Inference on Optimal Matching from Experimental Data

Debopam Bhattacharya,[*][†]

Dartmouth College.

First draft: February 15, 2006.

This draft: May 8, 2006.

Abstract

This paper analyzes nonparametric identification of optimal allocations in input matching problems and derives asymptotic distribution of resulting estimates under both point and set-identified situations. The key tools in the analysis are (i) the fundamental theorem of linear programming which makes the relevant parameter space discrete and (ii) Cramer's theorem for large deviations which implies that sampling error in the estimation of optimal *solutions* does not affect asymptotic distribution of the optimized *value.* This latter distribution is zero-mean normal when the solution is point-identified but both asymptotically biased and non-normal when the solution is set-identified. The

[*]Address for correspondence: 327, Rockefeller Hall, Dartmouth College, Hanover, NH 03755. e-mail: debopam.bhattacharya@dartmouth.edu.

[†]This research was motivated by the work of Graham, Imbens and Ridder (2005) on alternative matching procedures presented at the econometrics workshop at Harvard-MIT in the Fall of 2005. I am grateful to Geert Ridder for his encouragement at the early stages of this project. I also thank Guido Imbens and seminar participants at Dartmouth, MIT, Princeton and UBC for helpful comments and suggestions. Susan Schwarz provided invaluable help with computing. All errors are mine.

bias can be analytically corrected to produce confidence intervals for the maximum value with uniformly good coverage. A key contribution of this paper is to show that applicability of these techniques extends beyond linear maximands like the mean to quantiles of the outcome distribution which are *nonlinear* in the allocation probabilities. This work differs from the "treatment choice" literature as it concerns the efficient way to simultaneously allocate a large group of individuals under aggregate resource constraints and focuses on the asymptotic analysis of estimated optimal solutions and values for a fixed decision rule. I illustrate the methods using data from Dartmouth's random assignment of freshmen to dorm rooms, where Sacerdote (2001) detected significant contextual peer effects. Segregation by previous academic achievement and by race are seen to minimize mean enrolment into sororities and maximize mean enrolment into fraternities. Segregation appears to have no effect on mean and median freshman year GPA but increases the higher and decreases the lower percentiles of the GPA distribution for both men and women.

1 Introduction

Input-matching problems have recently been introduced in the econometrics literature by Graham, Imbens and Ridder (2005). In the present paper, I use insights from mathematical programming theory to analyze nonparametric identification of and asymptotic inference on optimal allocations in such problems. I show that when the objective of input-matching 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. This makes the computation easy and the asymptotic analysis nonstandard but elegant. I establish necessary and sufficient conditions for uniqueness of the population solution which are testable. Further, using Cramer's theorem for large deviations, I show that sampling error in estimating the optimal solution(s) has no effect asymptotically on the distribution of the estimated maximum value. This latter distribution is asymptotically zero-mean normal under uniqueness but under nonuniqueness it is both asymptotically biased and non-normal. These findings are combined to propose a valid method of inference on the maximum value.[1]

A further 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 in itself is a nonstandard exercise in that it cannot be interpreted as an M (or Z)-estimation problem, unlike e.g. quantile regressions. To the best of my knowledge, analysis of identification and estimation in quantile maximization problems has not been attempted before[2] and the current paper addresses these issues in the context of input matching, where quantile maximization is potentially an important policy objective.

To summarize, this paper makes four methodological contributions- (i) showing that optimal allocation problems have mathematical programming structures, (ii) using theoretical insights from the LP literature to characterize nonparametric identification of the population solution, (iii) using large deviations theory to derive asymptotic properties of plug-in type estimates in both the point and set-identified cases and (iv) showing that these results can be extended to the case where one is maximizing a quantile even if the quantile maximization problem is nonlinear. The paper also describes what are the general class of policy maximands to which these ideas can be applied.

The methodology presented here is relevant to many constrained allocation problems (see GIR (2005) for examples). But for expositional clarity this paper focusses

[1] Although the derivations of these asymptotic distributions rely on Cramer's theorem, they are not straightforward applications of it since the underlying estimates are *ratios* of sample averages and not simply sample averages.

[2] 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.

on the problem of choosing allocations to maximize mean and median outcome when peer-effects are important determinants of output, as elaborated below.

Analysis of peer effects has played an important role in both economic theory and econometrics. Conceptual issues involved in defining what constitutes "peer effects" and in the identification of these effects from nonrandomized field data were pioneered by Manski (1995). Subsequent research has addressed various methods of identifying peer effects from experimental (c.f. Sacerdote (2001)) as well as observational studies (c.f. Cooley (2006)) and tested the applicability of alternative models of peer interaction (c.f. Hoxby et al (2006)). This paper addresses a complementary question- namely, given the evidence on the magnitude and nature of peer effects, what is the socially "optimal" way for an outside planner to divide individuals into peer-groups? Obviously, the optimal grouping depends on what social criterion is to be optimized. But given a social welfare function, whether alternative grouping can affect aggregate social outcome will also depend on how peer effects interact with own effect in producing individual outcomes. Here, by "peer effects", I will mean "contextual effects" in Manski's terminology- the effect of peers' background on own outcome, controlling for own background.

To fix ideas, let us begin with the following allocation problem. 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 (section 4 below analyzes this case in details).

Assume that every dorm room can accommodate two students and the college can assign individuals to dorms based on an index of their previous academic achievement, say, SAT scores. For simplicity, assume that SAT score can take 3 distinct values- low, medium and high abbreviated by l,m,h. Denote the expected total score of a dorm room with each of 6 types of couples, denoted by $\mathbf{g} = (g_{hh}, g_{mm}, g_{ll}, g_{hl}, g_{hm}, g_{ml})'$. For instance, g_{ml} is the mean *per person* GPA score across all rooms which have one m-type and one l-type student. Also, denote the marginal distribution of SAT score for the current class by $\boldsymbol{\pi} = \pi_l, \pi_m, \pi_h$. Then an allocation is a vector $\mathbf{p} = (p_{hh}, p_{mm}, p_{ll}, p_{hl}, p_{hm}, p_{ml})'$, satisfying $p_{ij} \geq 0$ and $\mathbf{p'1} = 1$. Here p_{ij} (which equals p_{ji} by definition) denotes the fraction of dorm rooms that have one student of type i and one of type j, with $i, j \in \{h, m, l\}$. Then the authority's problem is defined by the following LP problem.

$$\max_{\{p_{ij}\}} \left[g_{hh}p_{hh} + g_{mm}p_{mm} + g_{ll}p_{ll} + g_{hl}p_{hl} + g_{hm}p_{hm} + g_{ml}p_{ml}\right]$$

s.t.

$$\begin{aligned} 2p_{hh} + p_{hl} + p_{hm} &= 2\pi_h \\ 2p_{mm} + p_{hm} + p_{ml} &= 2\pi_m \\ 2p_{ll} + p_{hl} + p_{ml} &= 2\pi_l = 2\left(1 - \pi_h - \pi_m\right) \\ p_{ij} &\geq 0,\ i, j \in \{h, m, l\}\,. \end{aligned}$$

The first set of linear constraints are just stating the budget constraint. For example, the first linear constraint simply says that the total number of students of h type in the dorm rooms (in every hh type room there are two h type students and hence the multiplier 2 appear before p_{hh}) should equal the total number of h type students that year. The first of these quantities is $N/2 \times (2p_{hh} + p_{hl} + p_{hm})$ if there are N students and hence $N/2$ dorm rooms. The second is $N \times \pi_h$. 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 h and one l type to estimate $\hat{g}_{hl}$ which will be a good estimate of the unknown g_{hl} if the sample size is large. This assumes that peer interaction is unaffected by whether allocations are made through general randomization (as with the pilot sample) or by randomization within covariate categories (as will be done by the planner). This assumption can fail to hold if, 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. It is also assumed that nature of interactions are unaffected by the aggregate proportions of H, M and L types, which can fail to hold if for instance, high types are more accommodating of low types if low types are a small minority but more antagonistic if high types are a minority.[3]

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

$$\max_{\mathbf{p}} \hat{\mathbf{g}}'\mathbf{p} \text{ s.t. } A\mathbf{p} = \boldsymbol{\pi},\ \mathbf{p} \geq \mathbf{0} \tag{1}$$

where

$$\begin{aligned}
\hat{\mathbf{g}} &= \left(\hat{g}_{hh}, \hat{g}_{mm}, \hat{g}_{ll}, \hat{g}_{hl}, \hat{g}_{hm}, \hat{g}_{ml}\right)' \\
\mathbf{p} &= \left(p_{hh}, p_{mm}, p_{ll}, p_{hl}, p_{hm}, p_{ml}\right)' \\
A &= \begin{pmatrix} 2 & 0 & 0 & 1 & 1 & 0 \\ 0 & 2 & 0 & 0 & 1 & 1 \\ 0 & 0 & 2 & 1 & 0 & 1 \end{pmatrix} \\
\boldsymbol{\pi} &= \left(\pi_h, \pi_m, \pi_l\right)'.
\end{aligned}$$

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

[3] I am grateful to Prof David Green of UBC for pointing out the second caveat.

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. Defining

$$\mu_{\tau,p} = \inf_{\mu} \left(\sum_{i=h,m,l} \sum_{j} p_{ij} F_{ij}(\mu) \geq \tau \right)$$

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

$$\max_{\mathbf{p}} \mu_{\tau,p} \text{ s.t. } A\mathbf{p} = \boldsymbol{\pi},\ \mathbf{p} \geq \mathbf{0}. \tag{2}$$

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

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

Typically, a planner would be interested in a "good" estimate of the optimal allocation, i.e the maximizer of (1) and (2) since that is the policy parameter. But the planner can also be interested in the respective maximum values because this gives her a benchmark against which to compare different rules in terms of their relative efficiencies. In particular, if there are political restrictions on which covariates can be used for room assignment, then one can evaluate the efficiency cost of such "non-discrimination" policies when one knows the maximum value attainable without the restrictive policy. The purpose of this paper is to analyze the population problem and derive statistical properties of the estimated optimal solutions when the size of the pilot sample is large.

