Identification problems and decisions under ambiguity: Empirical analysis of treatment response and normative analysis of treatment choice

Charles F. Manski*

Department of Economics and Institute for Policy Research, Northwestern University,
2003 Sheridan Road, Evanston, IL, 60208, USA

Abstract

Identification problems that arise in the empirical analysis of treatment response generate ambiguity about the identity of optimal treatment rules. This paper develops general themes about decisions under ambiguity, next specializes to problems of treatment choice under ambiguity, and then shows how identification problems induce ambiguity in treatment choice. The main ideas are given specific form through consideration of the treatment choice problem of a planner who observes the treatments and outcomes realized in a classical randomized experiment, but who does not observe the covariates of the experimental subjects. © 2000 Elsevier Science S.A. All rights reserved.

JEL classification: C10; D81

Keywords: Identification; Ambiguity; Treatment Response
1. Introduction

Economists have long associated decision making with optimization. We routinely suppose that decision makers with known choice sets act to maximize known objective functions.

A decision maker with a known choice set but an unknown objective function is said to face a problem of decision under ambiguity. The term ambiguity appears to originate with Ellsberg (1961), who sought to understand choice behavior in situations where the objective function depends on an unknown objective probability distribution. His famous experiment required subjects to draw a ball from either of two urns, one with a known distribution of colors and the other with an unknown distribution of colors. The term has since been adopted by Einhorn and Hogarth (1986), Camerer and Weber (1992), and other decision theorists. Much earlier, Keynes (1921) and Knight (1921) used the term uncertainty to describe these problems, but uncertainty has since come to be used to describe optimization problems in which the objective function depends on a known probability distribution. Some authors have used ignorance as a synonym for ambiguity (e.g., Arrow and Hurwicz, 1972; Maskin, 1979).1

The formal study of principles for decision making under ambiguity dates back at least to Wald (1950) and has recently become a prominent concern of decision theory. Nevertheless, applied economists engaged in positive and normative analysis of decision processes invariably characterize agents as solving optimization problems. Applied economists may perhaps believe ambiguity to be unusual in practice, or inconsequential, or analytically intractable. For whatever reason, the study of decisions under ambiguity has remained a peripheral concern of the profession.

This paper connects decisions under ambiguity with identification problems in econometrics. Considered abstractly, it is natural to make this connection. Ambiguity occurs when lack of knowledge of an objective probability distribution prevents a decision maker from solving an optimization problem. Empirical research seeks to draw conclusions about objective probability distributions by combining assumptions with observations. An identification problem occurs when a specified set of assumptions combined with unlimited observations drawn by a specified sampling process does not reveal a distribution of interest. Thus, identification problems generate ambiguity in decision making.2

---

1 There is also a literature on decision making with unknown subjective probability distributions. Here authors may refer to robust Bayesian analysis (e.g., Berger, 1985) or to decision making with lower/upper probabilities (e.g., Dempster, 1967,1968; Shafer, 1976), subjective probability domains (e.g., Manski, 1981), or imprecise probabilities (e.g., Walley, 1991).

2 Statistical problems of induction from finite samples to populations generate further ambiguity. Section 6 gives a brief discussion.
The particular connection that I develop relates econometric research on identification of treatment response to normative analysis of treatment choice. The canonical treatment choice problem to be considered here supposes that a planner must choose a treatment rule assigning a treatment to each member of a heterogeneous population of interest. For example, the planner might be a physician choosing medical treatments for each member of a population of patients or a judge deciding sentences for each member of a population of convicted offenders. The planner observes certain covariates for each person; perhaps demographic attributes, medical or criminal records, and so on. These covariates determine the set of non-randomized treatment rules that are feasible to implement: the set of feasible such rules is the set of all functions mapping the observed covariates into treatments. Each member of the population has a response function mapping treatments into a real-valued outcome of interest; perhaps a measure of health status in the case of the physician or a measure of recidivism in the case of the judge. The planner wants to choose a treatment rule that maximizes the population mean outcome; that is, the planner wants to maximize a utilitarian social function.

In this setting, an optimal treatment rule assigns to each member of the population a treatment that maximizes mean outcome conditional on the person’s observed covariates (see Section 3). A planner who knows these conditional means can implement an optimal rule, and thus faces a problem of treatment choice under uncertainty. A planner who lacks this knowledge faces a problem of treatment choice under ambiguity.

Econometricians and other methodologists have sought to determine the conclusions about treatment response in the population of interest that can be drawn from data on the (covariates, treatments, outcomes) realized by members of some previously treated population. A body of research shows that conditional mean responses under alternative treatments are identified if such data are combined with sufficiently strong assumptions (e.g., Maddala, 1983; Rosenbaum and Rubin, 1983; Heckman and Robb, 1985; Björklund and Moffitt, 1987). Abstracting from the statistical issues that arise in inference from finite samples, these identification results give conditions under which planners can implement optimal treatment rules.

Problems of treatment choice under ambiguity arise when the available data on the treated population and the assumptions that the planner finds credible do not suffice to identify mean treatment response in the population of interest. The program of research on non-parametric analysis of treatment response initiated in Manski (1989, 1990) characterizes such circumstances. The findings are sharp bounds on conditional mean responses under alternative treatments, the form of the bounds depending on the available data and the maintained assumptions. These bounds determine the nature of the ambiguity that the planner faces.

Here is the organization of the paper. Section 2 develops general themes about decisions under ambiguity. I evaluate the common practice,
recommended by Bayesian decision theorists and others, of transforming decisions under ambiguity into optimization problems by judicious modification of the planner’s objective function. I call attention to an important qualitative difference between optimization problems and decisions under ambiguity: whereas expansion of the choice set can do no harm in an optimization problem, it may do harm in a decision under ambiguity. The reason is that a decision maker facing ambiguity may not be able to order the newly feasible actions relative to those in the original choice set. Hence, expansion of the choice set may lead the decision maker to choose a newly feasible action that is inferior to the action he would choose otherwise.

Section 3 formalizes the canonical treatment choice problem. I derive the optimal treatment rule, namely maximization of mean response conditional on the covariates that the planner observes. I show how the general point about expansion of choice sets made in Section 2 manifests itself in the treatment choice setting. The more covariates that the planner observes, the larger is the set of feasible rules. Hence, observing more covariates may lead the planner to choose a newly feasible treatment rule that is inferior to the rule he would choose otherwise. I suggest that this possibility may help to explain the common empirical finding that statistical treatment rules using limited covariate information tend to outperform the subjective rules that clinicians apply using the extensive covariate information available to them. The section concludes with a discussion of decentralized treatment selection. Here I compare the planner’s choice problem with those of individuals who must select treatments for themselves.

Section 4 shows how identification problems induce ambiguity in treatment choice. A fundamental difficulty is that response functions are not directly observable. It is at most possible to observe the outcome that a person experiences under the treatment that this person actually receives. I review and elaborate on the simple analysis of this problem in Manski (1990). Here the treated population and the population of interest are taken to be distributionally identical, but no assumptions are made about the process generating realized treatments and outcomes. In this setting, random sample data on realized (covariates, treatments, outcomes) imply bounds on mean responses. The bounds are informative if the outcome of interest is itself a bounded variable. However the bounds for alternative treatments necessarily overlap one another. Hence, in the absence of maintained assumptions about the process generating realized treatments and outcomes, the planner cannot even partially order the feasible treatment rules. This result, when combined with the common difficulty of justifying strong maintained assumptions, makes a compelling argument that ambiguity is a fundamental problem of treatment choice in practice.

