_{j=1}^m r_j F_j(\mu_r). \quad (18)$$

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 (18). This implies that  $\mu_r \leq \mu_p$ .

Next,

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

implying that  $\mu_q < \mu_r$  since each  $F_j(\cdot)$  is non-decreasing. The inequality  $\stackrel{(1)}{<}$  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  $\mu_p$  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

$$\begin{aligned} \Pr(\hat{\mu}_p < \hat{\mu}_q) &= \Pr(\mu_q < \hat{\mu}_p < \hat{\mu}_q < \mu_p) + \Pr(\mu_q < \hat{\mu}_p < \mu_p < \hat{\mu}_q) \\ &\quad + \Pr(\hat{\mu}_p < \mu_q < \hat{\mu}_q < \mu_p) + \Pr(\hat{\mu}_p < \hat{\mu}_q < \mu_q < \mu_p). \end{aligned} \quad (19)$$

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  $\delta > 0$ , (19) is dominated by

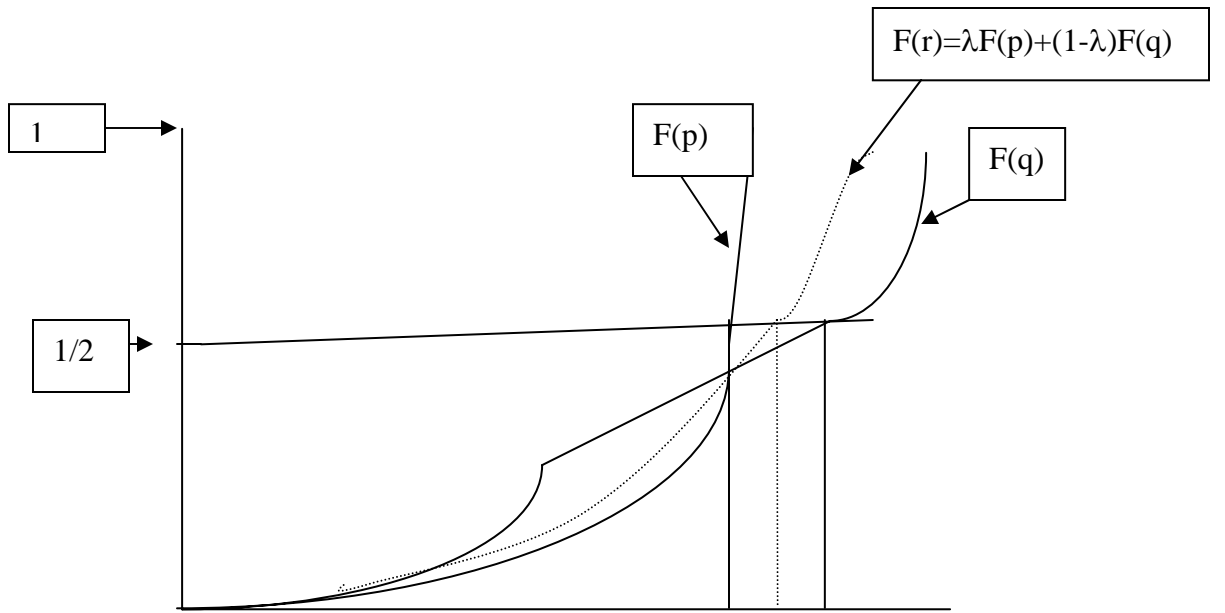
$$2\Pr(\delta < \hat{\mu}_q - \mu_q) + \Pr(\hat{\mu}_p - \mu_p < -\delta) + \Pr(2\delta < \hat{\mu}_q - \mu_q) + 2\Pr(\hat{\mu}_p - \mu_p < -2\delta).$$

Now, notice that

$$\begin{aligned} \Pr(\hat{\mu}_p - \mu_p < -\delta) &= \Pr\left(\frac{1}{2} < \hat{F}_p(\mu_p - \delta)\right) \\ &= \Pr\left(\frac{1}{2} - F(\mu_p - \delta) < \hat{F}_p(\mu_p - \delta) - F(\mu_p - \delta)\right) \\ &= \Pr\left(\frac{1}{2} - F(\mu_p - \delta) < \sum_{j=1}^m p_j \left(\frac{\frac{1}{n} \sum_{i=1}^n D_{ij} 1(y_i \leq \mu_p - \delta)}{\bar{d}_j} - F_j(\mu_p - \delta)\right)\right) \end{aligned}$$

$$\begin{aligned}
&\leq \Pr \left( \bigcup_{j=1}^m \left\{ \frac{\frac{1}{n} \sum_{i=1}^n D_{ij} 1(y_i \leq \mu_p - \delta)}{\bar{d}_j} - F_j(\mu_p - \delta) > \frac{\frac{1}{2} - F(\mu_p - \delta)}{m} \right\} \right) \\
&\leq \sum_{j=1}^m \Pr \left( \frac{1}{\bar{d}_j} \frac{1}{n} \sum_{i=1}^n \{D_{ij} 1(y_i \leq \mu_p - \delta) - \bar{d}_j F_j(\mu_p - \delta)\} > \frac{\frac{1}{2} - F(\mu_p - \delta)}{m} \right) \\
&\leq \sum_{j=1}^m \Pr \left( \frac{1}{n} \sum_{i=1}^n \{D_{ij} 1(y_i \leq \mu_p - \delta) - \bar{d}_j F_j(\mu_p - \delta)\} > \frac{\bar{m}}{m} \left( \frac{1}{2} - F(\mu_p - \delta) \right) \right) \\
&\quad + \sum_{j=1}^m \Pr(\bar{d}_j - \delta_j < \bar{m} - \delta_j).
\end{aligned}$$

