ISSN 1471-0498

DEPARTMENT OF ECONOMICS

DISCUSSION PAPER SERIES

EVALUATING TREATMENT PROTOCOLS USING DATA COMBINATION

Debopam Bhattacharya

Number 609
June 2012

Manor Road Building, Oxford OX1 3UQ

Evaluating Treatment Protocols using Data Combination*

Debopam Bhattacharya,
University of Oxford

March 15, 2012

Abstract: In real-life, individuals are often assigned to binary treatments according to existing treatment protocols. Such protocols, when designed with "taste-based" motives, would be productively inefficient in that the expected returns to treatment for the marginal treatment recipient would vary across covariates and be larger for discriminated groups. This cannot be directly tested if assignment is based on more covariates than the researcher observes, because then the marginal treatment recipient is not identified. We present (i) a partial identification approach to detecting such inefficiency which is robust to selection on unobservables and (ii) a novel way of point-identifying the necessary counterfactual distributions by combining observational datasets with experimental estimates. These methods can also be used to (partially) infer risk-preferences which may rationalize the observed treatment allocations. Specifically, existing healthcare datasets can be analyzed with the proposed tools to test the allocational efficiency of medical treatments. Using our methodology on data from the Coronary Artery Surgery Study in the US, which combined experimental and observational components, we find that after controlling for age, smokers in the observational dataset had to overcome a higher threshold of expected survival relative to non-smokers in order to qualify for surgery. Our methods are applicable when individuals cannot alter their potential treatment outcomes in response to the treatment regime, unlike in the case of law enforcement.

*Address for correspondence: debobhatta@gmail.com. I am grateful to four anaonymour referees, the editor and seminar participants at CEMMAP, Cambridge and Uppsala for comments and to Amitabh Chandra for pointing me to the CASS dataset. All errors are mine.

1 Introduction

In many real-life situations, external agencies assign individuals to treatments using covariate based protocols. For example, caseworkers assign the unemployed to job-training based on employment record, doctors refer patients to surgical or medical treatment based on clinical test results, colleges admit student applicants to academic programs based on test scores and so on. When protocols are chosen to maximize a functional (say, mean) of the marginal distribution of the resulting outcome subject to cost constraints, the protocol can be said to be "outcome-based". In the above examples, the outcomes can be post-program earnings, days of survival and performance in final examination, respectively. In all these cases, optimal protocol choice will seek to equate the returns to treatment for the marginal treatment recipient across covariate groups but this will typically cause average treatment rates to vary between groups. If, on the other hand, protocols are chosen to maximize a covariate weighted mean of the outcome, the protocol can be said to be covariate-based and the resulting between-group disparities in treatment rates at the optimal choice be regarded as having arisen from "taste-based" objectives. For example, consider the case where the treatment is assigning heart-attack patients to surgical treatment and the outcome is post-surgery mortality. Then, an outcome oriented protocol choice will aim to maximize mean days of survival. In contrast, a covariate-oriented protocol choice will seek to maximize mean weighted survival where the weights vary with covariate– rather than outcome– values such as race or gender of the patient.

A covariate-oriented protocol, as opposed to an outcome-oriented one, implies that the treatment, to be thought of as a scarce resource, is being assigned among individuals in a way that does not maximize its overall productivity, where productivity is measured solely in terms of the outcome. This idea has a long history in economics (Becker, 1957, Arrow, 1973) and suggests that distinguishing between the two types of protocols may be based on testing inefficiency of treatment assignment using outcomes data. Detecting such inefficiency in practice, however, is difficult because in many situations, the planner observes more characteristics than us, the researchers (Heckman, 1998). This makes it hard to rule out the possibility that the subgroup receiving seemingly sub-optimal levels of treatment does so because they are less endowed with some unobservable (to us) qualities which lower their expected outcome from treatment as perceived by the treatment assignors. The purpose of this paper is to show how a partial

identification approach can be used in this situation to test implications of efficient treatment assignment and, more generally, to infer which welfare functionals, defined on the marginal distribution of outcome, can rationalize observed treatment assignments by the planners.

We focus on the case where the treatment in question is binary but allow the outcome of interest to be either binary or continuous. Assume that an experienced planner observes for each individual a set of covariates and assigns him/her to treatment based on the expected gains from treatment, conditional on these covariate values and subject to an overall cost-constraint. In this set-up, a necessary condition for the planner's assignment to be productively efficient is that in every observable covariate group, the expected net benefit of treatment to the marginal treatment recipient(s) is weakly greater than a common threshold which, in turn, is weakly greater than the expected net benefit of the marginal treatment non-recipient and where marginal is defined in terms of the characteristics observed by the planner. The planner's assignment results in an observational dataset, where for each individual, we observe her treatment status, her outcome and costs conditional on her treatment status and a set of covariate values. The problem is to test outcome-oriented treatment assignment from these data.

Typically, a single observational dataset is inadequate for this purpose for two reasons. The first, already noted above, is that the planner can base treatment assignment on characteristics that are not observed by us. This makes it hard, if not impossible, to know who are the "marginal" treatment recipients (c.f., Heckman, 1998, Persico, 2009) and the identified expected outcome conditional on observed covariates among the treated (untreated) will exceed (be smaller than) the expected outcome for the marginal treatment recipient – the so-called treatment-threshold. This problem is traditionally referred to as the "inframarginality" problem which is the main obstacle to testing equality of treatment thresholds across population subgroups. Secondly, benefits are also hard to measure using observational data alone because counterfactual means are not observed.

In this paper, we discuss a new approach to detecting outcome-oriented allocation in such situations using the notion of partial identification. Our approach is motivated by the implication of outcome-oriented allocation that expected net benefits in every subset of treated individuals must weakly exceed expected net benefits in every subset of untreated individuals– a (conditional) moment inequality condition. These moment inequalities for subsets defined by covariates that the planner observes have testable implications for the (cruder) subsets based

on the covariates that we observe and intend to test for.[1] These implications can therefore be tested provided we can identify the relevant counterfactual means.

The method of identifying counterfactual means needed for our approach depends on the application and data at hand. In this paper, we suggest and use a novel method which involves combining the observational dataset with experimental or quasi-experimental evidence on treatment effects for subjects drawn from the same population.[2] Such data combination is directly feasible in healthcare applications where experimental datasets from clinical trials coexist with observational survey evidence for many medical conditions and their treatment. For example, the website: https://biolincc.nhlbi.nih.gov/studies/ contains a large number of studies with linked trial datasets for a variety of health conditions including diabetes, hypertension and AMI. The (non-experimental) treatments of these diseases are also routinely covered in medical and hospital surveys and thus provide the necessary observational data. These include the National Health and Nutrition Examination Survey, the National Health Interview Survey and so on. Occasionally, observational data are collected as part of the same study (see application below). In economic applications, implementing the proposed method is now logistically feasible, given the increasing use of large-scale field experiments in such studies. See below for some concrete suggestions about how to implement this in practice. This paper attempts to make the case for collecting such data and using them in conjunction with observational studies to understand how treatments are in fact assigned by real decision-makers in economic contexts.

It should also be noted that our partial identification based approach can only test *implications* of inefficiency and as such may fail to detect inefficient treatment assignment when it is present. However, when we have detected inefficiency, we can be sure (up to the size of our test) that it exists. This is true generically for hypothesis tests regarding partially identified parameters and should be considered as the price one has to pay for the lack of point-identification.

A final point is that in the context of describing the methodology, I deliberately refrain from using the term "prejudice". Neither I, nor the existing research in my reading, can pinpoint the behavioral source of any observed discrepancy from the economic ideal of identical marginals. As such, any such discrepancy is usually stated to have arisen from "taste-based" motives, by

[1] Which covariates we should test on is guided by the problem at hand– e.g., for gender disparities we analyze expected returns for treated and untreated males and for treated and untreated females.

[2] Below in subsection 4.3, we investigate the consequences of the failure of this assumption.

definition. In section 4 below, I further elaborate on this distinction between inefficient treatment assignment and prejudice – a more subtle distinction than the well-recognized difference between taste-based and statistical discrimination – which seems to have been overlooked in the literature.

Substantive assumptions: We now state the substantive assumptions which define our set-up. The first is that the planner is experienced in the sense that he can form correct expectations. The second is that the planner observes and can condition treatment allocation on all the characteristics (and possibly more than those) that we observe. Third, we observe the same outcomes and costs whose expectations– taken by the planner– should logically determine (productively efficient) treatment assignment in the observational dataset. Fourth, there are negligible externalities, i.e. where treating one individual has a significant impact on the outcome of another individual (c.f., Angelucci et al, 2009). Fifth, we have at our disposal an experimental dataset where the treatment was randomized and this experimental dataset was drawn from the same population as the observational dataset.

The fourth assumption is credible in, say, the case of job-training, mortgage approval or treatment of non-infectious diseases such as heart attack but less so in, say, academic settings or treatment of infections such as AIDS or malaria. Bhattacharya, 2009 considers roommate assignment in college where peer effects play a crucial role. The third assumption simply clarifies that the notion of productivity (with respect to which inefficiency is defined) must be fixed beforehand and it should be observable and verifiable. The second assumption defines the "selection on unobservables" problem. The first assumption–a "rational expectations" idea is standard for analyzing choice under uncertainty in applied microeconomics (c.f., the KPT paper cited below). It is part of our *definition* of efficiency, i.e., we are testing the joint hypothesis that the planner can calculate correct expectations *and* is allocating treatment efficiently, based on those calculations. This has been termed "accurate statistical discrimination" elsewhere in the literature (c.f., Pope and Sydnor, 2008, Persico, 2009, page 250). Correct expectations are more reasonable for treatments that are fairly routine– such as college admissions to well-established academic programs and less tenable for treatments that are relatively new, e.g., admission to a relatively new academic program.[3] Concerns for misallocation, especially along discriminatory lines, are more frequently voiced for routine treatments and therefore, it makes

[3]We will be concerned with expectations conditional on covariate values and so correct expectation is more credible the cruder the conditioning set. In our application, we consider a two-covariate conditioning set.

sense to concentrate on those for the purpose of the present paper. Notice that here we are describing the beliefs of a large central planning body who is experienced, rather than small individuals making one-time choice decisions. It is presumably less contentious to expect correct beliefs in the former case than in the latter. The fifth assumption is maintained throughout the analysis and is further discussed in the paragraph titled "Alternative designs and data issues" under section 3 below.

Plan of the paper: Section 2 discusses the contribution of the present paper in relation to the existing literature in economics and econometrics. Section 3 presents the partial identification methodology, discusses how counterfactuals may be identified via data combination, describes how a bounds analysis can help detect misallocation. Section 4 discusses some issues related to the interpretation of the results obtained with our methodology and also investigates the robustness of our methodology to the failure of the identical distribution assumption. Section 5 analyzes the complementary problem of inferring a planner's underlying risk-preferences which would justify the current allocations as efficient. Section 6 presents the empirical illustration and section 7 concludes. The appendix contains the proof of the main theorem.