A planner who can combine observations of the treated population with credible assumptions about the process generating realized treatments and outcomes may be able to deduce non-overlapping bounds for mean outcomes
under alternative treatments and, hence, may be able to order some treatment rules. In Section 5, I develop the main ideas through examination of a substantively and methodologically interesting identification problem that has not previously received attention. I examine the problem faced by a planner who observes the treatments and outcomes realized in a classical randomized experiment, but who does not observe the covariates of the experimental subjects. For example, the planner may be a physician who reads a medical journal report of the outcomes of a clinical trial. The physician may have extensive covariate information for his own patients but the journal report of the clinical trial may only report outcomes within broad risk-factor groups. We know that the optimal treatment rule is to choose for each person in the population of interest a treatment that maximizes mean outcome conditional on the covariates that the planner observes in this population. However the available experimental evidence, lacking covariate data, only reveals mean outcomes in the population as a whole, not mean outcomes conditional on covariates. Hence the planner faces a problem of treatment choice under ambiguity.

I focus on the simplest non-trivial case; that in which treatments, outcomes, and covariates are all binary. There are four feasible treatment rules in this setting; two rules assign the same treatment to all persons and two assign different treatments to persons with different covariate values. I apply a finding of Manski (1997a) to determine which of the four rules are dominated and, consequently, should not be chosen by the planner. It turns out that the dominated rules depend on a somewhat subtle interplay of the distribution of covariates in the population and the distributions of outcomes revealed by the randomized experiment. There are as many as three or as few as zero dominated treatment rules, depending on the configuration of these distributions. I use data from the Perry Preschool experiment to illustrate the findings.

The identification problems that are the concern of this paper are important sources of ambiguity in treatment choice, but they are not the only sources. Planners commonly observe a sample of the treated population, not the entire population. Hence they must contend with statistical problems of induction from finite samples to populations, which generate further ambiguity. Section 6 remarks on the classical and Bayesian prescriptions for dealing with statistical ambiguity.

2. Decisions under ambiguity

2.1. Basic ideas

We begin with a universe $A$ of actions, a choice set $C \subseteq A$, and a decision-maker who must choose an action from $C$. The decision-maker wants to maximize on $C$ an objective function $f(\cdot): A \to R$ mapping actions into
Throughout this paper, the term ‘welfare’ refers to the value of the objective function that the decision maker realizes. Thus, the welfare associated with action $i$ is $f(i)$. Some authors refer to $f(i)$ as ‘ex post’ welfare and to $h(i)$ as ‘ex ante’ welfare.

real-valued outcomes. The decision maker faces an optimization problem if he knows the choice set $C$ and the objective function $f(\cdot)$. He faces a problem of decision under ambiguity if he knows the choice set but not the objective function. Instead, he knows only that $f(\cdot) \in F$, where $F$ is some set of functions mapping $A$ into $R$.

Knowing that $f(\cdot) \in F$, how should the decision maker choose among the feasible actions? Clearly he should not choose a dominated action. Action $d \in C$ is said to be dominated (also inadmissible) if there exists another feasible action, say $c$, such that $g(d) \leq g(c)$ for all $g(\cdot) \in F$ and $g(d) < g(c)$ for some $g(\cdot) \in F$.

Let $D$ denote the undominated subset of $C$. How should the decision maker choose among the elements of $D$? Let $c$ and $d$ be two undominated actions. Then either $[g(c) = g(d), \forall g(\cdot) \in F]$ or there exist $g'(\cdot) \in F$ and $g''(\cdot) \in F$ such that $[g'(c) > g'(d), g''(c) < g''(d)]$. In the former case, $c$ and $d$ are equally good choices and the decision maker is indifferent between them. In the latter case, the decision maker cannot order the two actions. Action $c$ may yield a better or worse outcome than action $d$; the decision maker cannot say which. Thus the normative question ‘How should the decision maker choose?’ has no unambiguously correct answer.

2.2. Transforming decisions under ambiguity into optimization problems

Although there is no optimal choice among undominated actions, decision theorists have not wanted to abandon the idea of optimization. So they have proposed various ways of transforming the unknown objective function $f(\cdot)$ into a known function, say $h(\cdot) : A \rightarrow R$, that can be maximized on $D$. Three leading proposals – the maximin rule, Bayes decision rules, and imputation rules – are discussed here. Although these proposals differ in their details, they share a key common feature. In each case the solvable optimization problem, namely $\max_{i \in D} h(\cdot)$, differs from the problem that the decision maker wants to solve, namely $\max_{i \in D} f(\cdot)$. The welfare level that is attained under the solvable optimization problem is $f[\arg \max_{i \in D} h(\cdot)]$, not $\max_{i \in D} f(\cdot)$.\(^3\)

2.2.1. The maximin rule

Wald (1950) proposed that the decision maker should choose an action that maximizes the minimum welfare attainable under the functions in $F$. Formally,

\[ \text{Maximin rule: For each } d \in D, \text{ let } h(d) = \inf_{g(\cdot) \in F} g(d). \text{ Maximize } h(\cdot) \text{ on } D. \]

\(^3\)Throughout this paper, the term ‘welfare’ refers to the value of the objective function that the decision maker realizes. Thus, the welfare associated with action $i$ is $f(i)$. Some authors refer to $f(i)$ as ‘ex post’ welfare and to $h(i)$ as ‘ex ante’ welfare.
The maximin rule has a clear normative foundation in competitive games. In a competitive game, the decision maker chooses an action from $C$. Then a function from $F$ is chosen by an opponent whose objective is to minimize the realized outcome. A decision maker who knows that he is a participant in a competitive game does not face ambiguity. He faces the problem of maximizing the known function $h(\cdot)$ specified in the maximin rule.

There is no compelling reason why the decision maker should or should not use the maximin rule when $f(\cdot)$ is a fixed but unknown objective function. In this setting, the appeal of the maximin rule is a personal rather than normative matter. Some decision makers may deem it essential to protect against worst-case scenarios, while others may not. Wald himself did not contend that the maximin rule is optimal, only that it is 'reasonable.' Considering the case in which the objective is to minimize rather than maximize $f(\cdot)$, he wrote (Wald, 1950, p. 18): 'a minimax solution seems, in general, to be a reasonable solution of the decision problem.'

2.2.2. Bayes decision rules

Bayesian decision theorists assert that a decision maker who knows only that $f(\cdot) \in F$ should choose an action that maximizes some average of the elements of $F$. Formally,

Bayes decision rule: Place a $\sigma$-algebra $\Sigma$ and some probability measure $\pi$ on the function space $F$. Let $h(\cdot) \equiv \int g(\cdot) d\pi$. Maximize $h(\cdot)$ on $D$.

Bayesian decision theorists recommend that $\pi$ should express the decision maker's personal beliefs about where $f(\cdot)$ lies within $F$.

Bayesians offer various rationality arguments for use of Bayes decision rules. The most basic of these is that Bayes decision rules generally yield undominated actions provided that the expectations $\int g(\cdot) d\pi$ are finite (Berger, 1985, p. 253). This and other rationality arguments do not, however, fully answer the decision maker's bottom-line question: how well does the rule perform?

Consider, for example, the famous axiomatic approach of Savage (1954). Savage shows that a decision maker whose choices are consistent with a specified set of axioms can be interpreted as using a Bayes decision rule. Many decision theorists consider the Savage axioms, or other sets of axioms, to be a priori appealing. Acting in a manner that is consistent with these axioms does not, however imply that chosen actions yield good outcomes. Berger (1985, p. 121) calls attention to this, stating: 'A Bayesian analysis may be 'rational' in the weak axiomatic sense, yet be terrible in a practical sense if an inappropriate prior distribution is used.'