Several other problems have similar structure. For instance, consider a scenario where one observes student outcomes when teachers were randomly assigned to classes as in the well-known Project Star experiment. If one observes (mean) background characteristics of each class and characteristics of the teacher, one can estimate the mean outcome of each "type" of match. Using these estimates, one can devise the matching rule between teacher and class characteristics that should produce the largest expected test-score. The problem reduces mathematically to a finite LP problem where the RHS of the constraints now correspond to marginal distributions of teacher characteristics and class characteristics, respectively.

Randomized field-experiments are becoming more prevalent in the study of economic phenomena (c.f. Duflo (2005) and references therein). The methods presented in this paper shed light on how to use the results of these trials to design effective policies.

The plan of the paper is as follows. Section 2 describes contributions of this paper in relation to existing work. Section 3 describes the general problem for the mean and reviews the relevant results from LP theory. Section 4 deals with the degenerate case and section 5 with the general non-degenerate case. Section 6 analyzes the case where the maximand is a quantile rather than the mean. Section 7 discusses the application and section 8 concludes.

2 Relation to existing literature

The broad context for the work presented here, viz. input-matching problems, was first laid out by GIR (2005) who compare a fixed set of matching rules (like positive and negative assortative matching) to see which one yields the higher expected payoff. In that paper, they do not consider optimal allocations. Having a method for calculating the optima, as presented in the current paper, also enables one to compute the efficiency of a given allocation rule by comparing it to the maximum attainable. This is discussed in the remarks at the end of section 5 of this paper. In contrast to GIR (2005), however, this paper is concerned exclusively with discrete covariates. Since most real-life allocation problems will concern discrete covariates and continuous ones can be converted to discrete ones, this is not a large sacrifice in generality.[4]

[4] The nonstochastic version of mean based optimization in the continuous case can be related to the Monge-Kantorovich mass transportation problem which is well-known to be analytically challenging.

In independent work, GIR (2006) consider mean-based optimality from a decision theoretic standpoint when covariates are binary. In contrast, the current paper focuses on one specific albeit natural plug-in type decision rule for a general discrete valued covariate and, perhaps more importantly, considers the maximization of both means and quantiles. GIR take identification of the population optima as given while the current paper investigates point and set-valued identification and constructs statistical tests to distinguish between these cases. In fact, when the inputs are binary as in GIR (2006), it can be shown that either the population solution is unique or else all feasible allocations are optimal. For general discrete valued covariates, as discussed in the current paper, one can have nontrivial set-identified situations. This is further elaborated below in section 3 and remark 1. The method of analysis in the current paper is also fundamentally different from GIR (2006) in that it uses a unified programming theory based approach to identification and estimation of the optimal allocation. But unlike GIR (2006), this paper does not investigate decision theoretic issues associated with the problem. Thus the two papers are complementary to each other in terms of both objectives and methodology.

The work presented here is also related to the 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 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. Moreover, presence of the constraints makes the problem analytically different and ties the analysis to mathematical programming theory and large deviations principles. 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. However, in remark 5 below I briefly address the issue of choosing covariates.[5] 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.

3 LP preliminaries and uniqueness

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

In general, assume that there are M possible points of support of the covariate of interest and therefore a total of $m = M(M+1)/2$ possible types of room, indexed by the pair (j,k), $j = 1,..M$, $k = j,...M$. Let the vector of conditional mean of GPA obtained from a random assignment of the entire population be denoted by $g = (g_{jk})_{j=1,..M,k=j,...M}$. Let the proportion of incoming individuals (who are to be assigned to rooms) with value of covariate equal to k be denoted by π_k, $k = 1,...M$. Then the planner's problem, if she knew g, is referred to as the "population" problem and is given by

$$\begin{aligned} & \max_{\mathbf{p}=(p_{jk})_{j=1,..M,k=j,...M}} \mathbf{p}'g \\ & \text{s.t.} \\ 2p_{ll} + \sum_{j=1}^{l-1} p_{jl} + \sum_{k=l+1}^{M} p_{lk} &= 2\pi_l, \quad l = 1,...M \qquad (3) \\ p_{jk} &\geq 0, \; j = 1,..M, k = j,...M. \end{aligned}$$

The constraint set will be denoted by $\mathcal{P}$. Typically, the π's will be known and g's will not be known. A random sample is assumed to have been drawn from the same

[5] 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.

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 $\hat{g} = (\hat{g}_{jk}),\ j = 1,..M, k = j,...M$. The planner solves the problem in the previous display with g replaced by $\hat{g}$, which I will call the "sample problem". Whether this is the "optimal" action by the planner in the decision theoretic sense is an interesting and relevant question but is outside the scope of the current paper, which focusses on the asymptotic properties of *a* specific but "natural" action. For the asymptotic analysis of the decision theoretic optima in the treatment choice case see Hirano and Porter (2005). Note in passing that in my formulation, the constraint set $\mathcal{P}$ for the sample problem is identical to that of the population problem. I now review the idea of extreme points and their relation to solutions of LP problems. These ideas are crucial for my analysis. I will use the terminology of Luenberger (1984) for setting them out. Also, I will re-index the components of the g and corresponding p vectors by $k = 1, ..., m$.

Notice that the constraint set $\mathcal{P}$, defined by (3), is a convex polytope with extreme points corresponding to the set of basic solutions (e.g. Luenberger (1984) page 19). These extreme points or basic solutions are obtained by taking the independent columns of the constraint matrix, inverting the resultant matrix, post-multiplying by the vector $\boldsymbol{\pi}$ and adding zeros corresponding to the deleted columns. For the $M \times M(M+1)/2$ constraint matrix defined in (3), let $S = \left\{z_1, ...z_{|S|}\right\}$ denote the set of basic feasible solutions.[6] Then the fundamental theorem of LP implies that

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

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

[6] Clearly $|S| \leq \begin{pmatrix} M(M+1)/2 \\ M \end{pmatrix}$.

can simply concentrate on the countably finite set S rather than the much larger constraint set $\mathcal{P}$.

3.1 Uniqueness

There are three possible scenarios: (i) the "degenerate" case: $z_1'g = z_2'g = ... = z_{|S|}'g$, (ii) the "intermediate" case: $\exists\, z_1 \neq z_2 \neq z_3 \in \mathcal{P}$ such that $z_1, z_2 \in \{z^* \in S : z^{*\prime}g = \max_{z \in S}(g'z) \equiv v\}$ and $g'z_1 = g'z_2 = v > g'z_3$ and (iii) the "unique" case: $g\prime z_1 > g'z_j$ for all $j \neq 1$. It should be noted that uniqueness refers to the solution and not the maximum value since the latter will be unique if it is finite. I will deal separately with cases (i), (ii) and (iii) since the asymptotic distributions of the sample maximum value will be qualitatively different in these cases. In section 5.3, I will consider combining these results to construct confidence intervals which will have good coverage no matter which of these 3 cases is true.

Given that $\mathcal{P} = conv(S)$ and since $g'p$ is a linear function of p, case (i) above is equivalent to the entire set $\mathcal{P}$ being optimal. Case (ii) is equivalent to there existing vectors $p \neq q \neq r \in \mathcal{P}$ such that $p, q \in \{p^* : p^{*\prime}g = \max_{p \in \mathcal{P}}(g'p)\}$ and $g'p = g'q > g'r$. And case (iii) is equivalent to z_1 being the only solution in $\mathcal{P}$. Indeed, if p is a solution and p is not an extreme point, then $p = \lambda z_j + (1-\lambda) z_k$ for some $z_j, z_k \in S$ and $\lambda \in (0,1)$. But then we must have that both $g'z_j$ and $g'z_k$ are equal to $g'p$. So if there is a unique solution, it has to be at an extreme point. Moreover, if there are at least two distinct solutions, then there are at least two distinct extreme points which are solutions.

By way of illustration I now describe some scenarios where one can have situations (i), (ii) and (ii) using the example of the introduction. Case (i) will arise if for instance $g_{hh} = g_{hm} = g_{hl} = g_{mm} = g_{ml} = g_{ll}$. Case (ii) will arise if e.g. $g_{hh} = g_{hm} = g_{mm} = a > b = g_{hl} = g_{ml} = g_{ll}$. In that case, the value of the objective function will be $a - (a-b)\pi_l - (p_{hl} + p_{ml})(a-b)/2$ and so any allocation which gives nonzero weight to one or more of p_{hl}, p_{ml} will yield a strictly lower objective function than one which sets p_{hl}, p_{ml} equal to zero. Moreover, any allocation with $p_{hl} = p_{ml} = 0$, $p_{ll} = \pi_l$ and satisfies $2p_{hh} + p_{hm} = 2\pi_h$ and $2p_{mm} + p_{hm} = 2\pi_m$ with $p_{hh}, p_{hm}, p_{mm} \geq 0$ will be optimal and there will be at least two of them, viz. $p_{hh} = \pi_h$, $p_{mm} = \pi_m$, $p_{hm} = 0$ and $p_{hm} = \min\{\pi_h, \pi_m\}$, $p_{hh} = \frac{1}{2}(2\pi_h - p_{hm})$ and $p_{mm} = \frac{1}{2}(2\pi_m - p_{hm})$. Finally, case (iii) will arise if e.g. $g_{hm} < \min\{g_{hh}, g_{mm}\}$, $g_{hl} < \min\{g_{hh}, g_{ll}\}$ and $g_{ml} < \{g_{ll}, g_{mm}\}$.