2 Literature

Persico, 2009, provides a comprehensive survey of existing empirical approaches to the detection of taste-based discrimination in general settings. The approaches are varied and their applicability is usually context-specific. Here, we focus on detecting evidence of taste-based assignment of a binary treatment where the treater can be expected to observe more characteristics than the researcher. Our approach is based on using outcome data. In that sense, it is thematically close to Knowles, Persico, Todd, 2001 (KPT, henceforth) who examined the problem of detecting taste-based prejudice separately from statistical discrimination in the context of vehicle search by the police, using data on the search-outcome (hit rates).[4] KPT's key insight is that in law-enforcement contexts, potential treatment recipients can alter their behavior– and thus their potential outcome upon being treated– in response to the treater's behavior. This implies that equilibrium hit rates should be equalized across *observed* demographic groups under

[4]Related recent papers include Anwar and Fang (2006), Grogger and Ridgeway (2006), Antonovic and Knight (2009) and Brock et al (2011).

efficient search– a testable prediction. If hit rates are higher for one group, then the police is better-off searching that group more intensively and hence the group is better-off reducing the contraband activity. While the KPT approach applies to many situations of interest, especially ones involving law enforcement, it is not applicable to all situations of treatment assignment where misallocation is a concern. For example, it is very difficult– if not impossible– for patients to alter their potential health outcomes with and without surgery in response to the nature of treatment protocols used by doctors.

In another outcome-based approach, Pope and Sydnor (2008) seek to detect taste-based discrimination in peer-to-peer lending programs. PS use the facts that in these lending programs, (i) the researchers observe all the characteristics that the planners (lenders) observe and (ii) a competitive auction among lenders for funding each individual application drives interest rates so that every approved loan is at the "marginal" level of (expected) return. PS observe the actual returns on the approved loans and can test efficiency by comparing mean (and thus marginal) returns across race for approved loans. The peer-to-peer lending situation is different from job-training, medical treatment etc., where the same treatment protocol is used for all applicants and/or treatments are not allocated via a competitive bid, so that the PS approach cannot be used here (c.f., page 11 of PS).

In the medical setting, Anwar and Fang (2011) consider a test of taste-based prejudice in emergency room discharge using the re-admission rate as the outcome of interest. The key assumption is that physicians have at their disposal a continuous choice variable related to diagnostic tests which they can choose optimally in order to determine suitability for discharge. The identification strategy is then based on comparing the re-admission rates of patients of different race who had undergone the diagnostic test at the physician-optimized level of intensity. In our set-up, we do not have data on any such continuous choice variable available to physicians before they decide on surgery.

A second aim of the present paper is to infer what outcome-based objectives can rationalize observed treatment disparities across demographic groups. In that sense, it has some substantive similarities to a series of papers in the time-series forecasting literature which propose testing rationality of forecasts made by central agencies (c.f. Elliott, Komunjer and Timmerman (2005), Patton and Timmerman (2007) and references cited therein). The idea there is to (point) estimate parameters of a loss-function which rationalize the observed forecasts. The set-up in

that literature assumes that the action (i.e., the forecast) itself has no effect on the distribution of the realized future outcome. In contrast, the key issue in our set-up is that the action (the imposed treatment status) fundamentally determines which distribution the eventual outcome will be drawn from and so the methodology of forecast rationality tests cannot be used in our problem.

A recent set of papers in the econometrics literature have addressed the issue of how treatments should be assigned when only finite sample information is available to the *planner* regarding treatment effectiveness. This is relevant to those treatments that are relatively new, so that the planner is unlikely to know the actual distribution of outcomes with or without treatment– a situation usually termed "ambiguity" in the decision theory literature. See, for instance, Dehejia, 2005, Manski, 2004, 2005, Hirano and Porter, 2008 and Bhattacharya and Dupas, 2010. The present paper may be described as addressing the reverse problem. That is, when the treatment in question is routine and the planner can be expected to know the true outcome distributions (or at least able to form correct expectations), can *we* assess efficiency of the treatment assignment protocols using finite sample evidence, allowing for the possibility of selection on unobservables?

3 Methodology

Using the Neyman-Rubin terminology, denote outcome with and without treatment by Y_0 and Y_1, respectively and let $\Delta Y = Y_1 - Y_0$. We will allow for treatment effects to be negative, i.e., $\Pr(Y_1 - Y_0 < 0)$ may be positive. Analogously, define C_1 and C_0 as the potential costs corresponding to treatment 1 and treatment 0, respectively with $\Delta C \equiv C_1 - C_0 > A > 0$. The available budget per potential subject is denoted by c. This set-up captures the fact that treatment 1 is more expensive for everyone (e.g., an invasive surgery or administrative costs of dealing with loan payment) relative to treatment 0 (such as treatment with medicine or zero administrative costs incurred when loans are not approved). However, for some individuals, the more expensive treatment may be detrimental. In some applications, costs per se do not vary across individuals but treatment allocation is limited by capacity constraints. In such cases, we will let $C_1 \equiv 1$, $C_0 \equiv 0$ and let c denote the fraction of potential subjects who can be treated – i.e., the capacity constraint.

Let $W = (X, Z)$ denote the covariates observed by the planner, where the component Z is not observed by us. Let $\mathcal{X}, \mathcal{Z}$ denote the support of X, Z and $\mathcal{W} = (\mathcal{X} \times \mathcal{Z})$ denote the support of W. Let E denote expectations taken w.r.t. the planner's subjective probability distributions, which are assumed to be identical to the true probability distributions in the population. We will assume that all variables defined here have finite expectations. The planner's treatment allocation gives rise to the observational dataset, where for each individual, we observe her treatment status ($D = 1$ or 0), her outcome, Y and cost C which are respectively (Y_1, C_1) or (Y_0, C_0) depending on whether $D = 1$ or 0, and the set of covariates X. For any random variables U, V, let $F_{U|V}(u|v)$ denote the conditional C.D.F. of U at u given $V = v$ and $F_U(\cdot)$ denote the marginal c.d.f. of U.

From the planner's perspective, a treatment protocol is a function $p : \mathcal{W} \to [0, 1]$, specifying the probability of treatment for individuals with $W = w$. Each such protocol will give rise to a distribution of outcome Y, given by

$$F^p(y) = \int \left[p(w) F_{Y_1|W}(y|w) + \{1 - p(w)\} F_{Y_0|W}(y|w) \right] dF_W(w).$$

An outcome-based criterion is one where protocol p is preferred over protocol q if and only if $F^p(\cdot)$ is preferred over $F^q(\cdot)$. The latter preference could be captured by expected utility i.e. $U(p) = \int u(y) dF(y|p)$ or quantile utility $U(p) = F^{-1}(\tau|p)$ for some $\tau \in [0, 1]$ etc. The important point here is that the planner's preferences are over the *marginal* distribution of Y resulting from the protocol and not the distribution of Y, conditional on W and hence the term "outcome-based". For example, if the treatment is a job-training program and Y is post-program earning, an outcome oriented protocol choice will aim to maximize mean earnings. In contrast, a general covariate-weighted protocol will seek to maximize mean weighted earnings where the weights vary with covariate values, such as race or gender.

If the planner wants to maximize the possibly covariate-weighted mean outcome, her problem can be written as:

$$\max_{p(\cdot)} \int h(w) \left[p(w) E(Y_1|W = w) + \{1 - p(w)\} E(Y_0|W = w) \right] dF_W(w), \tag{1}$$

s.t.

$$\int \left[p(w) E(C_1|W = w) + \{1 - p(w)\} E(C_0|W = w) \right] dF_W(w) \leq c, \tag{2}$$

where the non-negative function $h(w)$ represents the welfare weight attached to a w-type subject. A purely outcome-oriented planner would set $h(\cdot)$ equal to a constant. A non-constant $h(\cdot)$

means that a subject's welfare is judged not just by her/his outcome but also her characteristics which is analogous to non-statistical discrimination or taste-based allocation.

We now state a theorem that characterizes the efficient treatment protocol. For this, we impose the following conditions on the problem.

Assumptions: (i) $C_0 \geq 0$ and $\Delta C > \bar{A} > 0$, w.p. 1, $h(w) > 0$ for all w, (ii) $E[C_0] < c$.

Assumption (i) says that treatment 1 is more expensive for everyone, (ii) says that the budget constraint is such that giving everyone treatment 0 will leave some budget unspent. These assumptions correspond to realistic situations faced by policy-makers. The budget constraint may or may not be binding in the optimal assignment situation, although binding budgets are probably more common in the real world.

Define the constant γ as

$$\begin{aligned} \gamma &: = \max\{\gamma^*, 0\}, \text{ where} \\ \gamma^* &= \inf\left\{ a : \int \left[\begin{array}{c} E(C_0|w) \times 1\{h(w) E(\Delta Y|W=w) \leq aE(\Delta C|W=w)\} \\ +E(C_1|w) \times 1\{h(w) E(\Delta Y|W=w) > aE(\Delta C|W=w)\} \end{array} \right] dF_W(w) \leq c \right\}. \end{aligned}$$

Such a γ^* must exist by condition (ii) since taking a to $+\infty$ will incur average cost equal to $E(C_0)$ which is strictly less than c. Therefore γ is a non-negative bounded constant.

Theorem 1 *Under conditions (i)–(ii), the unique solution to the problem*

$$\max_{p(\cdot)\in[0,1]} \int h(w) \left(p(w) E(Y_1|W=w) + (1-p(w)) E(Y_0|W=w) \right) dF_W(w),$$

s.t.

$$\int \{p(w) E(C_1|W=w) + \{1-p(w)\} E(C_0|W=w)\} dF_W(w) \leq c,$$

is of the form

$$p^*(w) = \begin{cases} 1 \text{ if } h(w) E(\Delta Y|W=w) > \gamma E(\Delta C|W=w), \\ q \text{ if } h(w) E(\Delta Y|W=w) = \gamma E(\Delta C|W=w), \\ 0 \text{ if } h(w) E(\Delta Y|W=w) < \gamma E(\Delta C|W=w), \end{cases}$$

Under conditions (i)–(ii), the solution to the problem

$$\max_{p(\cdot)\in[0,1]} \int h(w) \left(p(w) E(Y_1|W=w) + \int (1-p(w)) E(Y_0|W=w) \right) dF_W(w),$$

s.t.

$$\int \{p(w) E(C_1|W=w) + \{1-p(w)\} E(C_0|W=w)\} dF_W(w) \leq c,$$

is of the form

$$p^*(w) = \begin{cases} 1 \text{ if } h(w) E(\Delta Y|W=w) > \gamma E(\Delta C|W=w), \\ q \text{ if } h(w) E(\Delta Y|W=w) = \gamma E(\Delta C|W=w), \\ 0 \text{ if } h(w) E(\Delta Y|W=w) < \gamma E(\Delta C|W=w), \end{cases}$$

where $q \in [0,1]$ *satisfies*

$$\int \left(\begin{array}{c} \left\{ \begin{array}{c} 1(h(w) E(\Delta Y|W=w) > \gamma E(\Delta C|W=w)) \\ +q1(h(w) E(\Delta Y|W=w) = \gamma E(\Delta C|W=w)) \end{array} \right\} \times E(C_1|w) \\ +1\{h(w) E(\Delta Y|W=w) < \gamma E(\Delta C|W=w)\} \times E(C_0|w) \end{array} \right) dF_W(w) = c,$$

if

$$\Pr\{h(w) \times E(\Delta Y|W=w) = \gamma E(\Delta C|W=w)\} > 0,$$

and is equal to zero otherwise. If the budget constraint binds, then $\gamma = \gamma^* > 0$*; if the budget does not bind, then* $\gamma^* \leq 0$ *and* $\gamma = 0$*.*

In particular, if $h(W) E(\Delta Y|W)$ has a positive Lebesgue density on an open interval around γ, then

$$\Pr(h(W) E(\Delta Y|W) > \gamma E(\Delta C|W)) = \Pr(h(W) E(\Delta Y|W) \geq \gamma E(\Delta C|W)) = c$$

and $q = 0$. Then $p^*(w) = 1(h(W) E(\Delta Y|W) \geq \gamma E(\Delta C|W))$.