Even use of an 'appropriate' prior distribution $\pi$ does not imply that the decision maker should choose an action that maximizes the $\pi$-average of the functions in $F$. Suppose that $\pi$ has actually been used to draw $f(\cdot)$ from $F$; that
is, let $\pi$ describe an objective random process and not just the decision maker’s subjective beliefs. Even here, where use of $\pi$ as the prior distribution clearly is appropriate, Bayesian decision theory does not show that maximizing the $\pi$-average of $F$ is superior to other decision rules in terms of the outcome it yields. A decision maker wanting to obtain good outcomes might just as reasonably choose an action that maximizes a $\pi$-quantile of $F$ or some other parameter of $F$ that respects stochastic dominance (Manski, 1988).

2.2.3. Imputation rules

A prevalent practice among applied researchers is to act as if one does know $f(\cdot)$. One admits to not knowing $f(\cdot)$ but argues that pragmatism requires making some ‘reasonable,’ ‘plausible,’ or ‘convenient’ assumption. Thus one somehow imputes the objective function and then chooses an action that is optimal under the imputed function. Formally,

**Imputation rule:** Select some $h(\cdot) \in F$. Maximize $h(\cdot)$ on $D$.

Imputation rules are essentially Bayes rules placing probability one on a single element of $F$.

2.3. Ambiguity untransformed

Decision theorists have long sought to transform decisions under ambiguity into optimization problems. Yet the search for an optimal way to choose among undominated actions must ultimately fail. Let us face up to this. What then?

Simply put, normative analysis changes its focus from optimal actions to undominated actions. In optimization problems, the optimal actions and the undominated actions coincide, the decision maker being indifferent among all undominated actions. In decisions under ambiguity, there may be undominated actions that the decision maker cannot order.

This change of focus, albeit simple, has a qualitatively important implication. Let $c$ denote the action that the decision maker chooses from his choice set $C$. Consider the effect on welfare of adding a new feasible action, say $b \in A$, to the choice set. In an optimization problem, expansion of the choice set from $C$ to $C \cup b$ cannot decrease welfare because the decision maker will not choose $b$ if $f(b) < f(c)$. Under ambiguity, expansion of the choice set may decrease welfare. Suppose that $b$ neither dominates nor is dominated by the elements of $D$, so the new set of undominated actions is $D \cup b$. Then the decision maker may choose $b$ and it may turn out that $f(b) < f(c)$.

The possibility that expansion of the choice set may decrease welfare is familiar in the multiple-decision-maker settings considered in game theory, where expansion of choice sets can generate new inferior equilibria. To the best
of my knowledge, this possibility has not previously been recognized in the single-decision-maker settings considered in decision theory.

3. Treatment choice under uncertainty and ambiguity

3.1. The planner’s choice set and objective function

I now formalize the problem of treatment choice. I suppose that there is a finite set $T$ of treatments and a planner who must choose a treatment rule assigning a treatment in $T$ to each member of a population $J$. Each person $j \in J$ has a response function $y_{j}(t) : T \rightarrow Y$ mapping treatments into real-valued outcomes $y_{j}(t) \in Y$. A treatment rule is a function $\tau(\cdot) : J \rightarrow T$ specifying which treatment each person receives. Thus, person $j$’s outcome under rule $\tau(\cdot)$ is $y_{j}[\tau(j)]$.

The planner is concerned with the distribution of outcomes across the population, not with the experiences of particular individuals. With this in mind, I take the population to be a probability space, say $(J, \Omega, P)$, where $\Omega$ is the $\sigma$-algebra on which probabilities are defined and $P$ is the probability measure. Now the population mean outcome under treatment rule $\tau(\cdot)$ is

$$E\{y_{j}[\tau(j)]\} \equiv \int y_{j}[\tau(j)]dP(j).$$ (1)

I assume that the planner wants to choose a treatment rule that maximizes $E\{y_{j}[\tau(j)]\}$. This criterion function has both normative and analytical appeal. Maximization of a population mean outcome, or perhaps some weighted average outcome, is the standard utilitarian criterion of the public economics literature on social planning. The linearity of the expectation operator yields substantial analytical simplifications, particularly through use of the law of iterated expectations.

I suppose that the planner observes certain covariates $x_{j} \in X$ for each member of the population. The planner cannot distinguish among persons with the same observed covariates. Hence he cannot implement treatment rules that systematically differentiate among these persons. With this in mind, I take the feasible

---

4 This notation maintains the assumption of ‘individualistic treatment’ made commonly, albeit often only implicitly, in analyses of treatment response. Individualistic treatment means that each person’s outcome may depend on the treatment he receives, but not on the treatments received by other persons.
rules to be the set of functions mapping the observed covariates into treatments.\footnote{Although the planner cannot systematically differentiate among persons with the same observed covariates, he can randomly assign different treatments to such persons. Thus the set of feasible treatment rules in principle contains not only functions mapping covariates into treatments but also probability mixtures of these functions. Explicit consideration of randomized treatment rules would not substantively change the analysis of this paper, but would complicate the necessary notation. A simple implicit way to permit randomized rules is to include in \( z(x) \) a component whose value is randomly drawn by the planner from some distribution. The planner can then make the chosen treatment vary with this covariate component.}

To formalize this, let \( Z \) denote the space of all functions mapping \( X \) into \( T \). Then the feasible rules have the form

\[
\tau(j) = z(x), \quad j \in J,
\]

where \( z(\cdot) \in Z \). Let \( P[y(\cdot), x] \) be the probability measure on \( Y^T \times X \) induced by \( P(j) \) and let \( E[y(z(x))] \equiv \int y(z(x))dP[y(\cdot), x] \) denote the expected value of \( y[z(x)] \) with respect to this induced measure. The planner wants to solve the problem

\[
\max_{z(\cdot) \in Z} E[y[z(x)]].
\]  

3.2. Optimal treatment choice under uncertainty

It is easy to show that the optimal choice among the set \( Z \) of feasible non-randomized treatment rules assigns to each member of the population a treatment that maximizes mean outcome conditional on the person’s observed covariates. Let \( 1[\cdot] \) be the indicator function taking the value one if the logical condition in the brackets holds and the value zero otherwise. For each \( z(\cdot) \in Z \), use the law of iterated expectations to write

\[
E[y[z(x)]] = E \left\{ E[y[z(x)]|x] \right\} = E \left\{ \sum_{t \in T} E[y(t)|x] \cdot 1[z(x) = t] \right\}.
\]  

For each \( x \in X \), the bracketed expression on the right side is maximized by choosing \( z(x) \) to maximize \( E[y(t)|x] \) on \( t \in T \). Hence a treatment rule \( z^*(\cdot) \) is optimal if

\[
z^*(x) = \arg\max_{t \in T} E[y(t)|x], \quad x \in X.
\]  

The optimized population mean outcome is

\[
\bar{Y}^* \equiv E \left\{ \max_{t \in T} E[y(t)|x] \right\}.
\]
The planner is said to face a problem of treatment choice under uncertainty if he knows the conditional mean responses $E[y(t)|x], x \in X$ and thus can implement an optimal treatment rule.\(^6\)

It is easy to show, and important to keep in mind, that a planner facing a problem of treatment choice under uncertainty should use all the covariate information available to make treatment choices. Suppose that the planner were to use only $w(x)$, where $w(\cdot): X \rightarrow W$ is some many-to-one function of $x$. Then the feasible treatment rules would be those that give the same treatment to each person with covariates $w(x)$. These rules yield mean outcomes $E[y(t)|w(x)], t \in T$; hence the optimized mean outcome is $E\{\max_{t \in T} E[y(t)|w(x)]\}$. Using $x$, the optimized mean outcome is given in (6). The optimized mean outcome using covariates $x$ to choose treatments is necessarily at least as large as that achievable using the more limited covariates $w(x)$. Thus,

$$E\left\{\max_{t \in T} E[y(t)|x]\right\} \geq E\left\{\max_{t \in T} E[y(t)|w(x)]\right\}. \quad (7)$$

It is important to understand that this finding holds whatever the covariates $x$ may be. Empirical economists often distinguish between ‘exogenous’ and ‘endogenous’ covariates, asserting that only the former should be used as conditioning variables. This distinction is meaningful when covariates are used to identify treatment effects through the assumption of exogenous treatment selection (see Section 4.3). However it is meaningless here, where covariates are being used to determine feasible treatment rules. In the present context, the only relevant fact is that the set of feasible rules grows as more covariates are observed. Hence the optimal mean outcome achievable by the planner cannot fall, and may rise, as more covariates are observed.