Remark 1 *It turns out that when $M = 2$, one can have either situation (i) or situation (iii) but not situation (ii). To see this, denote the two categories as h and l with all notations analogous to the example above. Then the LP problem is $\max\{g_{ll}p_{ll} + g_{lh}p_{lh} + g_{hh}p_{hh}\}$ s.t. $2p_{ll} + p_{hl} = 2\pi_l$ and $2p_{hh} + p_{hl} = 2\pi_h$ and $p_{ll}, p_{hl}, p_{hh} \geq 0$. Solving out, the objective function equals $g_{ll}\pi_l + g_{hh}\pi_h - p_{hl}(g_{hh} + g_{ll} - 2g_{hl})/2$. So when $g_{hh} + g_{ll} - 2g_{hl}$ 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 non-trivial set-identified situation, of the type discussed in section 5 of this paper, will not arise there.*

Additional notation

For future use, rewrite

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

where D_{ij} is a dummy which equals 1 if the ith sampled individual is in room type j, $i = 1, ...n$ and $j = 1, ... |S|$. The expectation terms in the above display correspond to the combined experiment of drawing one random sample and making one random allocation of this drawn sample.

For the rest of this paper, I will assume that $\delta_j >> \bar{m}$ for all j. Then $\hat{g}$ is consistent for g and $\sqrt{n}(\hat{g} - g)$ 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.

A consistent estimate of the (j,k)th element of Σ is given by $\frac{1}{n}\sum_{i=1}^{n}s_{ij}s_{ik}$ where

$$s_{ij}=\frac{D_{ij}y_i}{\frac{1}{n}\sum_{i=1}^{n}D_{ij}}-\frac{\frac{1}{n}\sum_{i=1}^{n}D_{ij}y_i}{\left(\frac{1}{n}\sum_{i=1}^{n}D_{ij}\right)^2}\times D_{ij}.$$

4 Degenerate case

I now state and prove three propositions related to case (i). The first proposition states a necessary and sufficient condition for the degenerate case which can be tested, the second proposition sets out the asymptotic distribution theory for the statistic used to test degeneracy and the third discusses the asymptotic behavior of the sample maximum value under case (i).

Degeneracy, i.e. case (i) above comes from what might be called "additivity" of peer effects. Additive peer effects means that for some functions $\gamma_1(.)$ and $\gamma_2(.)$,

$$E(y|OB,RB)=\gamma_1(OB)+\gamma_2(RB) \tag{4}$$

where y is the individual outcome, OB and RB are one's own background covariate value and one's roommate's covariate value respectively. Additivity implies that all allocations would yield the same average effect overall and every feasible allocation will solve problem (3). However, the maximum (or minimum) value will still be unique. In other words, if our parameter of interest is the maximum value, then that parameter is a well-defined functional of g (given A,π) in that for every g, the mapping gives one number. However, the solution, i.e, the argmax or the argmin is not a well-defined function of g for every g in that it is a one-to-many mapping.

A natural statistic for testing additivity would be the difference between the sample maximum and the sample minimum. But the asymptotic distribution of these quantities is far from obvious under the null of additivity. Proposition 3 below will show that these distributions are not asymptotically centered at the population values. However, the following proposition shows that it is possible to bypass these complications altogether by establishing that additivity and degeneracy are equivalent and they are also equivalent to an easily testable rank condition. While it is easy to see that additivity implies degeneracy, the converse is probably less obvious.

Proposition 1 *Let* $A_{M\times\frac{M(M+1)}{2}}$ *be the matrix of constraints with* $M = rank\,(A)$ *and let* $B' = \left(\begin{array}{cc} g_{\frac{M(M+1)}{2}\times 1} & A'_{\frac{M(M+1)}{2}\times M} \end{array} \right)$. *Then the following statements are equivalent (i) (4) holds, (ii)* $rank\,(B) = M$ *and (iii)* $\min_{p\in\mathcal{P}}(g'p) = \max_{p\in\mathcal{P}}(g'p)$.

Proof. I will show that (i) implies (iii), (iii) implies (ii) and (ii) implies (i).

(i) implies (iii): Let wlog, let the points of support of the covariate be $1, 2, ...M$ with aggregate probabilities $\pi_1, ...\pi_M$. Then

$$\begin{aligned} g_1 &= E\,(y|OB = 1, RB = 1) = 2\,(\gamma_1\,(1) + \gamma_2\,(1)) \\ g_2 &= E\,(y|OB = 1, RB = 2) = (\gamma_1\,(1) + \gamma_2\,(2) + \gamma_1\,(2) + \gamma_2\,(1)) \\ & .... \\ g_M &= E\,(y|OB = 1, RB = M) = (\gamma_1\,(1) + \gamma_2\,(M) + \gamma_1\,(M) + \gamma_2\,(1)) \\ & ..... \\ g_{\frac{M(M+1)}{2}} &= E\,(y|OB = M, RB = M) = 2\,(\gamma_1\,(M) + \gamma_2\,(M))\,. \end{aligned}$$

Then for any $p_1, ...p_{\frac{M(M+1)}{2}}$, it can be seen by writing out that

$$\begin{aligned} & g_1p_1 + g_2p_2 + ... + g_{\frac{M(M+1)}{2}}\,p_{\frac{M(M+1)}{2}} \\ = \; & (\gamma_1\,(1) + \gamma_2\,(1)) \times 2\pi_1 + (\gamma_1\,(2) + \gamma_2\,(2)) \times 2\pi_2 + ...\,(\gamma_1\,(M) + \gamma_2\,(M)) \times 2\pi_M \end{aligned}$$

which does not depend on $p_1, ...p_{\frac{M(M+1)}{2}}$. This shows that (i)$\Longrightarrow$(iii).

(iii) implies (ii): To show that (iii) implies (ii), we will show that "not (ii)" implies "not (iii)". Let $m = M\,(M + 1)\,/2$. "Not (ii)" implies that the null space of B has strictly smaller dimension than the null space of A, since they have the same number of columns and rank of B is one more than the rank of A. So there exists δ such that $A\delta = 0$ and $g'\delta \neq 0$. Let $p \in \mathcal{P}$. We will construct a $q \in \mathcal{P}$ such that $g'q \neq g'p$. Now, for some $c \neq 0$, we will have $q = p + c\delta > 0$ (in particular, choose $c \geq \max\{-p_1/\delta_1, ... - p_m/\delta_m\}$. For this c, we have that $q > 0$, $Aq = Ap + cA\delta = \pi$ and $g'q = g'p + cg'\delta \neq g'p$ since $g'\delta \neq 0$. This contradicts (iii). This last proof assumes that $\mathcal{P}$ is not a singleton, which is easy to check.

(ii) implies (i): (ii) implies that g is a linear combination of the rows of A, say, $g = \sum_{j=1}^{M} \beta_j a_j$ where $a_1, ...a_M$ are the rows of A. Thus

$$\begin{aligned} g_1 &= g(1,1) = 2\beta_1, \\ g_2 &= g(1,2) = \beta_1 + \beta_2, \\ &\dots \\ g_M &= g(1,M) = \beta_1 + \beta_M \end{aligned}$$

and so on. Therefore, $g(i,j) = \beta_i + \beta_j$ which shows that (ii) implies (i) (equate β_i with $\gamma_1(i) + \gamma_2(i)$ in the previous notation). ■

Thus, one can test for degeneracy by testing whether $rank(B) = M$. This test is a little different from existing tests for rank of a matrix (e.g. Cragg and Donald (1997)) in that the matrix B has only one row- the first one corresponding to g- which is unknown and estimated. Our test is based on the following statistic

$$T_n = n \times \min_b (\hat{g} - A'b) \hat{\Sigma}^{-1} (\hat{g} - A'b),$$

where $\hat{\Sigma}$ is a consistent estimate of Σ. Under the null that rank of B is M, the value of T_n will be close to 0 and under the alternative of B being full-rank, its value should diverge to infinity. The next proposition describes the asymptotic distribution of T_n.

Proposition 2 *Assume that* $\sqrt{n}(\hat{g} - g) \xrightarrow{d} N(0, \Sigma)$*. Then under the null hypothesis that* $rank(B) = M$*,* $T_n \xrightarrow{d} \varkappa^2$ *with* $df = \frac{M(M+1)}{2} - M$*. Under the alternative, i.e.* $rank(B) = M + 1$*,* $T_n \xrightarrow{P} +\infty$*. So the test is consistent.*

Proof. By usual first order conditions, it is easy to check that

$$\arg\min_b (\hat{g} - A'b) \hat{\Sigma}^{-1} (\hat{g} - A'b) = \left[A\hat{\Sigma}^{-1}A'\right]^{-1} A\hat{\Sigma}^{-1}\hat{g}$$

whence

$$T_n = n \times \hat{g}' \left[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}A' \left[A\hat{\Sigma}^{-1}A'\right]^{-1} A\hat{\Sigma}^{-1}\right] \hat{g} = n\hat{g}'\hat{V}\hat{g}$$

where

$$\hat{V} = \left[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}A' \left[A\hat{\Sigma}^{-1}A'\right]^{-1} A\hat{\Sigma}^{-1}\right].$$

Let

$$V = \left[\Sigma^{-1} - \Sigma^{-1}A'\left[A\Sigma^{-1}A'\right]^{-1}A\Sigma^{-1}\right].$$

Observe that we can rewrite

$$\begin{aligned} T_n &= n\left(\hat{g}-g\right)'V\left(\hat{g}-g\right) + 2ng'\hat{V}\hat{g} - ng'\hat{V}g \\ &\quad + n\left(\hat{g}-g\right)'\left(\hat{V}-V\right)\left(\hat{g}-g\right). \end{aligned} \tag{5}$$