Proof. See appendix. ■

The theorem basically says that the planner should order individuals by their values of $h(W) E(\Delta Y|W)$ relative to $E(\Delta C|W)$ and first give treatment 1 to those values of W where $h(W) E(\Delta Y|W)$ is the largest relative to $E(\Delta C|W)$, then to those for whom it is the next largest and so on till the budget is exhausted. If the distribution of $h(W) E(\Delta Y|W) / E(\Delta C|W)$ has point masses, then there could be a tie at the margin, which is then broken by randomization (hence the probability q). In the absence of any point masses, the optimal protocol is of a simple threshold-crossing form.

Remark 1 *Under the homogeneous cost, capacity constraint model, the above result can be specialized to the following statement. This corresponds to the setting in our healthcare application*

below. In the homogeneous treatment cost case under capacity constraints, satisfying conditions (Ci) $C_0 = 0$ *and* $C_1 = 1$ *w.p. 1,* $h(w) > 0$ *for all* w*, (Cii)* $0 < c$*, the optimization problem*

$$\max_{p(\cdot)\in[0,1]} \int h(w) \times [(p(w) E(Y_1|W = w) + (1 - p(w)) E(Y_0|W = w))] dF_W(w),$$

s.t.

$$\int p(w) dF_W(w) \leq c,$$

has a unique solution of the form

$$p^*(w) = \begin{cases} 1 \text{ if } \beta(w) > \gamma, \\ q \text{ if } \beta(w) = \gamma, \\ 0 \text{ if } \beta(w) < \gamma, \end{cases}$$

where

$$\begin{aligned} \beta(w) &: = h(w) E(\Delta Y|W = w), \\ \gamma &: = \max\left\{\inf\left\{a : \int 1\{\beta(w) > a\} dF_W(w) \leq c\right\}, 0\right\}, \end{aligned}$$

and $q \in [0, 1]$ *satisfies*

$$\int (\{1(\beta(w) > \gamma) + q1(\beta(w) = \gamma)\}) dF_W(w) = c,$$

when $\gamma > 0$ *and is zero otherwise.*

In the rest of the paper, we will hold c fixed and suppress the dependence of γ on c in our notation.

Discussion: Defining the conditional return to treatment as

$$\mu(w) \equiv \frac{E[\Delta Y|W = w]}{E[\Delta C|W = w]},$$

it is easy to see that when $h(\cdot)$ is a constant, the solution of theorem 1 is of the form $1\{\mu(w) \geq \gamma\}$. So in this purely outcome oriented case, type w is treated when $\mu(w)$ (weakly) exceeds the fixed threshold γ but for the taste-based case, the corresponding threshold for $\mu(\cdot)$, i.e., $\frac{\gamma}{h(w)}$ varies by w and is lower for those w's whose outcomes are more important to the planner. In either case, the threshold represents the return to treatment for the marginal treatment recipient; in the outcome-based case, it stays constant across covariates W but in the taste-based case, it

varies with W. Thus, a test of taste-based assignment can be based on comparing the treatment thresholds for different covariate-groups and testing if they are equal. However, due to selection on the unobservables Z, the marginal treatment recipient and consequently the treatment threshold cannot be identified. We now show how certain inequalities implied by efficient treatment assignment may be useful for detecting taste-based allocation.

Testable Inequalities: First consider the outcome-oriented case. The solution above implies that treatment status D is a deterministic function of W. Accordingly, for $j = 0, 1$, let $\mathcal{W}^j = \{w \in \mathcal{W} : D(w) = j\}$. Let

$$\begin{aligned} \mathcal{X}^j &= \left\{x : (x,z) \in \mathcal{W}^j \text{ for some } z \in \mathcal{Z}\right\} \\ &\equiv \left\{x : \Pr(D = j | X = x) > 0\right\}. \end{aligned}$$

Since the planner 's subjective expectations are assumed to be consistent with true distributions in the population, we must have that

$$\begin{aligned} E(\Delta Y | x, z) &\geq \gamma E(\Delta C | x, z), \text{ for all } (x,z) \in \mathcal{W}^1, \\ E\left(\Delta Y | x', z'\right) &\leq \gamma E\left(\Delta C | x', z'\right) \text{ for all } \left(x', z'\right) \in \mathcal{W}^0. \end{aligned} \tag{3}$$

Note that these inequalities are strict when there is no fractional treatment allocation.

Now, since we do not observe Z, the inequalities in (3) are not of immediate use to us. However, an implication of (3), obtained by "integrating out" z, is potentially useful for detecting taste-based allocation. Indeed, (3) implies that for (almost) every $x \in \mathcal{X}^1$,

$$\begin{aligned} &\int_{z:D(x,z)=1} E(\Delta Y | X = x, Z = z)\, dF_{Z|X=x,D(x,Z)=1}(z|x) \\ \geq\ & \gamma \int_{z:D(x,z)=1} E(\Delta C | X = x, Z = z)\, dF_{Z|X=x,D(x,Z)=1}(z|x), \end{aligned}$$

i.e.

$$E[\Delta Y - \gamma \Delta C | D = 1, X = x] \geq 0, \text{ for all } x \in \mathcal{X}^1, \tag{4}$$

and similarly

$$E[\Delta Y - \gamma \Delta C | D = 0, X = x] \leq 0, \text{ for all } x \in \mathcal{X}^0. \tag{5}$$

In words, if the planner is outcome-oriented, then the net benefit from treatment for every subgroup (that the planner can observe) among the treatment recipients must weakly exceed the treatment threshold. Since this would have to hold for every subgroup among the treated,

it must also hold for groups (observed by us) constructed by aggregating these subgroups and averaging the gain across those subgroups. This leads to (4) and analogously for (5). This reasoning lets us overcome the problem posed by the planner observing more covariates than us and preserves the inequality needed for inference.

It follows now that if for some $a \neq b$, we have that

$$\frac{E\left[\Delta Y|D=0,X=b\right]}{E\left[\Delta C|D=0,X=b\right]} > \frac{E\left[\Delta Y|D=1,X=a\right]}{E\left[\Delta C|D=1,X=a\right]},$$

then we conclude that there is misallocation in terms of the mean outcome in a way that hurts type b people.

Counterfactuals: To be able to use the above inequalities to learn about γ, we need to identify the counterfactual mean outcomes $E\left(Y_0|X,D=1\right)$ and $E\left(Y_1|X,D=0\right)$ and the counterfactual mean costs $E\left(C_0|X,D=1\right)$ and $E\left(C_1|X,D=0\right)$. As outlined above, we propose identifying these means by supplementing the observational dataset with estimates from an experiment, where individuals are randomized in and out of treatment. If the observational and the experimental samples are drawn from the same population – an assumption we maintain – then combining them will yield the necessary counterfactual distributions. To see this, notice that for any $x \in \mathcal{X}^1$,

$$\begin{aligned}
\underbrace{P\left(Y_0 \leq y|X=x\right)}_{\text{known from expt}} &= P^{obs}\left(Y_0 \leq y|X=x\right) \\
&= P^{obs}\left(Y_0 \leq y|D=1,X=x\right) \times \underbrace{P^{obs}\left(D=1|X=x\right)}_{\text{known from obs}} \\
&\quad + \underbrace{P^{obs}\left(Y_0 \leq y|D=0,X=x\right)}_{\text{known from obs}} \times \underbrace{P^{obs}\left(D=0|X=x\right)}_{\text{known from obs}}. \qquad (6)
\end{aligned}$$

Similarly for any $x \in \mathcal{X}^0$,

$$\begin{aligned}
\underbrace{\Pr\left(Y_1 \leq y|x\right)}_{\text{known from expt}} &= \Pr\left(Y_1 \leq y|D=0,x\right) \times \underbrace{\Pr\left(D=0|x\right)}_{\text{known from obs}} \\
&\quad + \underbrace{\Pr\left(Y_1 \leq y|D=1,x\right)}_{\text{known from obs}} \times \underbrace{\Pr\left(D=1|x\right)}_{\text{from obs}}. \qquad (7)
\end{aligned}$$

Thus the two equalities above yield the counterfactual distributions $P\left(Y_0 \leq y|D=1,x\right)$ on $\mathcal{X}^1$ and $P\left(Y_1 \leq y|D=0,x\right)$ on $\mathcal{X}^0$. When we know the means but not the distribution of Y_1 and Y_0 from the experiment, we have to replace the c.d.f.'s in the previous displays by the corresponding

means, giving us, for instance, for any $x \in \mathcal{X}^0$,

$$\begin{aligned}\underbrace{E\left(Y_1|x\right)}_{\text{known from expt}} &= E\left(Y_1|D=0,x\right) \times \underbrace{\Pr\left(D=0|x\right)}_{\text{known from obs}} \\ &\quad + \underbrace{E\left(Y_1|D=1,x\right)}_{\text{known from obs}} \times \underbrace{\Pr\left(D=1|x\right)}_{\text{from obs}}.\end{aligned}$$

Bounds: Combining (4), (5), (6) and (7) yield the following bounds on γ:

$$\begin{aligned}\gamma_{lb} &= \sup_{x\in\mathcal{X}^0}\left(\frac{\underbrace{E\left(Y_1|X=x,D=0\right)}_{\text{from (6)}} - \underbrace{E\left(Y_0|X=x,D=0\right)}_{\text{from obs data}}}{\underbrace{E\left(C_1|X=x,D=0\right)}_{\text{from (6)}} - \underbrace{E\left(C_0|X=x,D=0\right)}_{\text{from obs data}}}\right), \\ \gamma_{ub} &= \inf_{x\in\mathcal{X}^1}\left(\frac{\underbrace{E\left(Y_1|X=x,D=1\right)}_{\text{from obs data}} - \underbrace{E\left(Y_0|X=x,D=1\right)}_{\text{from (7)}}}{\underbrace{E\left(C_1|X=x,D=1\right)}_{\text{from obs data}} - \underbrace{E\left(C_0|X=x,D=1\right)}_{\text{from (7)}}}\right). \end{aligned} \tag{8}$$

The bounds derived above essentially replace a minimum over finer subgroups (observed by the planner) by the minimum over groups (observed by us) of the subgroup averages. So one would expect the bounds to be wider when (i) the unobserved covariates have larger support making the average across subgroups further from the minimum or maximum across subgroups, and (ii) the observed covariates are correlated with the unobserved ones to a lesser extent.