Since  $\delta > 0$ , we have that  $\bar{m} \left( \frac{1}{2} - F_p(\mu_p - \delta) \right) > 0$  since  $F_p(\cdot)$  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  $E \{D_{ij} 1(y_i \leq \mu_p - \delta) - \bar{d}_j F_j(\mu_p - \delta)\} = E \{D_{ij} 1(y_i \leq \mu_p - \delta)\} - \bar{d}_j F_j(\mu_p - \delta) = 0$ . A similar argument works for the other terms in (19). ■



# Inferring Optimal Resource Allocation from Experimental Data: The Application

Debopam Bhattacharya

Dartmouth College

10th March, 2007.

Here I report the application of the methods developed in the paper "Inferring Optimal Allocations from Experimental Data" to a dataset of Dartmouth College undergraduates. Owing to confidentiality requirements, these data could not be made publicly available. Accordingly, at the suggestion of the editor, this application had to be removed from the main paper to avoid conflict with the data availability policy of the present journal and is being rewritten into a separate paper. It now appears below in its present form, together with the tables.

## 1 Application

I apply the 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  $Z$ . 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  $\Theta_n$  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. 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 (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. 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.

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. 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.

## 2 Appendix

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

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

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

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

Then

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

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

$$\begin{aligned} &p_{HH}E(Y|HH) + p_{HL}E(Y|HL) + p_{LL}E(Y|LL) \\ &= p_{HH}\gamma_1(HH) + p_{HL}\gamma_1(HL) + p_{LL}\gamma_1(LL) \\ &\quad + [\gamma_2(SS) \Pr(SS) + \gamma_2(SN) \Pr(SN) + \gamma_2(NN) \Pr(NN)] \end{aligned} \quad (1)$$

and the term within  $[\cdot]$  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

$$\begin{aligned}
& p_{HH}E^*(Y|HH) + p_{HL}E^*(Y|HL) + p_{LL}E^*(Y|LL) \\
= & p_{HH}\gamma_1(HH) + p_{HL}\gamma_1(HL) + p_{LL}\gamma_1(LL) \\
& + [\gamma_2(SS)\Pr^*(SS) + \gamma_2(NN)\Pr^*(NN)] . \tag{2}
\end{aligned}$$

Comparing (1) and (2), 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  $\phi_s$  then the second term in (1) 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 (2) 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 Characteristics**

Variable	Mean	SD	Range
<b>Women (N=428)</b>			
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
<b>Men (N=436)</b>			
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
Men (mean=0.53)							
2	0.07*	0.571	2	0.003 0.03 0.3	1 1 2	0.568, 0.574 0.568, 0.575	0.5681
3	0.68	0.56	8	.	.		
4	0.215	0.626	40				
Women (mean=0.47)							
2	0.053*	0.513	2	0.003 0.03 0.3	1 1 2	0.509, 0.515 0.509, 0.515	0.510
3	0.029*	0.54	8	0.003 0.03 0.3	1 2 8	0.534, 0.543	0.5351 0.5347
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 based on normal approximation. When the cardinality exceeds 1, column 7 contains the lower bounds corresponding to the conservative (based on corollary 1) 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)	(1)	(2)	(3)
# Categories	p-value	HH and LL: Max	HH and LL: Min
Men (mean=0.53)			
2	0.07*	0.51, 0.49	0.037, 0.00
Women (mean=0.47)			
2	0.05*	0.037, 0	0.52, 0.48
3	0.04*	0.00, 0.00	0.11, 0.28

Notes: This table shows the nature of allocations which maximize and minimize the probabilities of joining a frat/sorority corresponding to the 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	
Men (mean=0.53)	2	0.06	0.553	0.003	1	0.5505, 0.5557	.	
				0.03	1		0.5505, 0.5557	.
				0.3	2			0.5503
Women (mean=0.47)	2	0.03	0.479	0.003	1	0.476, 0.482	.	
				0.03	1		0.476, 0.482	.
				0.3	2			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 based on normal approximation. When the cardinality exceeds 1, column 7 contains the lower bounds corresponding to the conservative (based on corollary 1) one-sided CI.

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

(0)	(1) p-value	(2) HL: Max	(3) 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 minority 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 10<sup>th</sup>, 50<sup>th</sup> and 90 percentiles. Column 2 for men, for instance, shows that the maximized value of the 10<sup>th</sup> 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.