Under the null, $g = A'\delta$ for some δ and hence

$$\begin{aligned} g'\hat{V} &= \delta' A\left[\hat{\Sigma}^{-1} - \hat{\Sigma}^{-1}A'\left[A\hat{\Sigma}^{-1}A'\right]^{-1}A\hat{\Sigma}^{-1}\right] \\ &= \delta'\left[A\hat{\Sigma}^{-1} - A\hat{\Sigma}^{-1}A'\left[A\hat{\Sigma}^{-1}A'\right]^{-1}A\hat{\Sigma}^{-1}\right] \\ &= 0. \end{aligned}$$

Further, since $\hat{V} - V \xrightarrow{P} 0$ and $\sqrt{n}\left(\hat{g}-g\right) = O_p\left(1\right)$, it follows that

$$n\left(\hat{g}-g\right)'\left(\hat{V}-V\right)\left(\hat{g}-g\right) = o_p\left(1\right),$$

whence

$$T_n = n\left(\hat{g}-g\right)'V\left(\hat{g}-g\right) + o_p\left(1\right).$$

Using the breakup $\Sigma^{-1} = P'P$ with $PP' = I$, we have that under the null,

$$T_n = \left(\sqrt{n}P\left(\hat{g}-g\right)\right)'\left(\left[I - PA'\left[AP'PA'\right]^{-1}AP'\right]\right)\sqrt{n}P\left(\hat{g}-g\right).$$

Since $\sqrt{n}P\left(\hat{g}-g\right) \xrightarrow{d} N\left(0, I_{\frac{M(M+1)}{2}}\right)$ and $\left[I - PA'\left[AP'PA'\right]^{-1}AP'\right]$ is idempotent, we have by standard arguments that $T_n \xrightarrow{d} \varkappa^2$ with df equal to

$$\begin{aligned} & trace\left[I - PA'\left[AP'PA'\right]^{-1}AP'\right] \\ &= \frac{M(M+1)}{2} - tr\left(PA'\left[AP'PA'\right]^{-1}AP'\right) \\ &= \frac{M(M+1)}{2} - tr\left(PA'\left[AP'PA'\right]^{-1}AP'\right) \\ &= \frac{M(M+1)}{2} - tr\left(\left[AP'PA'\right]^{-1}AP'PA'\right) \\ &= \frac{M(M+1)}{2} - tr\left(I_M\right) = \frac{M(M+1)}{2} - M. \end{aligned}$$

Under the alternative,

$$g'\hat{V}g = \min_b (g - A'b)' \hat{\Sigma}^{-1} (g - A'b) > 0$$

w.p. 1. So under the alternative, continuing from (5),

$$\begin{aligned} T_n &= n(\hat{g} - g)' V (\hat{g} - g) + 2\sqrt{n} g'\hat{V}\sqrt{n}(\hat{g} - g) + ng'\hat{V}g \\ &\quad + n(\hat{g} - g)' \left(\hat{V} - V\right)(\hat{g} - g), \end{aligned}$$

which, for large enough n, will be dominated by $ng'\hat{V}g \to +\infty$, since $\sqrt{n}(\hat{g} - g) = O_p(1)$, $n(\hat{g} - g)' \left(\hat{V} - V\right)(\hat{g} - g) = o_p(1)$ and $n(\hat{g} - g)' V (\hat{g} - g) = O_p(1)$. ■

The asymptotic distribution of the sample maximum value under nonunique population solution is now described in the following proposition. Notice that in this degenerate case, all feasible solutions i.e. the entire set $\mathcal{P}$ are optimal and so the sample solution will lie within this set with probability 1. Conversely one would include all of $\mathcal{P}$ in the set of sample maximizers in order to make the Hausdorff distance between the sample solution and the population solution set converge to 0. I now turn to the distribution of the sample maximum value which is more interesting.

Proposition 3 *If $\hat{g} - g \xrightarrow{P} 0$, then $\hat{v}$ is consistent for v. Further, if $\sqrt{n}(\hat{g} - g) \xrightarrow{d} R \equiv N(0, \Sigma)$ and $\min_{p\in\mathcal{P}}(g'p) = \max_{p\in\mathcal{P}}(g'p) = v$, then*

$$\sqrt{n}(\hat{v} - v) \xrightarrow{d} W^{\max} = \max\left\{W_1, ...W_{|S|}\right\}$$

where $W_j = z_j'R$ for $j = 1, ...|S|$. $W^{\max}$ is not Gaussian in general.

Proof. Under degeneracy we have $v = z_1'g = ... = z_{|S|}'$ and continue to have that for all finite n,

$$\hat{v} = \sup_{p\in\mathcal{P}} p'\hat{g} = \max\left\{z_1'\hat{g}, ...z_{|S|}'\hat{g}\right\}.$$

Since $plim_{n\to\infty}\hat{g} = g$, $\max\left\{z_1'\hat{g}, ...z_{|S|}'\hat{g}\right\} \xrightarrow{P} \max\left\{z_1'g, ...z_{|S|}'g\right\} = v$, by the continuous mapping theorem. So $\hat{v}$ is consistent for v. Moreover,

$$\sqrt{n}(\hat{v} - v) = \max\left\{\sqrt{n}(z_1'\hat{g} - v), ...\sqrt{n}\left(z_{|S|}'\hat{g} - v\right)\right\} = \max\left\{z_1'\sqrt{n}(\hat{g} - g), ...z_{|S|}'\sqrt{n}(\hat{g} - g)\right\},$$

whence the conclusion follows by the continuous mapping theorem. ■

Although $W^{\max}$ does not have a pivotal distribution, its distribution can be simulated by first drawing a w from the (estimated) asymptotic distribution of $\sqrt{n}\,(\hat{g}-g)$, i.e. $N\left(0,\hat{\Sigma}\right)$, calculate the extreme points of the constraints set, viz. S and then calculate $\max\left\{z_1'w,...z_{|S|}'w\right\}$. Repeating this a large number of times should simulate the distribution of $W^{\max}$. However, this process can be very time-consuming if M is moderately large, implying that $|S|$ is very large. The following trick helps us reduce computation significantly. Solve problem (3) after replacing g by a draw from the (estimated) asymptotic distribution of $\sqrt{n}\,(\hat{g}-g)$, i.e. $N\left(0,\hat{\Sigma}\right)$. Repeat this a large number of times. Since the maxima will continue to be one of the extreme points, we will end up with the distribution of $W^{\max}$.

However, $E\left(W^{\max}\right)=\theta_0\neq 0$, in general and therefore a bias corrected estimate of v will be given by $\hat{v}_{BC}=\hat{v}-n^{-1/2}\tilde{\theta}_0$, where $\tilde{\theta}_0$ equals the mean of the simulated distribution of $W^{\max}$. To see why this bias arises, assume for simplicity that $|S|=2$, $z_1=(1,1)\,, z_2=(-1,-1)$. Let $\sqrt{n}\,(\hat{g}-g)\xrightarrow{d}(X_1,X_2)\simeq N_2\,(0,I)$. Then $\max\left\{z_1'\sqrt{n}\,(\hat{g}-g)\,,z_2'\sqrt{n}\,(\hat{g}-g)\right\}\xrightarrow{d}|X_1+X_2|$. But X_1+X_2 is a mean zero normal, so its absolute value has a strictly positive mean.

Thus a bias-corrected C.I. for v can be formed as follows. Choose d_L, d_H such that $1-\alpha=\Pr\left(d_L\leq W^{\max}\leq d_H\right)$. A level $(1-\alpha)$ confidence interval for v is then given by

$$CI_d=\left[\hat{v}-\frac{d_H}{\sqrt{n}},\,\hat{v}-\frac{d_L}{\sqrt{n}}\right]=\left[\hat{v}_{BC}-\frac{d_H-\tilde{\theta}_0}{\sqrt{n}},\,\hat{v}_{BC}+\frac{\tilde{\theta}_0-d_L}{\sqrt{n}}\right]. \tag{6}$$

Typically, $d_L<\theta_0<d_H$ implying that CI_d will be "centred around" $\hat{v}_{BC}$ (and not $\hat{v}$).

Remark 2 *It may be noted that when the solution is "Pitman-close" to being non-unique in the sense that* $\left\{z_1'g,...z_{|S|}'g\right\}=\left\{v+h_1/\sqrt{n},...v+h_{|S|}/\sqrt{n}\right\}$ *for some finite and fixed* $h_1,,,h_{|S|}$*, one gets (under le Cam type regularity conditions) that*

$$\begin{aligned}\sqrt{n}\,(\hat{v}-v) &= \max\left\{\sqrt{n}\,(z_1'\hat{g}-v)\,,...\sqrt{n}\left(z_{|S|}'\hat{g}-v\right)\right\}\\ &= \max\left\{z_1'\sqrt{n}\,(\hat{g}-g)-h_1,...z_{|S|}'\sqrt{n}\,(\hat{g}-g)-h_{|S|}\right\}\end{aligned}$$

which will still be $O_p(1)$, non-normal and asymptotically not centered at 0.

5 Non-degenerate case

I now turn to the non-degenerate cases (ii) and (iii). I will first consider asymptotic behavior of the maximum sample value $\hat{v}$ and then turn to the asymptotic properties of the solution.

5.1 Maximum value

Consistency of $\hat{v}$ for v can be established exactly as in proposition 3, above. To establish asymptotic distribution of $\hat{v}$, I will use the following lemma extensively. The proof of this lemma rests on Cramer's theorem for large deviations but is not completely straightforward because the $\hat{g}$'s are *ratios* of averages and not simply averages.