Simplified calculation: Observe that the lower bound calculation reduces to

$$\begin{aligned}& E\left(\Delta Y|D=0,X\right) \\ =\ & \frac{E\left(Y_1|X\right) - E\left(Y_1|D=1,X\right)\Pr\left(D=1|X\right)}{\Pr\left(D=0|X\right)} - E\left(Y_0|X,D=0\right) \\ =\ & \frac{E\left(Y_1|X\right) - E\left(Y_1|X,D=1\right)\Pr\left(D=1|X\right) - \Pr\left(D=0|X\right)E\left(Y_0|X,D=0\right)}{\Pr\left(D=0|X\right)} \\ =\ & \frac{E\left(Y_1|X\right) - E\left(DY_1|X\right) - E\left(\left(1-D\right)Y_0|X\right)}{\Pr\left(D=0|X\right)} = \frac{E\left(Y_1|X\right) - E\left(Y|X\right)}{\Pr\left(D=0|X\right)}.\end{aligned}$$

Similarly, for the upper bound,

$$E\left(\Delta Y|X,D=1\right) = \frac{E\left(Y|X\right) - E\left(Y_0|X\right)}{\Pr\left(D=1|X\right)}.$$

The bounds are then easily calculated as

$$\begin{aligned}\gamma_{ub} &= \inf_{x\in\mathcal{X}^1}\left\{\frac{E^{obs}\left(Y|X=x\right) - E^{\exp}\left(Y_0|X=x\right)}{E^{obs}\left(C|X=x\right) - E^{\exp}\left(C_0|X=x\right)}\right\} \\ \gamma_{lb} &= \sup_{x\in\mathcal{X}^0}\left\{\frac{E^{\exp}\left(Y_1|X=x\right) - E^{obs}\left(Y|X=x\right)}{E^{\exp}\left(C_1|X=x\right) - E^{obs}\left(C|X=x\right)}\right\}.\end{aligned}$$

Alternative designs and data issues: There are two different ways to perform the data combination exercise. In the first, the observational micro-data are combined with estimates obtained from an experimental study, conducted by other researchers. For this, one has to make sure that the observational group and the experimental group were drawn from the same population (see section 4.3 below for how the analysis can be modified when this assumption fails) and the same covariates were recorded in both cases.

A second possibility is to actually run an experiment, which can also be done in two ways. In the first, a sample of individuals is randomly divided into an experimental arm and a non-experimental one. The experimental arm individuals are randomly assigned to treatment and the observational arm ones are handed over to a planner who uses his/her discretion. This design was used in the CASS (1981) study in the US for studying the efficacy of coronary artery surgery. This is the set-up used to derive (6) and (7) above, motivated by our empirical application. The second way is as follows. First, present all the individuals to the planner and record his recommendations for treatment. This recommendation is recorded as $D = 1$ when recommended to have treatment and as $D = 0$, otherwise. Then we randomize actual approval across all applications (ignoring the planner 's recommendation) and observe the outcomes for each individual. The counterfactual $P(Y_0|D = 1, X)$ can then be obtained directly (i.e. without using (6) and (7)) from the outcomes of those who are approved by the planner but were randomized out of treatment. Conversely for $P(Y_1|D = 0, X)$.

The experimental approach requires significantly more work to implement but gives us the ideal set-up where the experimental and observational groups are ex-ante identical and the same variables can be recorded for both groups. The first method, where experimental results from existing studies are used instead of actually running an experiment, is applicable in many more situations. However, one is somewhat constrained by the outcomes and covariates that the original researchers had chosen.

In empirical microeconomic studies, the use of field experiments has now become widespread. Consequently, collecting the type of data outlined in the last but one paragraph (the second way) should be logistically straightforward. Such combined data, as we have tried to show above, are widely useful for inferring how treatments are assigned by real planners and what implicit covariate weighting justifies the observed treatment patterns.[5]

[5]The latter type of question was also previously addressed in the optimal tax literature by Stern (1987).

3.1 Misallocation

The bounds analysis presented above can be used to test whether there is misallocation of treatment both within and between demographic groups. To fix ideas, suppose $X = (X_1, female)$ and we are interested in testing if there is treatment misallocation within males and within females and then we want to test if treatment misallocation between males and females occurs in a way that hurts, say, females.

To do these tests, perform the above analysis separately for females and males and get the bounds

$$\Gamma_{fem} = \begin{pmatrix} \sup_{x \in Supp(X_1|fem=1,D=0)} \frac{E[\Delta Y|X_1=x,fem=1,D=0]}{E[\Delta C|X_1=x,fem=1,D=0]}, \\ \inf_{x \in Supp(X_1|fem=1,D=1)} \frac{E[\Delta Y|X_1=x,fem=1,D=1]}{E[\Delta C|X_1=x,fem=1,D=1]} \end{pmatrix}$$

and analogously Γ_{male}. Now, if Γ_{fem} (or Γ_{male}) is empty, then we conclude that there is misallocation within females (males). Further, if $\Gamma_{fem} \cap \Gamma_{male}$ is empty, then it implies that different thresholds were used for females and males and thus there is misallocation between males and females.

Intuition: To see why empty sets imply misallocation, notice that $\Gamma_{fem} \cap \Gamma_{male} = \phi$ means that there exist values a, b of X_1 such that either

$$\frac{E[\Delta Y|fem=0,D=1,X_1=a]}{E[\Delta C|fem=0,D=1,X_1=a]} < \frac{E[\Delta Y|fem=1,D=0,X_1=b]}{E[\Delta C|fem=1,D=0,X_1=b]}, \tag{9}$$

or

$$\frac{E[\Delta Y|fem=1,D=1,X_1=b]}{E[\Delta C|fem=1,D=1,X_1=b]} < \frac{E[\Delta Y|fem=0,D=0,X_1=a]}{E[\Delta C|fem=0,D=0,X_1=a]}. \tag{10}$$

The first inequality (9) means that the return to treatment among one subgroup (defined by $X_1 = a$) of treated males is less than that among one subgroup of untreated females (viz., those with $X_1 = b$) – i.e., b-type females are facing a higher threshold than a-type males. Similarly, (10) means that b-type males are being under-treated relative to a-type females.

Notice that the inequalities (10) or (9) can be interpreted and used directly without reference to a specific model of optimization or treatment allocation such as (2). However, the link with (1) gives our analysis a firm grounding in classical economic theory of choice under uncertainty.

Finally, it is worth pointing out that inequality (9) is fundamentally different from

$$\frac{E[\Delta Y|fem=0,D=1,X_1=a]}{E[\Delta C|fem=0,D=1,X_1=a]} < \frac{E[\Delta Y|fem=1,D=1,X_1=b]}{E[\Delta C|fem=1,D=1,X_1=b]}, \tag{11}$$

Instead of focusing on the design of an optimal tax structure which depended on subjective welfare weights, Stern investigated what underlying welfare weights would justify the existing income tax schedules.

i.e., we are comparing treated females with untreated males and not comparing treated females with treated males. The latter will in general reveal nothing about the sign of $\gamma_{male} - \gamma_{female}$ because of the so-called "inframarginality" problem (c.f., Persico (2009)), viz. all we know is that the LHS of (11) is larger than γ_{male} and the RHS is larger than γ_{fem}.

4 Interpretation and robustness

Our method of detecting non-outcome based treatment allocation is based on comparisons of expected returns across a specific covariate or set of covariates X and is valid under a specific substantive assumption of identical distributions. This raises two question about the interpretation and applicability of our methods, viz., (i) whether our method can pinpoint the true source of misallocation and (ii) how robust are our methods to the failure of the identical distribution assumption (assumption 5) stated in the introduction.

The first question is addressed using two illustrative scenarios. They illustrate why the discrepancy detected through the violation of inequalities, as described above, may arise from discrimination based on covariates "related to" X but not X itself. The implication is that when we have detected misallocation using X, we can conclude that there indeed is misallocation of treatment but that misallocation could have resulted from sources other than explicit prejudice of the planner against one or more subgroups defined through X.

Remark 2 *It should be noted that the notion of implicit discrimination (section 4.1) and the issue of compositional effects (section 4.2) discussed below, which prevent us from making a simplistic "prejudicial" interpretation, are not unique to our methodology and equally apply to all of the existing tests in the literature. For example, researchers who try to detect racial bias in law enforcement or credit approval do not consider whether differential gender composition by race can lead to misinterpretation of gender bias as racial bias. In other words, our method – like all other existing methods – can reject outcome-oriented efficiency of treatment assignment but cannot pinpoint the behavioral source of that inefficiency. Taste-based motives or explicit prejudice are sufficient but necessary for causing such discrepancy, as we discuss below. This distinction is very different from the well-recognized distinction between taste-based and statistical discrimination and, to the best of our knowledge, has been largely ignored in the existing literature.*

Remark 3 *The second question discussed in subsection 4.3, viz. that of robustness of our methods to the failure of identical distribution between the observational and experimental or quasi-experimental dataset is indeed unique to our methodology.*

4.1 Implicit discrimination

Suppose the two groups of interest are the rich and the poor. Assume identical treatment costs for now and suppose we detect an inequality of type (9):

$$E[\Delta Y|poor = 0,D = 1] < E[\Delta Y|poor = 1,D = 0],$$

which suggests that there is taste-based treatment assignment that hurts the poor. It is possible that this is brought about by a planner who practices taste-based discrimination against blacks but is not necessarily biased against the poor. The following scenario illustrates the point. Suppose it is the case that for two constants $\lambda_{bl} > \lambda_{wh}$, we have

$$\begin{aligned} E\left(\Delta Y|black,rich\right) &> \lambda_{bl} > E\left(\Delta Y|black,poor\right) \\ &> E\left(\Delta Y|white,rich\right) > E\left(\Delta Y|white,poor\right) > \lambda_{wh}. \end{aligned}$$

Suppose the planner observes both race and wealth status and thus assigns the rich blacks and all whites to treatment. Then we have that

$$\begin{aligned} E\left(\Delta Y|poor,D=0\right) &= E\left(\Delta Y|poor,black\right) \\ E\left(\Delta Y|rich,D=1\right) &= E\left(\Delta Y|rich,black\right)\times \Pr\left(black|rich\right) \\ &\quad + E\left(\Delta Y|rich,wh\right)\times \Pr\left(wh|rich\right) \\ &\simeq E\left(\Delta Y|rich,wh\right) \text{ if } \Pr\left(wh|rich\right)\simeq 1. \end{aligned}$$

Since it is the case that

$$E\left(\Delta Y|black,poor\right) > E\left(\Delta Y|white,rich\right),$$

we will conclude that

$$E\left(\Delta Y|poor,D=0\right) > E\left(\Delta Y|rich,D=1\right),$$

i.e., that there is misallocation which works against the poor. This will happen even if the DM is not explicitly discriminating against the poor. The root is of course the high positive correlation

between being white and rich. This shows that detecting misallocation that hurts a group we chose to test may not imply that the planner is practising taste-based allocation where taste dictates him to be biased for or against the characteristics which define the group chosen by us – it could arise from intentional discrimination against a positively correlated characteristic.

4.2 Inadvertent Discrimination

In this subsection, we provide another illustration which adds a second cautionary note to the interpretation of our results. This illustration also clarifies the role of the second assumption of the introduction – i.e., the researcher observes fewer covariates than the planner. For simplicity of exposition, we will assume here that ΔC is a constant equal to 1 (i.e., does not vary with any component of W), so that efficiency would imply that $D = 1 \Rightarrow E(\Delta Y|W) > \lambda$, where $\lambda = 1/\gamma$.

Suppose individuals are characterized by race (black/white) and gender (male/female). Suppose it is the case that

$$\begin{aligned} E(\Delta Y|fem, black) &> E(\Delta Y|male, White) > E(\Delta Y|male, black) \\ &> \lambda \\ &> E(\Delta Y|fem, white). \end{aligned} \tag{12}$$

Assume that the fraction of whites among women is high enough that

$$E(\Delta Y|male) > \lambda > E(\Delta Y|female). \tag{13}$$

That is, black females benefit a lot from treatment while females benefit the least. If white females are a much larger group than black females, then on average, females benefit less from treatment and hence (13) holds.