### 3.3. Undominated treatment choice under ambiguity

The planner faces a problem of treatment choice under ambiguity if he does not know the conditional mean response functions $E[y(t)|x], x \in X$. Suppose the planner knows only that the population (covariate, response function) distribution $P[x, y(t)]$ lies within a specified set $\Phi$ of possible (covariate, response function) distributions. The planner may then partition the feasible treatment rules into dominated and undominated subclasses. A feasible treatment rule $z(\cdot)$

---

\(^6\)The special case of treatment choice under certainty occurs if all persons with covariates $x$ have the same response function. In this case, knowing $E[y(t)|x], x \in X$ means knowing each person’s response function.
is dominated if there exists another feasible rule, say \( z'(\cdot) \), such that

\[
\int y[z(x)]d\phi \leq \int y[z'(x)]d\phi, \text{ all } \phi \in \Phi, \tag{8a}
\]

\[
\int y[z(x)]d\phi < \int y[z'(x)]d\phi, \text{ some } \phi \in \Phi. \tag{8b}
\]

A treatment rule \( z(\cdot) \) is undominated if no such \( z'(\cdot) \) exists. A planner facing a problem of treatment under ambiguity can eliminate dominated rules as sub-optimal but cannot choose optimally among rules that are undominated and unordered.

A generic feature of decisions under ambiguity is that expansion of the choice set may decrease welfare (see Section 2.3). In the treatment-choice setting, the choice set is the space of functions mapping covariates into treatments, so observation of additional covariates implies expansion of the choice set. Observation of additional covariates enables the planner to choose a treatment rule that more finely differentiates among the members of the population. The problem is that the planner, not knowing mean outcomes conditional on these covariates, may unwittingly use them to choose a worse treatment rule.

To see this, it suffices to consider the extreme case where the planner has no knowledge of \( P[x, y(\cdot)] \), so \( \Phi \) is the set of all (covariate, response function) distributions. In this case all feasible treatment rules are undominated and the planner may inadvertently choose the worst possible one. If only \( w(x) \) is observed, the feasible treatment rules give the same treatment to every person with covariates \( w(x) \). Then the worst possible rule yields \( E\{\min_{t \in T} E[y(t)|w(x)]\} \) as the population mean outcome. If \( x \) is observed, the feasible treatment rules give the same treatment to every person with covariates \( x \). Now the worst possible rule yields \( E\{\min_{t \in T} E[y(t)|x]\} \) as the population mean outcome. Thus,

\[
E\left\{\min_{t \in T} E[y(t)|x]\right\} \leq E\left\{\min_{t \in T} E[y(t)|w(x)]\right\}. \tag{9}
\]

Hence using \( x \) to choose treatments may decrease the population mean outcome achieved by the planner.

Juxtaposing the present analysis with that of Section 3.2 shows that the value of covariate information differs qualitatively in problems of treatment choice under uncertainty and ambiguity. Using the available covariates to choose treatments is always advantageous in the former case, however it may degrade the quality of decision making in the latter case.

The present analysis may shed some new light on the longstanding controversy about the relative merits of statistical and clinical approaches to outcome prediction and treatment choice (Meehl, 1954). Psychologists have reported that statistical treatment rules using a small set of covariates commonly yield better mean outcomes in practice than do clinicians using a large set of covariates to
choose treatments subjectively. It may be that ambiguity is a contributing factor. Clinicians often observe extensive covariate information for their patients, but they may lack knowledge about mean response conditional on these covariates. It may be that they use their extensive covariate information to unwittingly choose inferior treatment rules.

3.4. Decentralized treatment selection

To conclude this section, I ask whether it is desirable for the planner to decentralize treatment selection, leaving it to each member of the population to select his own treatment. Decentralization is clearly appealing if it yields a treatment rule dominating those that the planner can implement. Decentralization has some appeal even if it just yields an undominated treatment rule. After all, the planner cannot claim to know a better rule.

There are several reasons why the treatments self-selected by the members of the population may differ from those chosen by the planner. The population and the planner may have divergent objectives, observe different covariates, have different knowledge of treatment response, or use different criteria to select treatments given the information they possess. I want to focus on informational considerations, leaving aside the possibility of divergent objectives. To this end, let us assume that each person $j$ wants to maximize his own outcome $y_j(t)$ over $t \in T$. Then the population and planner have congruent objectives.

If each person were to know his own response function, decentralized treatment selection would maximize the mean outcome $E\{y_j[\tau(j)]\}$ over all treatment rules $\tau(\cdot): J \to T$. Thus, decentralization would yield the best possible result from the planner’s perspective. The question of interest is what happens if individuals do not know their own response functions.

Decentralization outperforms central planning if one supposes, as economists sometimes do, that individuals observe more covariates than planners observe, that individuals know the distribution of response conditional on their observed covariates (i.e., they have rational expectations), that they maximize expected utility, and that they are risk neutral. Formally, assume that person $j$ observes covariates $(x_j, w_j)$, knows the conditional response distribution $P[y(\cdot)|x_j, w_j]$, and acts to maximize $E[y(t)|x_j, w_j]$ on $t \in T$. Then decentralization achieves the mean outcome $E\{\max_{t \in T}E[y(t)|x, w]\}$. This value is at least as large as the maximum mean outcome attainable by a planner observing covariates $x$, namely $E\{\max_{t \in T}E[y(t)|x]\}$.

Unfortunately, these informational and behavioral assumptions are hard to substantiate. There is little reason to think that individuals generally have rational expectations, maximize expected utility, and are risk neutral. Moreover, the conventional economic assumption that individuals observe more covariates than do planners seems highly suspect in common treatment settings. Consider, for example, the situation of medical patients. Do patients know more
about their own health status than do examining physicians? In this and other common treatment settings, it seems more reasonable to think that individuals and planners observe overlapping but non-nested covariates.

Taken together, the many uncertainties about how individuals would choose their own treatments suggest enormous difficulty in reaching conclusions about the merits of decentralization relative to centralized treatment choice. The remainder of this paper restricts attention to the centralized treatment choice problem set out in Section 3.1.

4. Using observations of a treated population to learn about mean treatment response

Planners commonly have some information about the (covariate, response function) distribution $P[x, y(\cdot)]$. Ambiguity is a practical concern only if the available empirical evidence and credible maintained assumptions do not suffice to determine an optimal treatment rule.

The econometric literature on the analysis of treatment response begins from the premise that one can observe the (covariates, treatments, outcomes) realized by the members of a previously treated population that is identical in relevant respects to the population of interest (see Section 4.1). Unfortunately such empirical evidence does not per se imply that any feasible treatment rule is dominated (see Section 4.2). This easily proved result, when combined with the common difficulty of justifying strong maintained assumptions (see Section 4.3), makes a compelling argument that ambiguity is a fundamental problem of treatment choice in practice.