Lemma 1 *Assume that (i) $y_1, ...y_n$ and $D_{1j}, ...D_{nj}$ for each j are i.i.d. with finite mean, (ii) $\delta_j = E\left(\bar{d}_j\right) > \bar{m} > 0$ for all j and (iii) for each j, $E\left(e^{tw_{1j}}\right) < \infty$ for all $t \in \mathbb{R}$. If $z_1'g < z_2'g$, then $\Pr\left(z_1'\hat{g} > z_2'\hat{g}\right) \leq ce^{-\rho n}$ for some $c, \rho > 0$.*

Proof. Assume that $z_1'g < z_2'g$. Then

$$\begin{aligned}
&\Pr\left(z_1'\hat{g} > z_2'\hat{g}\right) = \Pr\left[(z_1 - z_2)'(\hat{g} - g) \geq (z_2 - z_1)' g\right] \\
&= \Pr\left[\sum_{j=1}^{m} (z_{1j} - z_{2j})(\hat{g}_j - g_j) \geq (z_2 - z_1)' g\right]
\end{aligned}$$

$$\begin{aligned}
&= \Pr\left[\sum_{j=1}^{m} (z_{1j} - z_{2j}) \left(\frac{\bar{y}_j - E(\bar{y}_j)}{\bar{d}_j} - \frac{E(\bar{y}_j) \times \left(\bar{d}_j - E\left(\bar{d}_j\right)\right)}{\bar{d}_j E\left(\bar{d}_j\right)}\right) \geq (z_2 - z_1)' g\right] \\
&\leq \Pr\left[\sum_{j=1}^{m} \frac{1}{\bar{d}_j} (z_{1j} - z_{2j})(\bar{w}_j - \omega_j) \geq (z_2 - z_1)' g,\ \bar{d}_1 > \bar{m}, ...\bar{d}_m > \bar{m}\right] \\
&\quad +1 - \Pr\left[\bar{d}_1 > \bar{m}, ...\bar{d}_m > \bar{m}\right]
\end{aligned}$$

The first probability

$$\begin{aligned}
&\leq \Pr\left[\left(\cup_{j=1}^{m}\left\{\frac{1}{\bar{d}_j}\left(z_{1j}-z_{2j}\right)\left(\bar{w}_j-\omega_j\right)\geq\frac{\left(z_2-z_1\right)'g}{m}\right\}\right)\cap\left(\bar{d}_1>\bar{m},...\bar{d}_m>\bar{m}\right)\right]\\
&= \Pr\left[\cup_{j=1}^{m}\left(\left\{\frac{1}{\bar{d}_j}\left(z_{1j}-z_{2j}\right)\left(\bar{w}_j-\omega_j\right)\geq\frac{\left(z_2-z_1\right)'g}{m}\right\}\cap\left(\bar{d}_1>\bar{m},...\bar{d}_m>\bar{m}\right)\right)\right]\\
&\leq \sum_{j=1}^{m}\Pr\left[\left\{\frac{1}{\bar{d}_j}\left(z_{1j}-z_{2j}\right)\left(\bar{w}_j-\omega_j\right)\geq\frac{\left(z_2-z_1\right)'g}{m}\right\}\cap\left(\bar{d}_1>\bar{m},...\bar{d}_m>\bar{m}\right)\right]\\
&\leq \sum_{j=1}^{m}\Pr\left[\left(z_{1j}-z_{2j}\right)\left(\bar{w}_j-\omega_j\right)\geq\frac{\bar{m}\left(z_2-z_1\right)'g}{m}\right]\\
&= \sum_{j=1}^{m}\Pr\left[\frac{1}{n}\sum_{i=1}^{n}\left(z_{1j}-z_{2j}\right)\left(w_{ij}-\omega_j\right)\geq\frac{\bar{m}\left(z_2-z_1\right)'g}{m}\right].
\end{aligned}$$

We now invoke the principle of large deviations. By condition (v) and Cramer's theorem (see, e.g. Hollander (2000), page 5) the first probability

$$\Pr\left[\frac{1}{n}\sum_{i=1}^{n}\left(z_{1j}-z_{2j}\right)\left(w_{ij}-\omega_j\right)\geq\frac{\bar{m}\left(z_2-z_1\right)'g}{m}\right]$$

is $O\left(e^{-\rho_1 n}\right)$ for some $\rho_1>0$ since $\left(z_2-z_1\right)'g>0$. The random variables D_{ij} are binary and so will satisfy Cramer's condition trivially. Therefore, the second probability

$$\begin{aligned}
&= 1-\Pr\left[\bar{d}_1>\bar{m},...\bar{d}_m>\bar{m}\right]\\
&= \Pr\left[\cup_{j=1}^{m}\left(\bar{d}_j<\bar{m}\right)\right]\leq\sum_{j=1}^{m}\Pr\left(\bar{d}_j<\bar{m}\right)=\sum_{j=1}^{m}\Pr\left(\bar{d}_j-\delta_j<\bar{m}-\delta_j\right)
\end{aligned}$$

is $O\left(e^{-\rho_2 n}\right)$ for some $\rho_2>0$ by Cramer's theorem since $\bar{m}-\delta_j<0$ for all j by condition (iii). Thus, we have shown that if $z_1'g<z_2'g$, then $\Pr\left(z_1'\hat{g}>z_2'\hat{g}\right)=O\left(e^{-\rho n}\right)$ for some $\rho>0$. ■

Now, I turn to the asymptotic distribution of the maximum value under case (ii). The distribution under case (iii) will be a special case of this. We have the following proposition.

Proposition 4 *Assume that every element of $\hat{g}$ is bounded with probability 1, $\sqrt{n}\left(\hat{g}-g\right)\xrightarrow{d}W\equiv N\left(0,\Sigma\right)$, $v=g'z_1=g'z_2=...g'z_J$ and $g'z_j<v$ for all $J<j\leq|S|$. Then*

$$\sqrt{n}\left(\hat{v}-v\right)=\max\left\{z_1'\sqrt{n}\left(\hat{g}-g\right),...,z_J'\sqrt{n}\left(\hat{g}-g\right)\right\}\times X_n+o_p\left(1\right),$$

where $X_n \xrightarrow{P} 1$, implying that $\hat{v}$ will, in general, have an asymptotically biased non-normal distribution as in proposition 3 above.

Proof. Define the event E as

$$E_n = \left\{ \max_{z \in \left\{ z_{J+1}, \ldots z_{|s|} \right\}} \hat{g}'z = \max_{z \in S} \hat{g}'z \right\}$$

and use the notation $1(E)$ to indicate the indicator for event E. Observe that

$$\begin{aligned}
\sqrt{n}(\hat{v} - v) &= \sqrt{n}\left[\max\left\{z_1'\hat{g}, \ldots z_{|S|}'\hat{g}\right\} - v\right] \\
&= \sqrt{n}\left(\max\left\{z_1'\hat{g}, \ldots z_J'\hat{g}\right\} - v\right) 1(E_n^c) + \sqrt{n}\left(\max_{z \in S} \hat{g}'z - v\right) \times 1(E_n) \\
&= \max\left\{z_1'\sqrt{n}(\hat{g} - g), \ldots z_J'\sqrt{n}(\hat{g} - g)\right\} \times 1(E_n^c) + \sqrt{n}\left(\max_{z \in S} \hat{g}'z - v\right) \times 1(E_n).
\end{aligned}$$

If one can show that $\sqrt{n}\Pr(E_n) \to 0$ as $n \to \infty$, then $1(E_n^c)$ will converge to 1 in mean and $\sqrt{n}(\max_{z \in S} \hat{g}'z - v) \times 1(E_n)$ will converge to zero in probability (since $(\max_{z \in S} \hat{g}'z - v)$ is bounded with probability 1) and so we will be done. But,

$$\sqrt{n}\Pr(E_n) \leq \sqrt{n}\Pr\left[\cup_{j=J+1}^{|S|}(\hat{g}'z_j \geq \hat{g}'z_1)\right] \leq \sqrt{n}\sum_{j=J+1}^{|S|}\Pr(\hat{g}'z_j \geq \hat{g}'z_1) \leq \sqrt{n}|S|ce^{-\rho n}$$

for some $\rho, c > 0$ by lemma 1. This implies the conclusion. ■

Corollary 1 *The case $J = 1$ corresponds to case (iii) above. When $J = 1$ and z_1 is the unique population maximizer, we will have $\sqrt{n}(\hat{v} - v) \xrightarrow{d} N(0, z_1'\Sigma z_1)$. So in this case $\hat{v}$ will have an asymptotic normal distribution, centered at zero.*

Estimating the distribution of $\hat{v}$ will be discussed in section 5.3 below.

5.2 Solution

I now describe the asymptotic properties of the solution in the non-degenerate case. Although it is possible to first describe case (ii) above and then derive (iii) as a special case, it appears that case (iii) can be of independent interest (e.g. the asymptotic distribution of the maximum value is normal, case (ii) does not arise with binary covariates etc.). So, first consider the case (iii), i.e. a unique solution exists to the population problem. The following proposition shows that the solution to the sample problem is consistent.

Proposition 5 *Assume that (i) all assumptions of lemma 1 hold, (ii) the population problem (3) has a unique solution* p_0*. Then the solution to the sample problem,* $\hat{p}$ *converges in probability to the solution of the population problem* p_0 *as* $n \to \infty$*.*

Proof. Suppose that $\hat{p}$, p_0 solve the problem under $\hat{g}$ and g respectively. We want to show that $p\lim_{n\to\infty}(\hat{p}-p_0)=0$.