Now suppose the planner ignores race and allocates treatment, based only on gender. Then $D = 1$ iff the individual is male and so it must be the case that

$$\begin{aligned} E(\Delta Y|D &= 0, Black) \\ &= E(\Delta Y|female, Black) \\ &> E(\Delta Y|male, White), \text{ by (12)} \\ &= E(\Delta Y|D = 1, White). \end{aligned} \tag{14}$$

Thus, we would conclude that there is misallocation which works against blacks *precisely because the planner is race-blind in his decision-making.*

Notice that this set-up violates the second assumption of the introduction. Here we observe race but the planner does not take into account race in making the allocation. This works against black females as they are treated the same as white females because of their gender. The scenario described here is quite stark in that we are detecting misallocation by race precisely because the planner is *not* taking race into account in making the allocation. A researcher concluding from (14) that there is *prejudice* against blacks will thus be dramatically mistaken. Notice that this "mistake" is very different from and more subtle than the mistake of interpreting statistical discrimination as taste-based discrimination.

4.3 Nonidentical distributions

As a final caveat, we consider the possibility that the observational sample and the experimental sample were drawn from different subsets of the population. For example, sometimes it is the case in medical trials that inherently sicker patients agree to be randomized. In this case, it is reasonable to expect that $E^{\exp}(Y_0|x) \leq E^{obs}(Y_0|x)$ and $E^{\exp}(Y_1|x) \leq E^{obs}(Y_1|x)$. Similarly, $E^{\exp}(C_0|x) > E^{obs}(C_0|x)$ and $E^{\exp}(C_1|x) > E^{obs}(C_1|x)$. Using the same steps as those leading to (6), one gets that

$$\begin{aligned} E^{obs}(Y_0|D=1,x) &= \frac{E^{obs}(Y_0|x) - P^{obs}(D=0|x) \times E^{obs}(Y_0|D=0,x)}{P^{obs}(D=1|x)} \\ &\geq \frac{E^{\exp}(Y_0|x) - P^{obs}(D=0|x) \times E^{obs}(Y_0|D=0,x)}{P^{obs}(D=1|x)} \\ &\equiv \bar{E}(Y_0|D=1,x), \end{aligned}$$

and similarly,

$$\begin{aligned} E^{obs}(Y_1|D=0,x) &= \frac{E^{obs}(Y_1|x) - P^{obs}(D=0|x) \times E^{obs}(Y_1|D=0,x)}{P^{obs}(D=1|x)} \\ &\geq \frac{E^{\exp}(Y_1|x) - P^{obs}(D=0|x) \times E^{obs}(Y_1|D=0,x)}{P^{obs}(D=1|x)} \\ &\equiv \bar{E}(Y_1|D=0,x). \end{aligned}$$

The quantities $\bar{E}(Y_1|D=0,x)$ and $\bar{E}(Y_0|D=1,x)$ are clearly identified. An analogous set of inequalities hold with Y replaced by C and the inequality sign reversed (since the experimental

group, being sicker will be more expensive to treat). These bounds can be used to detect misallocation. For instance, if it is the case that

$$\begin{aligned} & \frac{E^{obs}\left(Y_1|D=1,male\right)-\bar{E}\left(Y_0|D=1,male\right)}{E^{obs}\left(C_1|D=1,male\right)-\bar{E}\left(C_0|D=1,male\right)} \\ \leq & \frac{\bar{E}\left(Y_1|D=0,female\right)-E^{obs}\left(Y_0|D=0,female\right)}{\bar{E}\left(C_1|D=0,female\right)-E^{obs}\left(C_0|D=0,female\right)}, \end{aligned} \tag{15}$$

then it follows that

$$\begin{aligned} \gamma_{male} & < \frac{E^{obs}\left(\Delta Y|D=1,male\right)}{E^{obs}\left(\Delta C|D=1,male\right)} \\ & \leq \frac{E^{obs}\left(Y_1|D=1,male\right)-\bar{E}\left(Y_0|D=1,male\right)}{E^{obs}\left(C_1|D=1,male\right)-\bar{E}\left(C_0|D=1,male\right)} \\ & \leq \frac{\bar{E}\left(Y_1|D=0,female\right)-E^{obs}\left(Y_0|D=0,female\right)}{\bar{E}\left(C_1|D=0,female\right)-E^{obs}\left(C_0|D=0,female\right)} \\ & \leq \frac{E^{obs}\left(\Delta Y|D=0,female\right)}{E^{obs}\left(\Delta C|D=0,female\right)} \\ & \leq \gamma_{fem}. \end{aligned}$$

Thus, γ_{female} is larger, meaning that females are facing a larger threshold relative to males. However, since (15) implies (9), it will be harder to detect misallocation here compared to when the experimental and observational data came from identical populations. In other words, when we have only dominance and not identical distributions, we would get wider bounds that would make it harder to detect inefficiency. But if inefficiency is detected with wider bounds, then it would also have been detected with narrower bounds and is therefore conclusive.

5 Risk aversion

We now extend the analysis to include risk averse behavior by the planner – an issue which, to our knowledge, has been largely ignored in the existing empirical literature on testing treatment fairness – and transform the problem of detecting misallocation for a specific outcome to the problem of detecting the extent of risk aversion which justify the observed allocation as an efficient one. Specifically, we ask what risk-averse utility function(s) are consistent with efficient outcome-based allocation, given the data. To do this we consider a family of risk averse utility functions $u\left(\cdot,\theta\right)$, indexed by a finite dimensional parameter θ and the corresponding allocation

rule which is a generalization of (2): i.e., for some $\lambda > 0$,

$$D = 1 \Longrightarrow \frac{E\left(u\left(Y_1, \theta\right)|X, Z\right) - E\left(u\left(Y_0, \theta\right)|X, Z\right)}{E\left(C_1|X, Z\right) - E\left(C_0|X, Z\right)} \geq \lambda. \tag{16}$$

Examples of such utility functions are CRRA $u\left(Y, \theta\right) \equiv \frac{Y^{1-\theta}}{1-\theta}$ for $\theta \in (0, 1)$ and CARA $u\left(Y, \theta\right) \equiv -e^{\theta Y}$ for $\theta \geq 0$. Let $\Delta Y\left(\theta\right) \equiv u\left(Y_1, \theta\right) - u\left(Y_0, \theta\right)$.

When the planner 's subjective expectations are consistent with true distributions in the population, we have that

$$\frac{E\left(u\left(Y_1, \theta\right)|X, D = 1\right) - E\left(u\left(Y_0, \theta\right)|X, D = 1\right)}{E\left(C_1|X, D = 1\right) - E\left(C_0|X, D = 1\right)} \geq \lambda, \text{ w.p.1.}$$

As before, we do the analysis separately for males and females to get the bounded sets in terms of θ:

$$\left[L_f\left(\theta\right), U_f\left(\theta\right)\right] = \left\{ \left(\begin{array}{c} \sup_{x \in Supp(X_1|fem=1, D=0)} \frac{E[\Delta Y(\theta)|X_1=x, fem=1, D=0]}{E[\Delta C|X_1=x, fem=1, D=0]} \\ \leq \lambda \\ \leq \inf_{x \in Supp(X_1|fem=1, D=1)} \frac{E[\Delta Y(\theta)|X_1=x, fem=1, D=1]}{E[\Delta C|X_1=x, fem=1, D=1]} \end{array} \right) \right\}$$

and similarly, $\left[L_m\left(\theta\right), U_m\left(\theta\right)\right]$.

So the values of θ consistent with efficient allocation within gender are the ones for which

$$L_f\left(\theta\right) \leq U_f\left(\theta\right) \text{ and } L_m\left(\theta\right) \leq U_m\left(\theta\right). \tag{17}$$

Further, the values of θ which are consistent with efficient allocation across gender are the ones for which

$$\max\left\{L_f\left(\theta\right), L_m\left(\theta\right)\right\} \leq \min\left\{U_f\left(\theta\right), U_m\left(\theta\right)\right\}. \tag{18}$$

If the set of θ for which both (17) and (18) hold turns out to be empty, then no member of the corresponding family of utility functions will justify the observed allocation as an efficient one.

We now present an empirical illustration of the methodology developed in section 3. The illustration is based on the Coronary Artery Surgery Study (CASS), conducted in early 1980's in the US. A detailed description of the study design and its findings is provided in the CASS paper cited below. Here we provide a brief overview. The purpose of the present illustration is to show how our method performs in a real dataset that has the data combination flavor. A more substantive empirical analysis of these data is being conducted by the present author in an ongoing collaborative project (c.f., reference 7 below).[6]

[6] The CASS data may be obtained through online request at https://biolincc.nhlbi.nih.gov/studies/cass/?q=CASS.

The goal of the CASS study was to evaluate the effectiveness of coronary artery surgery versus medical therapy in patients with mild to moderate angina. Patients with severe angina were excluded from the study–bypass surgery was already known to improve longevity in such patients. The design involved dividing the patients into a trial arm where patients were randomized into or out of surgery and a non-trial arm where they were assigned to surgery by physician discretion. The stated goal of this design, deduced ex-post by the present author from the research paper cited below, was to check if outcomes with and without the treatment were different in the experimental arm from that in the observational arm. The study did not find any appreciable difference and it is unclear to the present author as to what this conclusion implies. Nonetheless, the study design is ideal for the objective of the present paper and provides a useful dataset for illustrating the usefulness of the methodology developed above.

Specifically, in the CASS study, all patients undergoing coronary angiography in participating sites and who showed indication of suspected or proven coronary artery disease were entered into a registry (about n = 25,000). Out of these 2,100 were medically eligible for randomization ($<$ 65 years, mild to moderate angina, etc.) Out of these 2100, about 1320 patients were not randomized and are referred to as randomizable patients and they constitute our observational group. 780 patients were evenly randomized into medical or surgical arms– the "randomized" patients constituting the experimental group. The specific surgical (medical) therapy given to a surgical (medical) patient was decided by the physician attending to the case. The primary endpoints of the study included death and myocardial infarction (heart attack), and secondary endpoints included evaluation of angina and quality of life. About 17 years of follow-up data for vital status were included. Due to some cross-over in the long-run,[7] we will refer to being assigned to surgery as the treatment. Also, we choose only males for our analysis. Females constitute less than 10% of the study sample and race is not recorded.

Summary statistics for some key variables is provided in figure 1. The variable lvscor is an index for how well the heart functions, with $5-8$ being a normal range; previousmi is a dummy for whether the patient had a previous heart attack and smoking is a dummy for whether the patient is currently smoking. Our outcome variable of interest, labeled "death", is the binary indicator for whether the patient died within 17 years from the date of treatment assignment.In

[7] 31 of the 390 patients randomized into the surgical group refused surgery and utilized medical therapy instead and about $\frac{1}{4}$ of the 390 patients in the medical arm elected to undergo surgical therapy in the long run.