4.1. The observability of response functions

Empirical analysis of treatment response faces a fundamental difficulty. Consider any person $j \in J$. By definition, treatments are mutually exclusive. Hence it is logically impossible to observe the full vector $[y_j(t), t \in T]$ of outcomes that person $j$ would experience under all treatments. It is at most possible to observe the outcome that $j$ realizes under the treatment that this person actually receives.\(^7\)

\(^7\)The mutual exclusivity of treatments has been a central theme of empirical research on the analysis of treatment effects. Mutual exclusivity of treatments is the reason why the term *experiment* is generally taken to mean a *randomized* experiment in which each person receives one randomly chosen treatment (Fisher, 1935), not a *controlled* experiment in which multiple treatments are applied to one person. A different perspective is found in the economic theory literature on revealed preference analysis. Here, it is sometimes assumed that treatments are not mutually exclusive or, equivalently, that persons receiving different treatments have the same response function. Varian (1982), for example, supposes that an analyst observes multiple realized (treatment, outcome) pairs for a given individual $j$. He investigates how these observations may be used to learn about $j$'s response function $y_j(\cdot)$. 
Even the realized outcome is observable only retrospectively, after a person’s treatment has been chosen. Nothing about response function \( y_j(\cdot) \) is observable prospectively, before the treatment decision. Facing this further difficulty, empirical researchers commonly (albeit often only implicitly) assume the existence of two populations having the same distribution of covariates and response functions, or at least the same conditional mean response functions. One is the population of interest, which I have denoted \( J \). The other is a treated population, say \( K \), in which treatments have previously been chosen and outcomes realized.

Let \( s(\cdot): K \to T \) denote the status quo treatment rule applied in the treated population. Then the realized (covariate, treatment, outcome) triples \( \{x_k, s(k), y_k[s(k)]; k \in K\} \) are observable. Under the assumption that populations \( J \) and \( K \) are distributionally identical, observation of the treated population reveals the distribution \( P[x, s, y(s)] \) of (covariate, treatment, outcome) triples that would be realized in the population of interest if treatment rule \( s(\cdot) \) were to be applied there. Knowledge of this distribution now becomes the basis for empirical analysis.

4.2. All feasible treatment rules are undominated

What is the set of undominated treatment rules given empirical knowledge of \( P[x, s, y(s)] \) but no maintained assumptions about the process generating realized treatments and outcomes? A straightforward extension of the analysis of Manski (1990) shows that this question has a simple but unpleasant answer: All feasible treatment rules are undominated.

Let \( K_0 \) and \( K_1 \) denote the lower and upper endpoints of the logical range of the response functions. If outcomes are binary, for example, then \( K_0 = 0 \) and \( K_1 = 1 \). If outcomes can take any non-negative value, then \( K_0 = 0 \) and \( K_1 = \infty \). For each \( t \in T \) and \( x \in X \), use the law of iterated expectations to write

\[
E[y(t)|x] = E[y(t)|x, s = t] \cdot P(s = t|x) + E[y(t)|x, s \neq t] \cdot P(s \neq t|x). \quad (10)
\]

Empirical knowledge of \( P[x, s, y(s)] \) implies knowledge of \( E[y(t)|x, s = t] \), \( P(s = t|x) \), and \( P(s \neq t|x) \) but reveals nothing about \( E[y(t)|x, s \neq t] \). We know only that the last quantity lies in the interval \( [K_0, K_1] \). Hence \( E[y(t)|x] \) lies within this sharp bound:

\[
E[y(t)|x, s \neq t]P(s = t|x) + K_0P(s \neq t|x) \leq E[y(t)|x]
\]

\[
\leq E[y(t)|x, s = t] \cdot P(s = t|x) + K_1P(s \neq t|x). \quad (11)
\]

Now let us compare two treatment rules. Under one rule, all persons with covariates \( x \) receive treatment \( t' \). Under the other rule, all such persons receive a different treatment, say \( t'' \). In the absence of any empirical evidence on
treatment response, we would be able to say only that $E[y(t')|x] - E[y(t)|x]$ lies in the interval $[K_0 - K_1, K_1 - K_0]$. With the available empirical evidence, (11) yields a narrower bound on $E[y(t'')|x] - E[y(t)|x]$. The sharp lower (upper) bound is the lower (upper) bound on $E[y(t'')|x]$ minus the upper (lower) bound on $E[y(t)|x]$. Thus

$$E[y(t'')|x, s = t'']P(s = t''|x) + K_0P(s \neq t'')$$

$$- E[y(t')|x, s = t']P(s = t'|x) = K_1P(s \neq t'|x)$$

$$\leq E[y(t'')|x, s = t'']P(s = t''|x) + K_1P(s \neq t''|x)$$

$$- E[y(t')|x, s = t']P(s = t'|x) - K_0P(s \neq t'|x).$$

(12)

This bound is a subset of the interval $[K_0 - K_1, K_1 - K_0]$. Its width is $(K_1 - K_0)[P(s \neq t''|x) + P(s \neq t'|x)]$, which can be no smaller than $(K_1 - K_0)$. Hence the lower bound in (12) is necessarily non-positive and the upper bound is necessarily non-negative. Thus the empirical evidence alone does not reveal which treatment, $t'$ or $t''$, yields the larger mean outcome. The same reasoning holds for all pairs of treatments and for all values of $x$. Hence all feasible treatment rules are undominated.

It is important to understand that this harshly negative finding does not imply that the planner should be paralyzed, unwilling and unable to choose a treatment rule. What it does imply is that, using empirical evidence alone, the planner cannot claim optimality for whatever treatment rule he does choose. The planner might, for example, apply the maximin rule. This calls for each person with covariates $x$ to receive the treatment that maximizes the lower bound in (11). The planner cannot claim that this rule is optimal, but he may find some solace in the fact that it fully protects against worst-case scenarios.

4.3. Using assumptions to identify mean treatment response

Although there are fundamental limits to the observability of response functions, there are no limits other than internal consistency to the assumptions that one can impose. Further conclusions about the mean response functions $E[y(t')|x], x \in X$ can be deduced, and ambiguity in treatment choice reduced, if empirical knowledge of $P[x, s, y(s)]$ is combined with maintained assumptions.

The prevailing practice in the econometrics literature on treatment response has been to combine observations of realized (covariates, treatments, outcomes) with assumptions strong enough to identify mean response functions. Researchers applying these strong assumptions, however, have commonly found it difficult to justify them. There is a need to face up to the fact that imposing assumptions that are not credible does not really eliminate ambiguity in treatment choice. I discuss the three main approaches below.
4.3.1. Exogenous treatment selection

Certainly the most well known and often used way to identify mean response functions is to impose the non-testable assumption\(^8\)

\[
E[y(t)|x] = E[y(t)|x, s = r]. \tag{13}
\]

Empirical knowledge of \(P[x, s, y(s)]\) implies knowledge of the right side of (13). Hence \(E[y(t)|x]\) is identified.

Researchers asserting assumption (13) may say that treatment selection is exogenous or random or ignorable conditional on \(x\). Researchers often say that conditioning on the covariates \(x\) ‘controls for’ treatment selection. This statement is misleading. It suggests that \(x\) is a treatment, a variable that can somehow be controlled.

The exogenous treatment selection assumption is well-motivated in classical randomized experiments (Fisher, 1935). Here the status quo treatment rule \(s(\cdot)\) involves a planner who randomly assigns treatments to the members of the treated population, all of whom comply with the assigned treatment. Hence \(s\) is necessarily statistically independent of \([x, y(\cdot)]\). Eq. (13) is an immediate consequence of this statistical independence.

The assumption is typically difficult to motivate in experiments that deviate from the classical ideal (Hausman and Wise, 1985; Heckman, 1992; Moffitt, 1992; Manski, 1996) and in non-experimental settings, especially those in which the status quo treatments are self-selected by the members of the treated population (Gronau, 1974). In these cases, the assumption is often no more than an imputation rule (Section 2.2). Researchers sometimes assert that the assumption does hold, or at least is a good approximation, if \(x\) is a sufficiently rich set of covariates. This assertion is usually made without a clear supporting argument.