Suppose $p\lim_{n\to\infty}(\hat{p}-p_0)\neq 0$. Then, we must have

$$p\lim_{n\to\infty}(\hat{p}-p_0)'\,g<0. \tag{7}$$

Observe that since p_0 solves the problem uniquely for g, we cannot have $p\lim_{n\to\infty}(\hat{p}-p_0)'\,g = (p\lim_{n\to\infty}\hat{p}-p_0)'\,g=0$ Otherwise, $p\lim_{n\to\infty}\hat{p}\neq p_0$ will be another solution because $p\lim_{n\to\infty}\hat{p}$ will belong to the parameter space (note that for each n, $\hat{p}$ satisfies the constraints of (3)). Also, since by definition of p_0, $g'p_0>g'\hat{p}$ for all n, we cannot have $p\lim_{n\to\infty}(\hat{p}-p_0)'\,g>0$. Now,

$$(\hat{p}-p_0)'\,\hat{g}=(\hat{p}-p_0)'\,g+(\hat{p}-p_0)'\,(\hat{g}-g)\,. \tag{8}$$

For n large enough, $(\hat{p}-p_0)'\,(\hat{g}-g)$ can be made arbitrarily close to 0 since the entries of $\hat{p},p_0$ lie within the unit cube and $\hat{g}$ is consistent for g (which follows from assumption (i) by WLLN and the continuous mapping theorem). In view of this and (7), for large enough n, the LHS of (8) can be made strictly negative. But then for this n, we have that $\hat{p}'\hat{g}<p_0'\hat{g}$ which contradicts the definition of $\hat{p}$ since p belongs to the parameter space and gives a strictly larger objective function. ∎

The sample solution is next shown to converge to the population solution at least exponentially fast- a consequence of the discreteness of the parameter space and the large deviation result of lemma 1.

Proposition 6 *Under the same assumptions as the previous proposition,* $\Pr(\hat{p}\neq p_0)\leq ce^{-\rho n}$ *as* $n\to\infty$ *for some* $c,\rho>0$*.*

Proof. If $p_0 = z_1$, i.e. $z_1'g > z_j'g$ for all $j = 2, ... |S|$, then $\Pr\left(z_1'\hat{g} > z_j'\hat{g}\right) \geq 1 - ce^{-\rho n}$ for any $j = 2, ... |S|$, by lemma 1. This implies that

$$\begin{aligned}\Pr\left(\hat{p} = z_1\right) &= \Pr\left(z_1'\hat{g} > z_2'\hat{g}, z_1'\hat{g} > z_3'\hat{g}, ...z_1'\hat{g} > z_{|S|}'\hat{g}\right) \\ &= 1 - \Pr\left(\cup_{j=1}^{|S|}\left\{z_1'\hat{g} < z_j'\hat{g}\right\}\right) \\ &\geq 1 - \sum_{j=1}^{|S|}\Pr\left(z_1'\hat{g} < z_j'\hat{g}\right) \\ &= 1 - O\left(|S|\, e^{-\rho n}\right) = 1 - O\left(e^{-\rho n}\right).\end{aligned}$$

Analogously, if $p_0 = z_j$, then $\Pr\left(\hat{p} = z_j\right) \geq 1 - ce^{-\rho n}$ for all $j = 1, 2, ... |S|$. ■

Now, I turn to the general case (ii) i.e. $v = g'z_1 = g'z_2 = ...g'z_J$ and $g'z_j < v$ for all $J < j \leq |S|$ where $1 \leq J \leq |S|$. In this case, one has to modify the sample maximization criterion so that the set of sample maxima converges to the set of population maxima $\{z_1, ...z_J\}$. The following proposition states this modification and establishes consistency in the Hausdorff sense. This proposition is analogous to Manski and Tamer's (2002) result on consistency for set identified parameters applied to a discrete parameter space.

Proposition 7 *Assume that* $\sqrt{n}\left(\hat{g} - g\right) \xrightarrow{d} W \equiv N\left(0, \Sigma\right)$ *and that* $v = g'z_1 = g'z_2 = ...g'z_J$ *and* $g'z_j < v$ *for all* $J < j \leq |S|$ *where* $1 \leq J \leq |S|$. *Define*

$$\Theta = \{z_1, ...z_J\}, \; \Theta_n = \left\{z^* \in S : \hat{g}'z^* \geq \max_{z \in S} \hat{g}'z - c_n\right\},$$

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

Remark 3 *The case* $J = |S|$ *has been treated separately above in proposition 3. The reasons for doing this are that in this (degenerate) case, (a) consistency is uninteresting since all feasible points are optimal and (b) the asymptotic distribution of* $\hat{v}$ *can be simulated without calculating the extreme points. But the previous proposition sets out the theory for a general* J, *which is also applicable to the degenerate case, i.e.* $J = |S|$ *as well as the unique solution case, i.e. for* $J = 1$, *described in propositions 5 and 6.*

Proof. Let $\hat{z} = \arg\max_{z \in S} \hat{g}'z$. Then using lemma 1, one can conclude just as in the previous proposition that $\Pr(\hat{z} \in \Theta)$ converges to 1 exponentially fast. Now, assume that $z \in \Theta$. Then

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

where equality $\overset{(1)}{=}$ comes from the fact that for $z, \hat{z} \in \Theta$, we have $g'z = g'\hat{z}$. Now, $\Pr(\hat{z} \in \Theta) \to 1$, $\sqrt{n}(\hat{g} - g)'(z - \hat{z}) = O_p(1)$ and $-\sqrt{n}c_n \to -\infty$ as $n \to \infty$. So the above probability tends to 1. This implies that $\Pr(z \in \Theta_n) \to 1$.

Conversely, suppose $z \notin \Theta$. Then $g'z < g'z_j$ for all $z_j \in \Theta$. Then

$$\begin{aligned}
\Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n) &= \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \in \Theta) + \Pr(\hat{g}'z > \hat{g}'\hat{z} - c_n, \hat{z} \notin \Theta) \\
&\overset{(1)}{=} \Pr(\hat{g}'z - g'z > \hat{g}'\hat{z} - g'\hat{z} + g'\hat{z} - g'z - c_n, \hat{z} \in \Theta) + o(1) \\
&= \Pr\left((\hat{g} - g)'(z - \hat{z}) > g'(\hat{z} - z) - c_n, \hat{z} \in \Theta\right) + o(1). \qquad (9)
\end{aligned}$$

Equality (1) comes from the fact that if $v = g'z_1 = g'z_2 = ...g'z_J$ and $g'z_j < v$ for all $J < j \leq |S|$ where $1 < J \leq |S|$, then

$$\Pr\left(\max_{j=1,...|S|} z_j'\hat{g} = z_k'\hat{g}, \text{ for some } k \notin \{1, ...J\}\right)$$

converges to zero. Indeed, one minus the above probability is at least

$$\Pr(z_k'\hat{g} > z_1'\hat{g}) = \Pr\left((z_k - z_1)'(\hat{g} - g) > (z_1 - z_k)'g\right)$$

and $(z_1 - z_k)'g > 0$ whence consistency of $\hat{g}$ yields the conclusion. Now, $\hat{z} \in \Theta$ implies $g'(\hat{z} - z) > 0$. Since $\Pr(\hat{z} \in \Theta) \to 1$ and $c_n \to 0$, using Cramer's theorem, we get that the probability in (9) goes to 0 at least exponentially fast. This shows that $z \notin \Theta$ implies $\Pr(z \in \Theta_n) \to 0$. ■

5.3 Asymptotic distribution, confidence intervals and bias correction

From propositions 4 and 7, the asymptotic distribution of $\hat{v}$ can be approximated as follows. Calculate Θ_n and approximate the asymptotic distribution of $U \equiv \sqrt{n}(\hat{v}-v)$, which is same as the asymptotic distribution of $\max_{z\in\Theta}\left\{\sqrt{n}\left(\hat{g}-g\right)'z\right\}$, by $\max_{z\in\Theta_n}\left\{\sqrt{n}\left(\hat{g}-g\right)'z\right\}$. The following proposition shows that approximating $\max_{z\in\Theta}\left\{\sqrt{n}\left(\hat{g}-g\right)'z\right\}$ by $\max_{z\in\Theta_n}\left\{\sqrt{n}\left(\hat{g}-g\right)'z\right\}$ does not entail any additional estimation error.

Proposition 8 *Under the assumptions of proposition 4 and 7, one has that*

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

Proof. If $\max_{z\in\Theta_n}\left\{\sqrt{n}\left(\hat{g}-g\right)'z\right\}\neq\max_{z\in\Theta}\left\{\sqrt{n}\left(\hat{g}-g\right)'z\right\}$ for any finite n, there will exist either a $z\in\Theta_n$ with $z\notin\Theta$ or a $z\notin\Theta_n$ with $z\in\Theta$ and the probability of both these events tend to 0 by proposition 7. ■

Since one can consistently estimate Σ, one can simulate the distribution of U by drawing observations $\{w_1,....w_N\}$ for some large N from a $N\left(0,\hat{\Sigma}\right)$ and calculating $\max_{z_j\in\Theta_n}\{w'z_j\}$ for each drawn w. The distribution of these maximum values will simulate the asymptotic distribution of $\sqrt{n}(\hat{v}-v)$. In particular, if $|\Theta_n|=1$ and $\Theta_n=z$, then this distribution is simply $N\left(0,z'\hat{\Sigma}z\right)$. The upper and lower α percentiles of this simulated distribution, denoted by c_u and c_l can be used to construct a confidence interval for v, given by $CI_{nd}=\left(\hat{v}-\frac{c_u}{\sqrt{n}},\hat{v}-\frac{c_l}{\sqrt{n}}\right)$. Note that due to the asymptotic bias of $\hat{v}$ in the general case where $|\Theta|>1$, this confidence interval will not be centered at $\hat{v}$.

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

Note that the way CI_{nd} is constructed implies that CI_{nd} is an asymptotically valid CI for v for all values of g. However, note that under additivity, one can construct CI_d (defined in section 4) which does not involve calculating the extreme points, as described in section 4. Since calculating extreme points is time consuming, especially for problems with large M, one can adopt a pretest method for calculating confidence intervals. This approach is followed in the application, below.

One first tests for additivity. Upon failure to reject the null of degeneracy, one calculates the confidence interval CI_d. Upon rejection of this null, one calculates Θ_n and the confidence interval CI_{nd}. This saves one the calculation of extreme points if the null of additivity were not rejected. However, one loses the size of the test of additivity from the coverage probability as is shown below. Also, recall from remark 1 in section 3 that when $M = 2$, the distribution of $\hat{v}$ is given by either proposition 3 or corollary 1 in the previous subsection. Construction of Θ_n is then unnecessary.