	Experimental		Observational		t-test
Variable	Mean	Std. Dev.	Mean	Std. Dev	p-value
death	0.36	0.48	0.34	0.47	
treatment	0.50	0.50	0.43	0.50	
unemployed	0.29	0.44	0.28	0.45	0.63
age	51.10	7.31	50.87	7.82	0.63
lvscor	7.55	2.90	7.46	2.96	0.54
previousmi	0.62	0.49	0.59	0.49	0.16
diabetes	0.09	0.28	0.06	0.24	0.04
stroke	0.01	0.12	0.01	0.10	0.52
smoking	0.40	0.49	0.42	0.47	0.46
N	704		1192		

Figure 1:

the ideal situation, the enrollment into the experimental arm should be random. However, in the CASS case, this seems to have been influenced to some extent by the physicians who were treating the patients. In terms of most observable characteristics however, the two groups seem very similar. They are very slightly different in terms of prior incidence of heart-attack and diabetes, for which the experimental group seems slightly sicker. As explained in the appendix part D, labelled "Non-identical distributions", this means that our bounds are still valid but wider. When we do detect different thresholds using these wider bounds, we would also have detected different thresholds under narrower bounds which would result if enrollment into the experimental arm were random.

In this application, we focus on testing efficiency in terms of survival outcomes. It seems unlikely that performing coronary artery surgery on AMI patients will involve a lot of variations in actual costs across patient types and capacity constraints are likely to be the key reason for rationing treatments. In this sense, the problem now corresponds to the setting of corollary 1, above.

As for the key covariates of interest, health insurance coverage is potentially a key factor affecting treatment status, especially since all individuals in this dataset are under 65 (and hence not covered by Medicare). While the dataset does not record HI status, we may regard employment status, which is recorded here, as a crude proxy for (employer-provided) HI cover.

In this context, we would expect that insurance considerations might lead the non-employed to receive the treatment less frequently than their potential health outcomes might dictate. This is the first hypothesis we seek to test.

Using survival gain due to surgery as the outcome of interest, our bounds are (recall corollary 1):

$$\begin{aligned}
\gamma_{lb} &= \sup_{x\in\mathcal{X}^0} E\left(Y_1 - Y_0 | D=0, X=x\right) = \sup_{x\in\mathcal{X}^0}\left\{\frac{E^{\exp}\left(Y_1|X=x\right) - E^{obs}\left(Y|X=x\right)}{\Pr^{obs}\left(D=0|X=x\right)}\right\}, \\
\gamma_{ub} &= \inf_{x\in\mathcal{X}^1} E\left(Y_1 - Y_0 | D=1, X=x\right) = \inf_{x\in\mathcal{X}^1}\left\{\frac{E^{obs}\left(Y|X=x\right) - E^{\exp}\left(Y_0|X=x\right)}{\Pr^{obs}\left(D=1|X=x\right)}\right\}.
\end{aligned}$$

We first consider the case where the groups of interest are the unemployed versus employed and use q quantiles of age to narrow the bounds. That is, calculate the sample analog of the following bounds: for the unemployed,

$$\begin{aligned}
\gamma_{lb}^{unem} &= \max_{x\in(1,...,q)}\left\{\frac{E^{\exp}\left(Y_1|age_q=x, unem=1\right) - E^{obs}\left(Y|age_q=x, unem=1\right)}{\Pr^{obs}\left(D=0|age_q=x, unem=1\right)}\right\}, \\
\gamma_{ub}^{unem} &= \min_{x\in(1,...,q)}\left\{\frac{E^{obs}\left(Y|age_q=x, unem=1\right) - E^{\exp}\left(Y_0|age_q=x, unem=1\right)}{\Pr^{obs}\left(D=1|age_q=x, unem=1\right)}\right\}.
\end{aligned}$$

Similarly, for the employed:

$$\begin{aligned}
\gamma_{lb}^{emp} &= \max_{x\in(1,...,q)}\left\{\frac{E^{\exp}\left(Y_1|age_q=x, unem=0\right) - E^{obs}\left(Y|age_q=x, unem=0\right)}{\Pr^{obs}\left(D=0|age=x, unem=1\right)}\right\}, \\
\gamma_{ub}^{emp} &= \min_{x\in(1,...,q)}\left\{\frac{E^{obs}\left(Y|age_q=x, unem=0\right) - E^{\exp}\left(Y_0|age_q=x, unem=0\right)}{\Pr^{obs}\left(D=1|age_q=x, unem=0\right)}\right\}.
\end{aligned}$$

The hypothesis of interest is whether $\gamma^{unem} > \gamma^{emp}$ which will be implied by $\gamma_{lb}^{unem} \geq \gamma_{ub}^{emp}$.

Methods of inference for such max or min type estimates are now well-known (c.f., Chernozhukov et al (2003), Rosen (2008), Andrews and Soares (2010) etc.). In the above example, we have for each $x = 1, ..., q$ the $4q$ moment inequalities:

$$\begin{aligned}
\gamma^{unem} - \frac{E^{\exp}\left(Y_1|age_q=x, unem=1\right) - E^{obs}\left(Y|age_q=x, unem=1\right)}{\Pr^{obs}\left(D=0|age_q=x, unem=1\right)} &\geq 0, \\
\frac{E^{obs}\left(Y|age_q=x, unem=1\right) - E^{\exp}\left(Y_0|age_q=x, unem=1\right)}{\Pr^{obs}\left(D=1|age_q=x, unem=1\right)} - \gamma^{unem} &\geq 0, \\
\gamma^{emp} - \frac{E^{\exp}\left(Y_1|age_q=x, unem=0\right) - E^{obs}\left(Y|age_q=x, unem=0\right)}{\Pr^{obs}\left(D=0|age=x, unem=1\right)} &\geq 0, \\
\frac{E^{obs}\left(Y|age_q=x, unem=0\right) - E^{\exp}\left(Y_0|age_q=x, unem=0\right)}{\Pr^{obs}\left(D=1|age_q=x, unem=0\right)} - \gamma^{emp} &\geq 0. \quad (19)
\end{aligned}$$

Using any of the methods cited above, one can obtain a joint confidence set C_n satisfying, say,

$$\lim_{n\to\infty} \Pr\left[\begin{pmatrix} \gamma^{unem} \\ \gamma^{emp} \end{pmatrix} \sqsubseteq C_n\right] \geq (1-\alpha).$$

For example, let $\theta = \begin{pmatrix} \gamma^{unem} \\ \gamma^{emp} \end{pmatrix}$ and the sample analog of the LHS expressions in (19) by the $4q$-vector $\bar{m}_n(\theta)$ with estimated covariance matrix $\hat{\Sigma}(\theta)$. Then Andrews and Soares (2010) suggest a confidence region $C_n = \{\theta : T_n(\theta) \leq c_{1-\alpha}(\theta)\}$ based on the test-statistic

$$T_n(\theta) = \inf_{t\in\mathbb{R}^{4q}_+} \left(\sqrt{n}\bar{m}_n(\theta) - t\right)' \hat{\Sigma}_n^{-1}(\theta) \left(\sqrt{n}\bar{m}_n(\theta) - t\right),$$

and using a critical value $c_{1-\alpha}(\theta)$ obtained through generalized moment selection, which "estimates" from the data which inequality constraints bind and uses those to calculate the critical value. In particular, for the jth moment inequality, one checks whether $\sqrt{n}\bar{m}_{nj}(\theta)/\hat{\sigma}_{nj}(\theta) > (\ln n)^{1/2}$, where $\hat{\sigma}_{nj}(\theta)$ is simply the estimated standard deviation of $\bar{m}_{nj}(\theta)$. The final critical value is then computed using bootstrap resamples of the data and using only those moments which could not be rejected as holding with equality (See Andrews and Soares (2010), sec 4.2, for details of computation).

A confidence set for the threshold difference is then given by

$$I_n = \left\{ a : a = \gamma^{unem} - \gamma^{emp}, \begin{pmatrix} \gamma^{unem} \\ \gamma^{emp} \end{pmatrix} \sqsubseteq C_n \right\}.$$

If this confidence interval does not contain zero, then one rejects the null hypothesis of equal thresholds and incurs a type 1 error probability bounded above by $(1-\alpha)$. These are the confidence intervals we compute below.

Given the simple nature of the above bounds, a referee has suggested calculating confidence intervals for the difference using one-sided confidence intervals for the individual bounds. For example, we know that $\gamma^{unem} \in [\gamma^{unem}_{lb}, \gamma^{unem}_{ub}]$ satisfying for all $x = 1, ...q$,

$$\gamma^{unem}_{lb} \geq m_{lx},\ \gamma^{unem}_{ub} \leq m_{ux},$$

where for $x = 1, ...q$,

$$\begin{aligned} m_{lx} &= \frac{E^{\exp}(Y_1|age\_q = x, unem = 1) - E^{obs}(Y|age\_q = x, unem = 1)}{\Pr^{obs}(D = 0|age\_q = x, unem = 1)}, \\ m_{ux} &= \frac{E^{obs}(Y|age\_q = x, unem = 1) - E^{\exp}(Y_0|age\_q = x, unem = 1)}{\Pr^{obs}(D = 1|age\_q = x, unem = 1)}. \end{aligned}$$

Choose constants c_{lx}, c_{ux} for $x = 1, ...q$, such that

$$\Pr\left[c_{lx} \geq \sqrt{n}\left(\hat{m}_{lx} - m_{lx}\right), c_{ux} \leq \sqrt{n}\left(\hat{m}_{ux} - m_{ux}\right)\right] \to (1-\alpha)^{1/q}.$$

Now since the sample analog estimates for m_{lx} for different values of x are asymptotically independent (since they are based on independent random samples drawn from different population strata indexed by x), we have for $\beta = (1-\alpha)^{1/q}$ that

$$\begin{aligned}
& \Pr\left[\cap_{x=1}^{q}\left\{m_{lx} \geq \hat{m}_{lx} - \frac{c_{lx}}{\sqrt{n}}, m_{ux} \leq \hat{m}_{ux} - \frac{c_{ux}}{\sqrt{n}}\right\}\right] \\
= & \Pr\left[\cap_{x=1}^{q}\left\{c_{lx} \geq \sqrt{n}\left(\hat{m}_{lx} - m_{lx}\right), c_{ux} \leq \sqrt{n}\left(\hat{m}_{ux} - m_{ux}\right)\right\}\right] \\
\to & \Pi_{x=1}^{q} \Pr\left[c_{lx} \geq \sqrt{n}\left(\hat{m}_{lx} - m_{lx}\right), c_{ux} \leq \sqrt{n}\left(\hat{m}_{ux} - m_{ux}\right)\right] \\
\to & \beta^q = (1-\alpha).
\end{aligned}$$

Set

$$\hat{L}^{unem} = \sup_{x=1,...,q}\left\{\hat{m}_{lx} - \frac{c_{lx}}{\sqrt{n}}\right\}, \hat{U}^{unem} = \inf_{x=1,...,q}\left\{\hat{m}_{ux} - \frac{c_{ux}}{\sqrt{n}}\right\}.$$

Then we have

$$\begin{aligned}
& \Pr\left(\hat{L}^{unem} \leq \gamma^{unem} \leq \hat{U}^{unem}\right) \\
\geq & \Pr\left(\hat{L}^{unem} \leq \gamma_{lb}^{unem}, \hat{U}^{unem} \geq \gamma_{ub}^{unem}\right) \\
= & \Pr\left[\cap_{x=1}^{q}\left\{m_{lx} \geq \hat{m}_{lx} - \frac{c_{lx}}{\sqrt{n}}, m_{ux} \leq \hat{m}_{ux} - \frac{c_{ux}}{\sqrt{n}}\right\}\right] \to 1-\alpha.
\end{aligned}$$

This yields $\left(\hat{L}^{unem}, \hat{U}^{unem}\right)$ as a $(1-\alpha)$-level confidence interval for γ^{unem}. Similarly, one can get a $(1-\alpha)$-level confidence interval $\left(\hat{L}^{emp}, \hat{U}^{emp}\right)$ for γ^{emp}. Then, by independence, one gets the confidence set $\left(\hat{L}^{unem} - \hat{U}^{emp}, \hat{U}^{unem} - \hat{L}^{emp}\right)$ for $\gamma^{unem} - \gamma^{emp}$ with level $(1-\alpha)^2$. So if we start with $\alpha = \sqrt{0.95} = 0.974$, then we will get a 95% asymptotic level confidence region eventually. However, note that this method essentially provides a confidence interval not for $\gamma^{unem} - \gamma^{emp}$ but for $\gamma_{lb}^{unem} - \gamma_{ub}^{emp}$, whereas the Andrews and Soares method provides a CI for $\gamma^{unem} - \gamma^{emp}$, which is likely to be tighter.