4.3.2. Latent-variable models

When the status quo treatments are self-selected, it is easier to argue that treatment selection is not exogenous than to find a credible alternative assumption that identifies mean outcomes. Some researchers have proposed latent-variable models that jointly explain treatment and response. These models make assumptions about the form of the distribution \(P[s, y(\cdot)|x]\) of status quo treatments and response functions, conditional on the covariates. If the assumptions are sufficiently strong, combining them with empirical knowledge of \(P[x, s, y(s)]\) identifies the mean outcomes \(E[y(t)|x]\). See, for example, Maddala (1983), Björkland and Moffitt (1987), and Heckman and Honore (1990).

The use of latent-variable models to identify treatment effects has been quite controversial. Some researchers have regarded these models as ill-motivated.

---

\(^8\) Assumption (13) is not testable because \(E[y(t)|x, s \neq r]\) is not observable. Hence there is no empirical basis for refutation of the hypothesis \(E[y(t)|x, s \neq r] = E[y(t)|x, s = r]\), which implies (13).
imputation rules whose functional form and distributional assumptions lack foundation (e.g., Lalonde, 1986; Wainer, 1989). Others have viewed them as credible assumptions (e.g., Heckman and Hotz, 1989).

4.3.3. Instrumental variable assumptions and constant treatment effects

In situations where outcomes are continuous rather than discrete, mean outcomes can be identified by combining an instrumental variable assumption with the assumption of constant treatment effects. The classical research on linear response models that began in the 1920s and crystallized by the early 1950s invokes these assumptions (Hood and Koopmans, 1953).

An instrumental variable assumption holds that mean response is constant across sub-populations defined by different values of some covariate. The assumption is non-trivial if the status-quo treatments do vary with this covariate. The constant-treatment-effect assumption is that the response functions \( y_j(t), j \in J \), are parallel to one another. That is, there exists a function \( v(\cdot) : T \to \mathbb{R} \) and a set of real constants \( \alpha_j, j \in J \), such that

\[
y_j(t) = v(t) + \alpha_j.
\]

The controversy surrounding latent-variable models re-appears in applications that assume constant treatment effects. Whereas applied researchers sometimes feel that they can plausibly assert an instrumental variable assumption, the assumption of constant treatment effects usually strains credibility. In particular, this assumption implies that it is optimal to assign the same treatment to every member of the population, namely the treatment that maximizes \( v(\cdot) \) on \( T \).

Recent research stresses that the population may be heterogeneous. Treatment effects may vary from person to person and the optimal treatment rule may assign treatments that vary across persons with different covariate values.\(^9\) See Bloom (1984), Heckman and Robb (1985), Björkland and Moffitt (1987), Robinson (1989), Manski (1990), Imbens and Angrist (1994), Manski

---

\(^9\) For example, a longstanding concern of labor economics is to determine the effect of union membership on wages. There are two treatments, with \( t = 1 \) denoting union membership and \( t = 0 \) otherwise. Let \( y_j(1) \) be the wage that person \( j \) would earn if she were a union member and \( y_j(0) \) be the wage that \( j \) would earn as a non-member. Then constant treatment effects mean that the union wage differential \( y_j(1) - y_j(0) \) is the same for all \( j \in J \).

Is it plausible to assume that union membership gives the same wage increment to all workers? Union contracts are often thought to tie wages and job security more closely to seniority than to merit. If so, then within a given job category, the less productive workers should experience a larger union wage differential than do the more productive ones. It may be that the less productive workers have positive union wage differentials and the more productive workers have negative differentials.
5. Treatment choice using experiments without covariate data

The discussion in Section 4 suggests a stark tension. Observation of a treated population reveals something about mean treatment response but not enough to conclude that any rule is dominated. Empirical knowledge combined with strong assumptions can identify mean response, but these assumptions are only occasionally credible.

A planner who can combine observations of the treated population with credible assumptions about the process generating realized treatments and outcomes may be able to deduce non-overlapping bounds for mean outcomes under some alternative treatments and, hence, may be able to partially order the feasible treatment rules. There is a vast range of interesting situations yielding bounds that are potentially non-overlapping. Some of these may be investigated using the various bound results reported in Balke and Pearl (1999), Horowitz and Manski (1998), Hotz et al. (1997), Manski (1990, 1994–1997a, b), Manski and Nagin (1998), Manski et al. (1998), Manski and Pepper (1999), Robins (1989), and Robins and Greenland (1996). In this Section I develop the main ideas through examination of one situation of substantive and methodological interest. I examine the problem faced by a planner who observes the treatments and outcomes realized in a classical randomized experiment, but who does not observe the covariates of the experimental subjects.

5.1. The planner’s problem

In Section 4.3, I observed that the exogenous treatment selection assumption (13) is credible if the members of the treated population are subjects in a classical randomized experiment. The discussion there assumed that the (covariates, treatments, outcomes) realized by the experimental subjects are observable. Here I consider the problem of treatment choice when the planner observes subjects’ realized treatments and outcomes but not their covariates $x$.

This informational situation is common in medical and other settings when new treatments are approved for use following a period of experimentation. Consider, for example, a physician who must choose treatments for a population of heterogeneous patients. Physicians commonly have extensive covariate information – medical histories, diagnostic test findings, and demographic attributes – for the patients that they treat. Physicians often know the outcomes of randomized clinical trials evaluating new treatments. The medical journal articles that report the findings of clinical trials, however, do not usually report extensive covariate information for the subjects of the experiment. Articles
reporting on clinical trials usually describe outcomes only within broad risk-factor groups.\textsuperscript{10}

To grasp the essence of the planner’s problem, it suffices to consider the simplest non-trivial setting; that in which treatments, outcomes, and covariates are all binary. Thus I henceforth suppose that there are two treatments, say $t = 0$ and $t = 1$. The outcome $y(t)$ is binary, taking the values $y(t) = 0$ and $y(t) = 1$; hence $E[y(t)|x] = P[y(t) = 1|x]$. The covariate $x$ is also binary, taking the values $x = a$ and $x = b$.

Even in this simple setting, analysis of the planner’s problem turns out to be complex. There are four feasible treatment rules. These rules and their mean outcomes are

- **Treatment rule** $\tau(0, 0)$: All persons receive treatment 0. The mean outcome is $M(0, 0) = P[y(0) = 1]$.

- **Treatment rule** $\tau(1, 1)$: All persons receive treatment 1. The mean outcome is $M(1, 1) = P[y(1) = 1]$.

- **Treatment rule** $\tau(0, 1)$: Persons with $x = a$ receive treatment 0 and persons with $x = b$ receive treatment 1. The mean outcome is $M(0, 1) = P[y(0) = 1|x = a]P(x = a) + P[y(1) = 1|x = b]P(x = b)$.

- **Treatment rule** $\tau(1, 0)$: Persons with $x = a$ receive treatment 1 and persons with $x = b$ receive treatment 0. The mean outcome is $M(1, 0) = P[y(1) = 1|x = a]P(x = a) + P[y(0) = 1|x = b]P(x = b)$.

5.2. The dominated treatment rules

Which of the four feasible treatment rules are dominated? The experiment reveals $M(0, 0)$ and $M(1, 1)$. Thus rule $\tau(0, 0)$ is dominated if $M(0, 0) < M(1, 1)$.