The overall confidence interval is, in the pretest case,

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

where c is the critical value used in the T_n-based test of degeneracy, described in proposition 2. The estimator of the maximum value is simply $\hat{\theta} = \hat{g}'\hat{p}$, no matter whether the null of degeneracy is rejected or not. The following proposition describes the probability that $\hat{I}$ covers the maximum value v.

Proposition 9 *Under the assumptions of propositions 3, 6 and 7, for all g,*

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

where α' is the size of the test described in proposition 2.

Proof. For values of g such that $g \notin \mathcal{R}(A)$, i.e.$rank\left(B\right) = M + 1$, we have

$$\begin{aligned} & \Pr\left(v \in \hat{I} | g \notin \mathcal{R}(A)\right) \\ = \; & \Pr\left(T_n \geq c,\, c_l \leq \sqrt{n}\left(\hat{g}'\hat{p} - v\right) \leq c_u | g \notin \mathcal{R}(A)\right) \\ & + \Pr\left(T_n < c,\, d_l \leq \sqrt{n}\left(\hat{g}'\hat{p} - v\right) \leq d_u | g \notin \mathcal{R}(A)\right). \end{aligned} \tag{10}$$

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

$$\Pr\left(A_1 \cap A_2\right) \geq \Pr\left(A_1\right) + \Pr\left(A_2\right) - 1. \tag{11}$$

So the first term in (10) is at least

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

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

$$\Pr\left(v \in \hat{I} | g \notin \mathcal{R}(A)\right) \geq 1 - \alpha \tag{12}$$

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

For values of g such that $rank(B) = M$, i.e. $g \in \mathcal{R}(A)$, we have

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

Using (11), the second probability is at least

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

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

$$\Pr\left(v \in \hat{I} | g \in \mathcal{R}(A)\right) \geq 1 - \alpha - \alpha'. \tag{13}$$

From (12) and (13), the conclusion follows. ■

Before ending analysis of the mean maximization case, a few remarks are in order.

Remark 4 ***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.*[7] *If this weighting function is* $u(.)$, *e.g. a concave "utility", define* $g_j = E(u(y)|j)$, *i.e. the expected value of* $u(y)$ *across rooms 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.*

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

Remark 5 ***Which covariates:*** *Obviously, including more covariates or refining the support of a given covariate leads to higher maxima in both the population (i.e. with known g's) and in the sample (with estimated $\hat{g}$'s) since the maxima is sought over a larger set. So when one knows average outcome for every potential combination of all covariates, using all covariates will yield the largest maximum mean outcome- which we will call $g'p_0$. But for a finite sample size, finer categorization implies smaller precision in the estimation of $\hat{g}$'s making it more likely that the MSE of the estimated maximum will become larger when using more covariates. These issues are discussed in details in Manski (2004) in the treatment choice context.*

Remark 6 ***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. Moreover, 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.*

6 Quantiles

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

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

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

Then the population maximization problem can be stated as

$$\max_{p \in \mathcal{P}} \mu_p \tag{14}$$

Examining uniqueness of the population problem and finding asymptotic properties of the sample based estimate may first seem highly complicated due to the nonlinearity of μ_p in $\{p_j\}$. I will argue below that several of the key insights for the mean case actually carry over to the median and the analysis is quite similar to the mean case although this is hardly obvious to start with.

The first proposition will establish that the "extreme point" idea, which is key to the mean analysis above, also extends to the median. This will be key to the rest of this section.

Proposition 10 *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 (14).*

Proof. The idea of the proof is the observation that if $r = \lambda p + (1-\lambda) q \in \mathcal{P}$ with $\lambda \in (0,1)$, then the median corresponding to the weighting r lies "between" the medians corresponding to the allocations p and q. To see this, WLOG assume that $\mu_p > \mu_q$.

Suppose $\mu_r > \mu_p$. Then

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

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

$$\sum_{j=1}^{m} r_j F_j\left(\mu_p\right) > \sum_{j=1}^{m} r_j F_j\left(\mu_r\right). \tag{15}$$

Now, since $\mu_r > \mu_p$, one has that for each j, $F_j\left(\mu_r\right) \geq F_j\left(\mu_p\right)$ implying $\sum_{j=1}^{m} r_j F_j\left(\mu_p\right) \leq \sum_{j=1}^{m} r_j F_j\left(\mu_r\right)$- which contradicts (15). This implies that $\mu_r \leq \mu_p$.

Next,

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

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

Remark 7 *Clearly, one can generalize the previous proposition to any functional* $\nu(.)$ *of the CDF which satisfies*

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

for every $\lambda \in [0,1]$ *and* $F_p(.) \equiv \sum_{j=1}^{m} p_j F_j(.)$. *It is not clear at this point if there is an alternative and more familiar characterization of the set of all such functionals* $\nu(.)$. *Means and quantiles seem to be the two leading examples which are obvious policy targets. But inter-quantile differences do not satisfy this property in general.*

It follows from the above proposition that just like in the mean case, one can focus on the extreme points of the constraint set which can be calculated a priori. To my knowledge, this insight is new in regards to the programming literature as well. Further, notice that the only property of the F_j's used in the above proof is that they are nondecreasing. So exactly the same result will hold for the sample problem since the estimates $\hat{F}_j$ are also nondecreasing.

Thus, all the results for the mean case can be adapted to those for quantiles, once one can establish a large deviation result for the median, which is done in lemma 2, below. The only difference is that there does not seem to be a result analogous to proposition 1. Hence a test for degeneracy in the quantile case has to be based on extreme values and will, therefore, be harder to implement than the test described in proposition 2. In particular, one can test $\mu_{z_1} = ... = \mu_{z_{|S|}}$ based on the estimates $\hat{\mu}_{z_1}, ...\hat{\mu}_{z_{|S|}}$. Since the $\hat{\mu}_z$'s are asymptotically normal, this is equivalent to testing the equality of several normal means, which is a well-studied problem in multivariate statistics.

We now turn attention to asymptotic properties of the estimated maximum medians. Let v denote the maximum value corresponding to (14) as before with sample counterpart denoted by $\hat{v}$. Using the implication that $\hat{v} = \max\left\{\hat{\mu}_{z_1}, ...\hat{\mu}_{z_{|S|}}\right\}$ and using continuity of the max function, consistency of the maximum *value* follows by continuous mapping theorem since sample medians are well-known to be consistent for population medians. The above argument holds independently of whether the population *solution* is unique.

I now state a large deviations type result which will play the role that lemma 1 played for analysis of the mean. Note that for the estimated median, we have

$$0.5=\sum_{j=1}^{m}p_j\left(\frac{\frac{1}{n}\sum_{i=1}^{n}D_{ij}1\left(y_i\leq\hat{\mu}_p\right)}{\frac{1}{n}\sum_{i=1}^{n}D_{ij}}\right)\equiv\sum_{j=1}^{m}p_j\hat{F}_j\left(\hat{\mu}_p\right),$$

and let $F_p(.)$ denote the CDF corresponding to weighting p, i.e. $F_p(y)=\sum_{j=1}^{m}p_jF_j(y)$.

Lemma 2 *Let $p,q\in S$ be two allocations with corresponding medians satisfying $\mu_p>\mu_q$. Let $\hat{\mu}_p$, $\hat{\mu}_q$ denote their estimates. Assume that for every allocation p, F_p admits a density around its median μ_p. Then for some $c,\rho>0$,*

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

Proof. Notice that

$$\begin{aligned}\Pr\left(\hat{\mu}_p<\hat{\mu}_q\right) &= \Pr\left(\mu_q<\hat{\mu}_p<\hat{\mu}_q<\mu_p\right)+\Pr\left(\mu_q<\hat{\mu}_p<\mu_p<\hat{\mu}_q\right)\\ &\quad+\Pr\left(\hat{\mu}_p<\mu_q<\hat{\mu}_q<\mu_p\right)+\Pr\left(\hat{\mu}_p<\hat{\mu}_q<\mu_q<\mu_p\right). \qquad (16)\end{aligned}$$

The first of these probabilities can be further decomposed into

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

Rewriting $\frac{1}{2}\left(\mu_p-\mu_q\right)$ as $\delta>0$, (16) is dominated by

$$2\Pr\left(\frac{1}{2}\left(\mu_p-\mu_q\right)<\delta\right)+\Pr\left(\hat{\mu}_p-\mu_p<-\delta\right)+\Pr\left(\delta<\hat{\mu}_q-\mu_q\right)+2\Pr\left(\hat{\mu}_p-\mu_p<-\delta\right).$$

Now, notice that

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

Since $\delta > 0$, we have that $\bar{m}\left(\frac{1}{2} - F_p\left(\mu_p - \delta\right)\right) > 0$ since $F_p\left(.\right)$ admits a density in a neighborhood of the median; also $\bar{m} - \delta_j < 0$ for all j. Cramer's theorem implies that both probabilities go to zero exponentially fast since

$$
E\left\{D_{ij} 1\left(y_i \leq \mu_p - \delta\right) - \bar{d}_j F_j\left(\mu_p - \delta\right)\right\} = E\left\{D_{ij} 1\left(y_i \leq \mu_p - \delta\right)\right\} - \delta_j F_j\left(\mu_p - \delta\right) = 0.
$$

A similar argument works for the other terms in (16). ■

Now consider the implication of the above lemma.