The following table reports a two-sided 95% confidence interval, using the Andrews and Soares method (2010) for the differences in threshold values for two different choices of group-identity. A confidence interval that contains only positive values indicates that the first group is facing a higher threshold for treatment– i.e., there is taste-based allocation against the first group.

95% CI for threshold differences

q	nonemployed-employed	smoker-nonsmoker
2	(-0.18, 0.163)	(-0.051, 0.195)
10	(-0.151, -0.004)	(0.012, 0.292)

Figure 2:

In column 2, we report the results corresponding to group identity by employment status and for $q = 2$ and $q = 10$. When using $q = 2$, we have higher sampling precision but also a higher extent of non-identification; a larger q lowers estimate precision but brings us closer to the true threshold. Specifically, Column 2 of the table suggests that the treatment threshold for the employed is *higher* than that for the non-employed, contrary to our original hypothesis. One possible reason for this is that invasive procedures require longer recovery periods which may be easier to implement when the person is not in employment. Further, as we hinted above, not being employed does not necessarily mean that the individual has no health insurance coverage– they may receive Medicaid or be covered through a spouse's job.

In column 3, we report the results where unemployment status is replaced by smoking status. The hypothesis of interest is that smokers are set a higher threshold in survival gains in order to qualify for treatment. The graph below[8] plots the estimated lower bound for smokers, the upper bound for nonsmokers and the difference between the two for different deciles of age (the first decile is 40 years, the median is 52 and the 9th decile is 61).These curves are based on sample analog estimates; so they are consistent but biased in finite samples (c.f. Manski and Pepper (2000)) and the dip of the upper bound curve below zero is likely arising from the well-known downward finite sample bias for estimates defined as the maximum of several underlying estimates. Nonetheless, the graph shows some suggestive evidence that the threshold difference exists at all ages. Combining with the confidence interval of the previous table, we can conclude that smokers are indeed facing higher thresholds. It is conceivable that this happens due to worse "quality" of life for smokers or because they are likely to suffer from heart-attack in the

[8] We thank an anonymous referee for suggesting this.

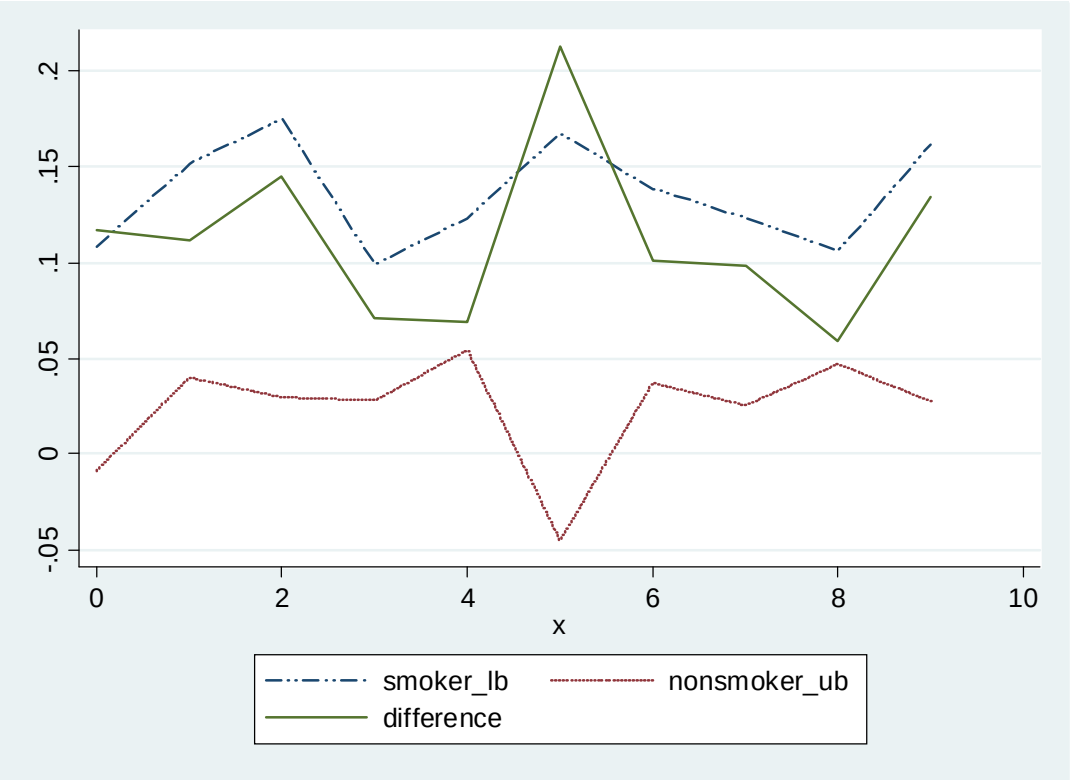

Figure 3:

q=10, outcome=-exp(-y*theta)

Theta	95% CI for smoker-nonsmoker
0.001	(0.09, 0.174)
0.002	(0.05, 0.207)
0.003	(0.013, 0.092)
0.004	(-0.01,0.068)
0.005	(-0.033, 0.143)
0.01	(-0.027, 0.0003)

Figure 4:

future thereby raising costs.

A third plausible explanation, as suggested by a referee, is that smokers who are less health-conscious and the employed who have higher opportunity cost of receiving invasive treatments (requiring longer hospital stays) opt for less invasive procedures. Although this shifts some of the allocational inefficiencies away from the physicians' responsibilities toward patient choice, the evidence of *net* inefficiency is still present, irrespective of which direction it arises from.

Bounds on risk aversion: We now use the above data to find bounds on the risk aversion coefficient in a CARA family of utilities, $u(y) = -e^{-\theta y}$, $\theta > 0$ as outlined in section 5.1 above. To do this, we now define the outcome as the number of days of survival after the medical/surgical procedure which is capped above at 17 years. Our key regressor of interest is smoking status, as in the previous table. Using values of $\theta \in [0.001, 0.01]$, we check if the inequalities in (17) and (18) are satisfied but, as in the above exercises, we use means across the deciles of age, rather than minimum and maximum over those deciles. The purpose of this exercise is to see for what values of the risk aversion parameter θ, can the observed treatment assignment be justified as having arisen from (expected) utility maximization.

The results are shown in figure 4. For each value of θ, we report the 95% confidence interval for the difference in thresholds for smokers from that of non-smokers. A CI containing only positive values indicates that for the given value of θ, smokers are facing a higher threshold.The table shows that when the value of the planner's risk aversion parameter θ exceeds about 0.004, we cannot reject the hypothesis that smokers and nonsmokers are facing the same expectational

threshold in order to get the treatment. However, for values of $\theta \leq 0.003$, one can reject the hypothesis of equal threshold in favor of the hypothesis that smokers face a higher threshold. As θ gets close to zero, the utility function moves towards a risk-neutral one and the implied threshold difference between smokers and non-smokers is then seen to increase. This suggests that the smokers' outcomes with the surgery are more uncertain so that higher degrees of risk-aversion eliminate the threshold differences observed under risk-neutrality.

6 Conclusion

The treatment effects literature in econometrics has, justifiably, focused on identifying impacts of the treatment from observational studies. However, when individuals are assigned to treatment by external agencies, the procedure of treatment assignment becomes a matter of significant political and social concern. This is especially the case when treatment rates vary significantly by socioeconomic characteristics, raising concerns about unfairness. Such concerns warrant a formal statistical analysis and evaluation of existing treatment protocols – a relatively less researched topic in econometrics. The present paper attempts to contribute to this topic. In particular, in this paper, we have defined and analyzed the problem of detecting taste-based allocation of a binary treatment through a partial identification approach and using a novel data combination method to learn the necessary counterfactuals. The latter methodology, though nonstandard, is somewhat similar in spirit to using validation data in measurement error analysis (c.f., Chen Hong and Tamer (2004)), in that we are using experimental estimates to "validate" observational studies. Such data combination is most easily feasible when observational and experimental data are collected as part of the same study, as in the empirical example analyzed above. It also extends straightforwardly to situations where independent observational and experimental studies exist on the efficacy of a treatment such as in the healthcare examples cited in the introduction.

Our analysis in this paper is based on an expected utility framework. In the case of non-binary outcomes, there are alternative approaches that are worth investigating. One would be to consider the notion of loss aversion. A crucial feature of the analysis presented above is that inequalities in terms of variables that the planner observes are preserved when aggregated across unobservables– a version of the law of iterated expectations for inequalities. This feature holds

for expected utilities–including the case where the sub-utility function exhibits loss aversion–but may not be shared by all the alternative criteria for treatment assignment, e.g., if the goal were to minimize the variance of outcome in the population. There are also possible extensions of the analysis that would relax the assumption of correct expectations and incorporate some form of learning by the planner. Another extension is to consider heterogeneity in assignment protocols used by different treaters. This latter direction is currently being investigated in the context of university admissions by Bhattacharya et al (2011).

References

[1] Andrews, D.W.K. & Gustavo Soares (2010): "Inference for Parameters Defined by Moment Inequalities Using Generalized Moment Selection," Econometrica, Econometric Society, vol. 78(1), pages 119-157.

[2] Angelucci, M & G. De Giorgi (2009): "Indirect Effects of an Aid Program: How do Cash Injections Affect Ineligibles' Consumption?", American Economic Review, 99(1), pp. 486-508.

[3] Antonovics, Kate L. and Brian G. Knight, (2009): "A New Look at Racial Profiling: Evidence from the Boston Police Department," Review of Economics and Statistics, 2009, 91, 163–177.

[4] Anwar, Shamena and Hanming Fang (2006): "An Alternative Test of Racial Profiling in Motor Vehicle Searches: Theory and Evidence," American Economic Review, 96, 127–151.

[5] Anwar, Shamena and Hanming Fang (2011): "Testing for the Role of Prejudice in Emergency Departments using Bounceback Rates," NBER working paper, number 16888.

[6] Arrow, K. (1973): The theory of discrimination, in "Discrimination in labor markets", Princeton.

[7] Becker, Gary (1957): The economics of discrimination, University of Chicago Press.