\textsuperscript{10}There seem to be two reasons why experimental outcomes are commonly reported without much covariate information. First, researchers often seem to assume that treatment effects are constant across the population, or at least that mean treatment effects do not vary in sign across different subpopulations. (I say ‘seem to’ because the assumption is typically implicit, not explicit.) Given two treatments $t = 0$ and $t = 1$, suppose one knows a priori that there are only these two possibilities: either $\{E[y(1)|x] \geq E[y(0)|x], \forall x \in X\}$ or $\{E[y(1)|x] \leq E[y(0)|x], \forall x \in X\}$. Then collection of data on the covariate $x$ is not necessary to determine an optimal treatment rule. It suffices to learn if $E[y(1)]$ exceeds $E[y(0)]$.

Second, there is the matter of sampling variability. Researchers often perform randomized experiments with samples of subjects that are large enough to yield statistically precise findings for unconditional mean outcomes but are not large enough to yield precise findings for mean outcomes conditional on covariates. Findings conditional on covariates commonly go unreported if they do not meet conventional criteria for statistical precision.
and rule \( \tau(1, 1) \) is dominated if \( M(1, 1) < M(0, 0) \). The planner is indifferent between these two rules if \( M(0, 0) = M(1, 1) \).

The experiment does not reveal \( M(0, 1) \) and \( M(1, 0) \). Manski (1997a, Proposition 7), however, shows that the experiment in the treated population and the planner’s knowledge of the covariate distribution in the population of interest imply sharp bounds on these quantities. The derivation begins from the fact that

\[
P[y(0) = 1] = P[y(0) = 1|x = a]P(x = a) + P[y(0) = 1|x = b]P(x = b) = (15a)
\]

\[
P[y(1) = 1] = P[y(1) = 1|x = a]P(x = a) + P[y(1) = 1|x = b]P(x = b).
\]

(15b)

Consider (15a). The planner knows \( P[y(0) = 1] \) and \( P(x) \). The unknown quantities \( P[y(0) = 1|x = a] \) and \( P[y(0) = 1|x = b] \) both lie in the interval \([0, 1]\). These restrictions yield informative bounds on \( P[y(0) = 1|x = a] \) and \( P[y(0) = 1|x = b] \). Similarly, (15b) yields bounds on \( P[y(1) = 1|x = a] \) and \( P[y(1) = 1|x = b] \). The sharp bounds on \( M(0, 1) \) and \( M(1, 0) \) then follow immediately. These turn out to be (Manski, 1997a, Proposition 7)\(^{11}\)

\[
\max\{0, P[y(1) = 1] - P(x = a)\} + \max\{0, P[y(0) = 1] - P(x = b)\} \\
\leq M(0, 1) \leq \min\{P(x = b), P[y(1) = 1]\} + \min\{P(x = a), P[y(0) = 1]\}
\]

(16a)

\[
\max\{0, P[y(1) = 1] - P(x = b)\} + \max\{0, P[y(0) = 1] - P(x = a)\} \\
\leq M(1, 0) \leq \min\{P(x = a), P[y(1) = 1]\} + \min\{P(x = b), P[y(0) = 1]\}
\]

(16b)

The form of these bounds depends on the ordering of \( P[y(0) = 1] \), \( P[y(1) = 1] \), \( P(x = a) \), and \( P(x = b) \). Henceforth I assume without loss of generality that \( P[y(0) = 1] \leq P[y(1) = 1] \) and \( P(x = a) \leq P(x = b) \). Then there are six distinct orderings to be considered. For each ordering, application of (16a) and (16b) yields the bounds on \( M(0, 1) \) and \( M(1, 0) \). These bounds determine which treatment rules are dominated.

\(^{11}\) It might have been conjectured that \( M(0, 1) \) and \( M(1, 0) \) must lie in the interval \([M(0, 0), M(1, 1)]\). This is correct if treatment 0 is inferior to treatment 1 in both of the subpopulations \( \{x = a\} \) and \( \{x = b\} \); that is, if \( P[y(0) = 1|x = a] \leq P[y(1) = 1|x = a] \) and \( P[y(0) = 1|x = b] \leq P[y(1) = 1|x = b] \). However the conjecture is not correct if the ordering of the treatments differs across the two subpopulations. It is this possibility that gives rise to the surprisingly complex bounds on \( M(0, 1) \) and \( M(1, 0) \) reported in (16).
Rule $\tau(0, 0)$ is dominated by $\tau(1, 1)$ if $P[y(0) = 1] < P[y(1) = 1]$. The other results follow:

*Case 1:*

\[
P[y(0) = 1] \leq P[y(1) = 1] \leq P(x = a) \leq P(x = b)
\]

\[
0 \leq M(0, 1) \leq P[y(1) = 1] + P[y(0) = 1].
\]

Rules $\tau(0, 1)$, $\tau(1, 0)$, and $\tau(1, 1)$ are undominated.

*Case 2:*

\[
P[y(0) = 1] \leq P(x = a) \leq P[y(1) = 1] \leq P(x = b)
\]

\[
P[y(1) = 1] - P(x = a) \leq M(0, 1) \leq P(x = b) + P[y(0) = 1].
\]

\[
0 \leq M(1, 0) \leq P(x = a) + P[y(0) = 1].
\]

Rules $\tau(0, 1)$ and $\tau(1, 1)$ are undominated. Rule $\tau(1, 0)$ is dominated by rule $\tau(1, 1)$ if $P(x = a) + P[y(0) = 1] < P[y(1) = 1]$.

*Case 3:*

\[
P[y(0) = 1] \leq P(x = a) \leq P(x = b) \leq P[y(1) = 1]
\]

\[
P[y(1) = 1] - P(x = a) \leq M(0, 1) \leq P(x = b) + P[y(0) = 1].
\]

\[
P[y(1) = 1] - P(x = b) \leq M(1, 0) \leq P(x = a) + P[y(0) = 1].
\]

Rule $\tau(1, 1)$ is undominated. Rule $\tau(0, 1)$ is dominated by rule $\tau(1, 1)$ if $P(x = b) + P[y(0) = 1] < P[y(1) = 1]$. Rule $\tau(1, 0)$ is dominated by rule $\tau(1, 1)$ if $P(x = a) + P[y(0) = 1] < P[y(1) = 1]$.

*Case 4:*

\[
P(x = a) \leq P[y(0) = 1] \leq P[y(1) = 1] \leq P(x = b)
\]

\[
P[y(1) = 1] - P(x = a) \leq M(0, 1) \leq P[y(1) = 1] + P(x = a).
\]

\[
P[y(0) = 1] - P(x = a) \leq M(1, 0) \leq P(x = a) + P[y(0) = 1].
\]

Rules $\tau(1, 1)$ and $\tau(0, 1)$ are undominated. Rule $\tau(1, 0)$ is dominated by rule $\tau(1, 1)$ if $P(x = a) + P[y(0) = 1] < P[y(1) = 1]$.

*Case 5:*

\[
P(x = a) \leq P[y(0) = 1] \leq P(x = b) \leq P[y(1) = 1]
\]

\[
P[y(1) = 1] - P(x = a) \leq M(0, 1) \leq 1.
\]

\[
P[y(1) = 1] + P[y(0) = 1] - 1 \leq M(1, 0) \leq P(x = a) + P[y(0) = 1].
\]
This does not imply that $q(1, 1)$ is the optimal rule. The fact that $q(1, 1)$ is always undominated only means that, given the available information, the planner cannot reject the hypothesis that $q(1, 1)$ is the optimal rule.

Case 6:

$$P(x = a) \leq P(x = b) \leq P[y(0) = 1] \leq P[y(1) = 1]$$

$$P[y(1) = 1] + P[y(0) = 1] - 1 \leq M(0, 1) \leq 1.$$  

Rules $q(0, 1)$, $q(1, 0)$, and $q(1, 1)$ are undominated.