Proposition 11 *Assume that* $\sqrt{n}\left(\hat{\mu}_z - \mu_z\right) \xrightarrow{d} N\left(0, \sigma_z^2\right)$, *for which sufficient conditions are standard. Also suppose that* $\mu_{z_1} = ... = \mu_{z_J} = v$ *for some* $1 \leq J \leq |S|$. *Then*

$$
\sqrt{n}\left(\hat{v} - v\right) = \max\left\{\sqrt{n}\left(\hat{\mu}_{z_1} - \mu_{z_1}\right), ...\sqrt{n}\left(\hat{\mu}_{z_J} - \mu_{z_J}\right)\right\} + o_p\left(1\right).
$$

In particular, if $J = 1$ *and* $z_1 = \arg\max_{j=1,,,,|S|}\left\{\mu_{z_j}\right\}$, *then*

$$
\sqrt{n}\left(\hat{v} - v\right) = \sqrt{n}\left(\hat{\mu}_z - \mu_z\right) + o_p\left(1\right) \xrightarrow{d} N\left(0, \sigma_z^2\right).
$$

Proof. Exactly analogous to proposition 4, replacing lemma 1 by lemma 2. ■

The asymptotic properties of the estimated solutions will be exactly analogous to the mean case and will depend on whether the population solution is unique.

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

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

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

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

7 Application

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

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

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 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 conclusions of propositions 4 and 7, I will 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. One may view the requirement of $c_n \to 0$ to be of "greater priority" than $\sqrt{n}c_n \to \infty$ in the conservative sense that the former implies that $\Pr(\Theta_n \subset \Theta) \to 1$. 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).

The second column in table 1 reports the value of the test statistic T_n and the 90% and 95% critical points of the corresponding $\varkappa^2$ distribution (df=1, 3 and 6 for 2, 3 and 4 categories respectively). When I cannot reject the null of additive effects, I report the maximum value calculated from the sample, the bias corrected one next to it and a 95% confidence interval for the maximum value. When the outcome of interest is freshman year GPA, it is seen from table 1 that we cannot reject the null of additive effects. The value of the test statistic is always smaller than the 90% critical value. The confidence intervals are to be interpreted as pretest type and hence the exact coverage probability is less than 95% as shown in proposition 9.

In table 2A, the same exercise is repeated for the outcome "joining a Greek organization". It is seen from table 2A that the test statistic for additivity is significantly large (marked with an asterisk) in almost all cases, leading to rejection of additivity. I report the 95% confidence interval for the minimum value together with the sample maximum minus the minimum values. The next three columns report the number of vertices (equal to $|S|$ in our notation above), the tolerance level c_n and the cardinality of Θ_n for each choice of c_n. The corresponding 95% CI are reported in the last column. When we have 2 categories, we cannot have a nontrivial set-identified situation as pointed out in remark 1 in section 3. So for this case, when I reject additivity, I calculate CI based on the normal approximation and when I do not reject it, I calculate the pretest CI based on the distribution of proposition 3. These are called "normal" and "pretest" in table 2A.

Concentrating on the panel for men, it is seen that the additivity is rejected for both small (2) and large number (4) of categories but not for the intermediate one (3). This is likely a consequence of the fact that one loses precision as the number of categories rises leading to a decrease in the value of the test statistic but it becomes harder to "fit" a b to a larger number of $\hat{g}$'s as described in the definition of T_n.

Table 2B describes the nature of the maximizing allocations corresponding to the cases where additivity was rejected in table 2A. Columns 3 and 4 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.

Similar exercises with race are reported in tables 3A-C. 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, Native Indian and Asian Indian) or not. These results are reported in table 3 when the outcome is joining a Greek house. There seems to be significant nonadditivity 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.

Table 4 shows that when race is classified as white and others, there does not seem to be any evidence of non additivity in (race-induced) peer effects for either GPA or joining fraternities.[8]

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

[8] Finer categorization of "others" into "blacks" and "others" caused problems in estimation since the number of rooms with two black men is one- which makes it impossible to estimate the requisite variance.

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 based on methods of testing multivariate normal means, as mentioned in section 7, 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.

8 Conclusion

This paper draws on insights from the mathematical programming literature to study identification and estimation of optimal allocations in input-matching 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. Further, it is shown that asymptotic properties of the estimated maximum value depend on uniqueness of the population solution which is investigated in details. A further contribution of the paper is to extend the analysis to maximizing quantiles of the outcome distribution and showing that almost all the insights from mean-maximization carry over to the quantile-maximization case even if the quantile problem, unlike the mean problem, is nonlinear.

Randomized field experiments, which are becoming increasingly popular among applied microeconomists, provide the ideal source of data to which these methods can be applied for designing optimal policies. In nonrandomized observational studies, one could use instrumental variables to estimate underlying structural functions which could, in turn, be utilized to compute optimal allocations. So the applicability of methods developed in this paper are not restricted exclusively to randomized settings.

When these methods are applied to the randomized setting of room allocation for Dartmouth freshmen, the analysis reveals several interesting conclusions. First, segregation by previous background seems to affect social behavior of average men and women in exactly opposite ways. For women, this leads to lowest propensity of joining social organizations like sororities whereas for men, the propensity to join fraternities is maximized. This likely results from the average woman's tendency to find close companionship among other women with whom they can identify. When the outcome of interest is freshman year GPA, it is seen that for both men and women, segregation by prior academic achievement leads to maximizing higher percentiles and minimizing lower percentiles of the GPA distribution without affecting the mean or the median. This highlights the importance of considering the effects of allocation rules on quantiles of the distribution if equity is to be an important goal of policy-making.

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

References

[1] Cooley, Jane (2006): Desegregation and the Achievement Gap: Do Diverse Peers Help?, Duke University.

[2] Cragg, J. G and Donald, S. G. (1997): Inferring the Rank of a Matrix, Journal of Econometrics, vol. 76, no. 1-2, Jan.-Feb. 1997, pp. 223-50.

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

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

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

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

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

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

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

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

[11] Hoxby, C. and G. Weingarth (2006): Taking Race Out of the Equation: School Reassignment and the Structure of Peer Effects. Manuscript.

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

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

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

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

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

[17] 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.

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

[19] 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.

Table 0: Sample Characterestics

Variable	Mean	SD	Range
Women (N=428)			
ACA	202.68	12.65	156, 228
White	0.69	0.46	0, 1
Non-minority	0.87	0.33	0, 1
Freshman_GPA	3.23	0.39	1.56, 4.0
Sorority	0.47	0.5	0, 1
Men (N=436)			
ACA	205.55	12.99	151, 231
White	0.72	0.45	0, 1
Non-minority	0.91	0.29	0, 1
Freshman_GPA	3.15	0.45	1.15, 3.9
Fraternity	0.53	0.48	0,1

Table 1: Degeneracy, Y=GPA, X=ACA

# categories	Test stat	Critical values 90%, 95%	Sample Max	Bias-Corrected Max	95% CI	Max-Min
Men (mean=3.15)						
2	0.59	2.71, 3.84	3.17	3.16	3.13, 3.19	0.023
3	0.318	6.25, 7.81	3.16	3.16	3.15, 3.17	0.05
4	7.4	10.64, 12.59	3.205	3.17	3.14, 3.20	0.11
Women (mean=3.235)						
2	0.01	2.71, 3.84	3.238	3.23	3.21, 3.25	0.004
3	0.26	6.25, 7.81	3.25	3.24	3.245, 3.252	0.1
4	3.98	10.64, 12.59	3.275	3.25	3.22, 3.28	0.15

Table 2A: Y=Joining Frat, X=ACA

# Categories	test-statistic	Sample Min	Max-Min	# vertices	Toler	Cardinality	"95%" CI
Men (mean=0.53)							
2	3.05*	0.487	0.084	2	.	.	0.483, 0.490^^
3	1.43	0.51**	0.05	8	.	.	0.505, 0.511***
4	9.82*	0.463	0.163	37	0.003	2	0.458, 0.468
					0.03	18	0.463, 0.470
					0.3	37	0.463,0.470
Women (mean=0.47)							
2	3.65*	0.423	0.09	2			0.419, 0.427^^
3	8.25*	0.38	0.16	8	0.003	1	0.377, 0.385
					0.03	2	0.376, 0.384
					0.3	8	0.377, 0.386
4	6.7	0.399^	0.106	41	0.003	4	0.395, 0.404
					0.03	10	0.397, 0.404
					0.3	41	0.399, 0. 408

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

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

# Categories	test-statistic	HH and LL: Max	HH and LL: Min
Men (mean=0.53)			
2	3.05*	0.51, 0.49	0.037, 0.00
4	9.82*	0.01, 0.24	0.00, 0.00
Women (mean=0.47)			
2	3.65*	0.037, 0	0.52, 0.48
3	8.25*	0.00, 0.00	0.11, 0.28

Table 3A: X=Race: Non-minorities (L), minorities (H), Y=GPA

	Test-statistic	HH and LL: Max	HH and LL: Min
Men (mean=0.53)	3.3*	0.114, 0.885	0, 0.77
Women (mean=0.47)	4.66**	0, 0.78	0.11, 0.89

Table 3B: =Race: Non-minorities (L), minorities (H), Y=Joining Frat

		Critical values			
	test-statistic	90%, 95%	Sample Min	"95%" CI	Max-Min
Men (mean=0.53)	3.3*	2.71, 3.84	0.3	0.23, 0.36	0.077
Women (mean=0.47)	4.66**	2.71, 3.84	0.09	0.05, 0.13	0.04

Table 3C: =Race: Non-minorities (L), minorities (H), Y=Joining Frat

	test-statistic	HH and LL: Max	HH and LL: Min
Men (mean=0.53)	3.3*	0.114, 0.885	0, 0.77
Women (mean=0.47)	4.66**	0, 0.78	0.11, 0.89

Table 4: Race=White,Others

	test-statistic	Critical values 90%, 95%
Y=Freshman GPA		
Men (mean=3.15)	0.28	2.71, 3.84
Women (mean=3.235)	0.33	2.71, 3.84
Y=Joining frat		
Men (mean=0.53)	0.26	2.71, 3.84
Women (mean=0.47)	1.27	2.71, 3.84

Table 5: Maximal allocations: Y=GPA, X=ACA, # Categories=2

	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