[8] Bhattacharya, D. (2009): "Inferring Optimal Peer Assignment from Experimental Data", Journal of the American Statistical Association, Jun 2009, Vol. 104, No. 486: pages 486–500.

[9] Bhattacharya, D. & Dupas, P. (2010): Inferring Efficient Treatment Assignment under Budget Constraints", NBER working paper number 14447.

[10] Bhattacharya, D., S. Kanaya, and M. Stevens (2011): Are University Admissions Academically Fair?, mimeo., Oxford University.

[11] Brock, William A., J Cooley, S. Durlauf and S. Navarro (2011): "On the Observational Implications of Taste-Based Discrimination in Racial Profiling,", forthcoming, Journal of Econometrics.

[12] CASS (1984): Circulation, Journal of American College of Cardiology, vol. 3, pp.114-128, published by the American College of Cardiology Foundation.

[13] Chen, X., H. Hong and E. Tamer (2005): "Measurement Error Models with Auxiliary Data", Review of Economic Studies, 72, pp. 343–366.

[14] Chernozhukov, V., Han Hong & E. Tamer (2007): Estimation and Confidence Regions for Parameter Sets in Econometric Models, Econometrica, Econometric Society, vol. 75(5), pages 1243-1284.

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

[16] Elliott, G., I. Komunjer and A. Timmerman (2005): Estimation and Testing of Forecast Rationality under Flexible Loss. Review of Economic Studies, 72, pp. 1107-1125.

[17] Heckman, J. (1998): Detecting discrimination, Journal of Economic Perspectives-Volume 12, Number 2, Pages 101-116.

[18] Hirano, K. and J. Porter (2009): "Asymptotics for Statistical Treatment Rules", *Econometrica,* vol. 77(5), pages 1683-1701.

[19] Knowles, Persico and Todd (2001): Racial Bias in Motor Vehicle Searches: Theory and Evidence, Journal of Political Economy,109 (11) pp. 203-229.

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

[21] Manski, C. (2005): Social choice with partial knowledge of treatment response, Princeton University Press.

[22] Manski, C. & John V. Pepper (2000): "Monotone Instrumental Variables, with an Application to the Returns to Schooling," Econometrica, Econometric Society, vol. 68(4), pages 997-1012.

[23] Patton, Andrew J. & Timmermann, Allan (2007): "Testing Forecast Optimality Under Unknown Loss," Journal of the American Statistical Association, vol. 102, pages 1172-1184.

[24] Persico, N (2009): "Racial Profiling? Detecting Bias Using Statistical Evidence", Annual Review of Economics, volume 1.

[25] Pope, D. and Sydnor (2008): What's in a Picture? Evidence of Discrimination from Prosper.com, forthcoming in Journal of Human Resources, 2010.

[26] Rosen, A. (2008): Confidence Sets for Partially Identified Parameters that Satisfy a Finite Number of Moment Inequalities, Journal of Econometrics, vol. 146, issue 1, pp. 107-117.

[27] Stern, N. (1977), 'Welfare Weights and the Elasticity of the Marginal Valuation of Income', in Artis, M. and Nobay, R. (eds.) Studies in Modern Economic Analysis, Oxford: Basil Blackwell.

[28] Tamer, E. (2008): "Partial Identification In Econometrics", The Annual Review of Economics, 2010.

7 Appendix:

Proof of theorem 1:

Conditions: (i) $C_0 \geq 0$ and $\Delta C > \bar{A} > 0$, w.p. 1, (ii) $E[C_0] < c$.

Define the constant γ as

$$\begin{aligned} \gamma^* &: = \inf \left\{ a : \int \left[\begin{array}{c} E(C_0|w) \times 1\{h(w) E(\Delta Y|W=w) \leq aE(\Delta C|W=w)\} \\ +E(C_1|w) \times 1\{h(w) E(\Delta Y|W=w) > aE(\Delta C|W=w)\} \end{array} \right] dF_W(w) \leq c \right\} \\ \gamma &: = \max\{\gamma^*, 0\}. \end{aligned}$$

Such a γ exists, by conditions (ii) since taking a to ∞ will satisfy condition (ii). Also, $\gamma \geq 0$ by definition.

Under conditions (i)–(ii), the solution to the problem

$$\max_{p(\cdot)\in[0,1]} \int h(w) \left(p(w) E(Y_1|W=w) + \int (1-p(w)) E(Y_0|W=w) \right) dF_W(w),$$

s.t.

$$\int \{p(w) E(C_1|W=w) + \{1-p(w)\} E(C_0|W=w)\} dF_W(w) \leq c,$$

is of the form

$$p^*(w) = \begin{cases} 1 \text{ if } h(w) E(\Delta Y|W=w) > \gamma E(\Delta C|W=w), \\ q \text{ if } h(w) E(\Delta Y|W=w) = \gamma E(\Delta C|W=w), \\ 0 \text{ if } h(w) E(\Delta Y|W=w) < \gamma E(\Delta C|W=w), \end{cases}$$

where $q \in [0,1]$ satisfies

$$\int \left(\begin{array}{c} \left\{ \begin{array}{c} 1(h(w) E(\Delta Y|W=w) > \gamma E(\Delta C|W=w)) \\ +q1(h(w) E(\Delta Y|W=w) = \gamma E(\Delta C|W=w)) \end{array} \right\} \times E(C_1|w) \\ +1\{h(w) E(\Delta Y|W=w) < \gamma E(\Delta C|W=w)\} \times E(C_0|w) \end{array} \right) dF_W(w) = c,$$

if

$$\Pr\{h(w) \times E(\Delta Y|W=w) = \gamma E(\Delta C|W=w)\} > 0.$$

and is equal to zero otherwise. If the budget constraint binds and $\gamma^* > 0$, then $\gamma = \gamma^* > 0$; if the budget does not bind, then $\gamma^* \leq 0$ and $\gamma = 0$.

In particular, if $h(W) E(\Delta Y|W)$ has a positive Lebesgue density on an open interval around γ, then

$$\Pr(h(W) E(\Delta Y|W) > \gamma E(\Delta C|W)) = \Pr(h(W) E(\Delta Y|W) \geq \gamma E(\Delta C|W)) = c$$

and $q = 0$. Then $p^*(w) = 1(h(W) E(\Delta Y|W) \geq \gamma E(\Delta C|W))$.

Proof. Consider any feasible rule $p(\cdot)$ which differs from $p^*(\cdot)$. The welfare difference

between using p and p^* is given by

$$\begin{aligned}
& W(p^*) - W(p) \\
= & \int h(w) E(\Delta Y|W=w) \{p^*(w) - p(w)\} dF_W(w) \\
= & \int h(w) E(\Delta Y|W=w) \{p^*(w) - p(w)\} 1\{p^*(w)=1\} dF_W(w) \\
& + \int h(w) E(\Delta Y|W=w) \{p^*(w) - p(w)\} 1\{p^*(w)=q\} dF_W(w) \\
& + \int h(w) E(\Delta Y|W=w) \{p^*(w) - p(w)\} 1\{p^*(w)=0\} dF_W(w) \\
\geq & \int \{1-p(w)\} \gamma E(\Delta C|W=w) 1\{p^*(w)=1\} dF_W(w) \\
& + \int \gamma E(\Delta C|W=w) \{q - p(w)\} 1\{p^*(w)=q\} dF_W(w) \\
& - \int p(w) \gamma E(\Delta C|W=w) p(w) 1\{p^*(w)=0\} dF_W(w) \qquad (20) \\
= & \int \{p^*(w) - p(w)\} E(\Delta C|W=w) dF_W(w).
\end{aligned}$$

There are two cases – (a) where $p^*(w)$ makes the budget constraint bind with $\gamma > 0$, and (b) $\gamma = 0$ and the budget constraint does not bind, i.e.,

$$p^*(w) = \begin{cases} 1 \text{ if } h(w) E(\Delta Y|W=w) \geq 0, \\ 0, \text{ otherwise.} \end{cases},$$

and

$$\begin{aligned}
\int \{1\{E(\Delta Y|W=w) \geq 0\} E(C_1|W=w)\} dF_W(w) &< c, \\
\gamma &= 0.
\end{aligned}$$

In case (a), since p is feasible we have from the budget constraint that

$$\begin{aligned}
\int p^*(w) E(\Delta C|W=w) dF_W(w) &= c - E(C_0) \\
&\geq \int p(w) E(\Delta C|W=w) dF_W(w),
\end{aligned}$$

implying that

$$\int \{p^*(w) - p(w)\} E(\Delta C|W=w) dF_W(w) \geq 0. \qquad (21)$$

Substituting in (20), implies that $W(p^*) \geq W(p)$. Since $E(\Delta C|W=w) > \bar{A} > 0$ w.p. 1, (20) can be zero if and only if $p(w) = 0$ whenever $p^*(w) = 0$ and $p(w) = 1$ whenever $p^*(w) = 1$.

Then $p = p^*$ except when $p^*(w) = q$. Then the budget constraint implies that when $p^*(w) = q$, we must have $p(w) < q$. This, in turn, implies that the inequality in (21) and hence in (20) must be strict. So either inequality (20) or inequality (21) is strict which implies the desired result that $W(p^*) > W(p)$.

Next consider the case (b) where the budget constraint does not bind. Then $\gamma = 0$ and $p^*(w) = 1(h(W) E(\Delta Y|W) \geq 0)$. Then

$$\begin{aligned}
& W(p^*) - W(p) \\
= & \int h(w) E(\Delta Y|W = w) \{p^*(w) - p(w)\} dF_W(w) \\
= & \int h(w) E(\Delta Y|W = w) \{1 - p(w)\} 1\{p^*(w) = 1\} dF_W(w) \\
& + \int \{-h(w) E(\Delta Y|W = w)\} p(w) 1\{p^*(w) = 0\} dF_W(w) \\
= & \int h(w) E(\Delta Y|W = w) \{1 - p(w)\} 1\{p^*(w) = 1, , h(w) E(\Delta Y|W = w) > 0\} dF_W(w) \\
& + \int h(w) E(\Delta Y|W = w) \{1 - p(w)\} 1\{p^*(w) = 1, h(w) E(\Delta Y|W = w) = 0\} dF_W(w) \\
& + \int \{-h(w) E(\Delta Y|W = w)\} p(w) 1\{p^*(w) = 0\} dF_W(w) \\
= & \int h(w) E(\Delta Y|W = w) \{1 - p(w)\} 1\{p^*(w) = 1, h(w) E(\Delta Y|W = w) > 0\} dF_W(w) \\
& + \int \{-h(w) E(\Delta Y|W = w)\} p(w) 1\{p^*(w) = 0\} dF_W(w).
\end{aligned}$$

The first term $h(w) E(\Delta Y|W = w)$ within the first integral is strictly positive when $p^*(w) = 1$ and $h(w) E(\Delta Y|W = w) > 0$ and the first term in the second integral is non-negative since $h(w) E(\Delta Y|W = w) \leq 0$ when $p^*(w) = 0$. Therefore, first integral is strictly positive unless $p(w)$ equals 1 whenever $p^*(w) = 1$. In the latter case, $p(w)$ must be positive on some subset of positive probability where $p^*(w) = 0$ and $h(w) E(\Delta Y|W = w) < 0$. But this will make the second integral strictly positive. So either the first integral or the second integral is strictly positive and both integrals are non-negative, which proves the assertion. Due to the strict inequality, it also follows that the solution p^* is unique. ■