Cases 1–6 show that as many as three or as few as zero treatment rules are dominated, depending on the empirical values of $P[y(0) = 1]$, $P[y(1) = 1]$, $P(x = a)$, and $P(x = b)$. The one constancy is that rule $q(1, 1)$ is always undominated. Indeed, $q(1, 1)$ is always the maximin rule.

5.3. An empirical illustration: The Perry Preschool Project

An empirical illustration helps to see the range of possibilities. Beginning in 1962, the Perry Preschool Project provided intensive educational and social services to a random sample of low-income black children in Ypsilanti, Michigan. The project investigators also drew a second random sample of such children, but provided them with no special services. Subsequently, it was found that 67% of the treatment group and 49% of the control group were high-school graduates by age 19 (see Berrueta-Clement et al., 1984).

Let $t = 1$ denote the Perry Preschool treatment and $t = 0$ denote the ‘no special services’ control treatment. Let $y(t) = 1$ if a child receiving treatment $t$ is a high school graduate by age 19 and $y(t) = 0$ otherwise. Abstracting from sampling variability and ignoring some attrition from the experiment, the outcome data reveal that $P[y(0) = 1] = 0.49$ and $P[y(1) = 1] = 0.67$.

Consider the situation of a planner, perhaps a social worker, who is charged with making preschool treatment choices for low-income black children in Ypsilanti and whose objective is to maximize the high school graduation rate. The planner can assign each child to the Perry Preschool treatment or not. Suppose that the planner observes a binary covariate that describes each member of the population. For the sake of concreteness, let the covariate indicate the child’s family status, with $x = a$ if the child has an intact two-parent family and $x = b$ otherwise.

---

12 This does not imply that $q(1, 1)$ is the optimal rule. The fact that $q(1, 1)$ is always undominated only means that, given the available information, the planner cannot reject the hypothesis that $q(1, 1)$ is the optimal rule.
The available outcome data reveal that treatment rule \( \tau(0, 0) \), wherein no children receive the Perry Preschool treatment, is dominated by rule \( \tau(1, 1) \), wherein all children receive preschooling. The conclusions that the planner can draw about rules \( \tau(0, 1) \) and \( \tau(1, 0) \) depend on the covariate distribution \( P(x) \).

Suppose that half of all children have intact families, so \( P(x = a) = P(x = b) = 0.5 \). Then Case 3 of Section 5.2 holds. The bounds on mean outcomes under rules \( \tau(0, 1) \) and \( \tau(1, 0) \) are

\[
0.17 \leq M(0, 1) \leq 0.99, \quad 0.17 \leq M(1, 0) \leq 0.99.
\]

These bounds imply that rules \( \tau(0, 1) \) and \( \tau(1, 0) \), which reverse one another’s treatment assignments, have an enormously wide range of potential consequences for high school graduation. The best case for \( \tau(0, 1) \) and the worst for \( \tau(1, 0) \) both occur if the (unknown) graduation probabilities conditional on covariates are

\[
P[y(0) = 1|x = a] = 0.98, \quad P[y(1) = 1|x = a] = 0.34,
\]

\[
P[y(0) = 1|x = b] = 0, \quad P[y(1) = 1|x = b] = 1.
\]

These graduation probabilities, which yield \( M(0, 1) = 0.99 \) and \( M(1, 0) = 0.17 \), are consistent with the experimental evidence that \( P[y(0) = 1] = 0.49 \) and \( P[y(1) = 1] = 0.67 \). They describe a possible world in which preschooling is necessary and sufficient for children in non-intact families to complete high school, but substantially hurts the graduation prospects of children in intact families. There is another possible world with the reverse graduation probabilities, one in which \( M(0, 1) = 0.17 \) and \( M(1, 0) = 0.99 \). Hence rules \( \tau(0, 1) \), \( \tau(1, 0) \), and \( \tau(1, 1) \) are all undominated.

The planner faces a much less ambiguous choice problem if most children have non-intact families. Suppose that \( P(x = a) = 0.1 \) and \( P(x = b) = 0.9 \). Then Case 4 of Section 5.2 holds. The bounds on mean outcomes under rules \( \tau(0, 1) \) and \( \tau(1, 0) \) are

\[
0.57 \leq M(0, 1) \leq 0.77, \quad 0.39 \leq M(1, 0) \leq 0.59.
\]

These bounds are much narrower than those obtained when half of all children have non-intact families. The upper bound on \( M(1, 0) \) is 0.59, which is less than the known value of \( M(1, 1) \), namely 0.67. Hence treatment rule \( \tau(1, 0) \) is dominated. Recall that rule \( \tau(0, 0) \) is also dominated. Thus, although the planner does not observe graduation probabilities conditional on covariates, he can nevertheless conclude that the 90% of children who have non-intact families should receive preschooling. The only ambiguity about treatment choice concerns the 10% of children who have intact families. Treatment rules \( \tau(0, 1) \) and \( \tau(1, 1) \) are undominated. Thus, in the absence of other information, the planner cannot determine whether children in intact families should or should not receive preschooling.
6. Postscript: Statistical ambiguity

The ambiguity in treatment choice studied in this paper arises out of identification problems. In practice, planners may observe only a sample of the treated population. The problem of induction from sample to population then generates further ambiguity.

The econometrics literature on treatment response suggests that a planner should use sample data and assumptions that identify mean response functions to compute a point estimate of the mean response function conditional on the observed covariates. The planner can then choose for each person a treatment that maximizes the estimated conditional mean outcome. This practice is usually motivated by reference to classical asymptotic theory. If the estimates of mean responses are (weakly or strongly) consistent, then the treatment choices implied by those estimates converge (in probability or almost surely) to optimal treatment choices.

There is no escaping the fact that classical statistical theory, whether asymptotic or finite-sample, gives at most indirect guidance to a planner who must make treatment choices using sample data. The planner’s problem is to make good treatment choices given the particular data sample that he observes. Classical statistics, however, aims to characterize the sampling distributions of estimates under maintained assumptions about the population and the sampling process. Efforts to use information about these sampling distributions to guide treatment choices inevitably require a leap of deductive logic.

Bayesian decision theory puts forward an internally coherent program for decision making given sample data. A Bayesian planner begins by specifying a prior subjective probability distribution over all possible population distributions of covariates and response functions. He uses sample data to update this subjective distribution via Bayes’ Theorem. He then chooses a treatment rule that maximizes subjective expected utility.

A Bayesian planner avoids the classical leap of deductive logic, but achieves coherency only by modifying the planner’s objective function. The planner must accept the probabilistic logic of Bayesian analysis and specify a prior subjective distribution. He must specify a cardinal function transforming population mean outcomes into utilities and accept the idea that he should maximize subjective expected utility. After all of this, unfortunately, the Bayesian approach still does not answer the bottom-line question that I posed in Section 2.2: How well does the chosen treatment rule perform in maximizing the population mean outcome?

As I see it, the normative analysis of treatment choice using sample data remains wide open as a subject of inquiry.

Acknowledgements

This paper integrates two circulated earlier by the author. ‘Treatment Under Ambiguity’ was prepared for presentation at the Santa Fe Institute conference
on Fundamental Limits to Knowledge in Economics and is available as Santa Fe Institute Working Paper 96-10-078. ‘Treatment Choice in Heterogeneous Populations Using Experiments without Covariate Data’ was prepared for presentation at the Fourteenth Conference of the Association for Uncertainty in Artificial Intelligence and is available in the conference proceedings. This research was supported in part by National Science Foundation grants SBR-9223220 and SBR-9722846. I have benefitted from the comments of William Brock, Gary Chamberlain, Steven Durlauf, Arthur Goldberger, and John Pepper.

References


