TREATMENT UNDER AMBIGUITY

Charles F. Manski
Department of Economics
University of Wisconsin-Madison

September 1996

This paper was prepared for presentation at the 1996 Sante Fe Institute conference on Fundamental Limits to Knowledge in Economics. The research reported here is supported by National Science Foundation Grant SBR92-23220. I have benefitted from the comments of William Brock and Steven Durlauf.

Abstract

Economists have long associated decision making with optimization. The decision maker chooses an action from a known choice set C. The chosen action maximizes a known real-valued objective function $f(\cdot): C \rightarrow R$. Optimization assumes enough knowledge of C and $f(\cdot)$ to determine an optimal action. Suppose the decision maker knows C but not $f(\cdot)$. He knows only that $f(\cdot) \in F$, where F is a specified set of functions mapping C into R. Then the decision maker may not have enough information to determine an optimal action. This is a problem of decision under *ambiguity*.

After introducing basic themes about decision under ambiguity, I examine the problem of treatment choice. A social planner must choose a *treatment rule* assigning a treatment to each member of a population. Each person has some observed covariates and a *response function* mapping treatments into real-valued outcomes. The planner wants to choose treatments that maximize the population mean value of the outcome.

It has been conventional to assume that the planner knows (or at least can estimate) the population distribution of response functions conditional on covariates. With this knowledge, the planner faces a problem of decision under uncertainty and can choose an optimal treatment rule. There are, however, fundamental and practical limits to the knowledge of response functions that planners commonly possess. Thus planners choosing treatment rules ordinarily face problems of decision under ambiguity. This paper gives the key theoretical findings and considers the implications for treatment choice.

1. Introduction

Economists have long associated decision making with optimization. There is a universe A of actions. A decision maker chooses an action from a known choice set $C \subset A$. The chosen action maximizes on C a known real-valued objective function $f(\cdot): A \to R$.

Optimization assumes knowledge of C and $f(\cdot)$, or at least enough knowledge to determine an optimal action. Suppose that the decision maker knows the choice set but does not know the objective function. He knows only that $f(\cdot) \in F$, where F is a specified set of functions mapping A into R. Then the decision maker may not have enough information to determine an optimal action. This is a problem of decision under *ambiguity*.[1]

Economists have long recognized that decision makers may face ambiguity. See, for example, Knight (1921), Arrow and Hurwicz (1972), Maskin (1979), and Manski (1981). Nevertheless, study of the subject has remained a peripheral concern of the profession. The prevailing view seems to be that ambiguity is unusual or, perhaps, inconsequential.

In this paper I use a simple class of decision problems of considerable practical importance to show that ambiguity is both common and consequential. This is the problem of treatment choice studied by economists evaluating social programs, public health researchers comparing alternative medical treatments, and policy analysts more generally. The standard formalization of the problem

---

[1] The term *ambiguity* appears to originate with Ellsberg (1961), who used it to describe decision problems in which the objective function depends on an unknown probability distribution. The term has since been adopted by Einhorn and Hogarth (1986), Camerer and Weber (1992), and others. Much earlier, 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. Other authors have used *vagueness* and *ignorance* as synonyms for ambiguity.

supposes that a planner must choose a *treatment rule* assigning a treatment to each member of a population. Each person has some observed covariates and an unobserved *response function* mapping treatments into real-valued outcomes. The planner wants to choose treatments that maximize the population mean value of the outcome.

It has been conventional to assume that the planner somehow knows (or at least can estimate) the population distribution of response functions conditional on covariates. With this knowledge, the planner faces a problem of decision under uncertainty and can choose an optimal treatment rule. My recent program of research on the identification of treatment effects shows that there are limits, fundamental and practical, to the knowledge of response functions that planners commonly possess (Manski, 1990, 1994, 1995, 1996a, 1996b, 1996c). Thus planners choosing treatment rules ordinarily face problems of decision under ambiguity.

Section 2 develops basic themes about decision under ambiguity. Section 3 reviews the standard formulation of the planner's problem as treatment under uncertainty and then examines the problem of treatment under ambiguity. Section 4 shows that planners commonly face ambiguity of specific forms. Section 5 considers the desirability of the planner foregoing centralized selection of treatments and instead allowing the members of the population to select their own treatments. Section 6 gives conclusions.

Before proceeding, it is perhaps necessary to say that I use the term "knowledge" in the sense of the standard deductive logic of scientific inference. The decision maker draws logical conclusions by combining empirical evidence with maintained assumptions. These conclusions constitute knowledge.

2. The Basics of Ambiguity

Knowing that $f(\cdot) \in F$, how should the decision maker choose among the elements of the choice set C? 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 $[g(c) = g(d)$, all $g(\cdot) \in F]$ 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.

2.1. Rules Transforming Decisions under Ambiguity into Optimization Problems

Although there is no optimal decision under ambiguity, 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. Three leading proposals -- the maximin rule, Bayes rules, and imputation rules -- are discussed below. Although these proposals differ in their details, they share a key common feature. In each case, the solvable optimization problem $\max_{i \in D} h(\cdot)$ differs from the problem that the decision maker wants to solve, namely $\max_{i \in D} f(\cdot)$. The attained welfare level is $f[\operatorname{argmax}_{i \in D} h(\cdot)]$, not $\max_{i \in D} f(\cdot)$.

The Maximin Rule

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

Maximin Rule: For each $d \in D$, let $h(d) \equiv \inf_{g(\cdot) \in F} g(d)$. Maximize $h(\cdot)$ on D.

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."

Bayes Rules

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

Bayes Rule: Place a σ-algebra Σ and some probability measure π on the function space F. Let $h(\cdot) \equiv \int g(\cdot)d\pi$. Maximize $h(\cdot)$ on D.

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

Bayesians offer various procedural rationality arguments for use of a Bayes rules. These arguments do not, however, answer the question most relevant to a decision maker: 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 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) calls attention to this, stating (page 121): "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 π does not imply that the decision maker should choose an action that maximizes the π-average of the functions in F. Suppose that π has actually been used to draw $f(\cdot)$ from F; that is, let π describe an objective random process and not just the decision maker's subjective beliefs. Even here, where use of π as the prior distribution clearly is appropriate, Bayesian decision theory does not show that maximizing the π-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 alternatively choose an action that maximizes the π-median of F (see Manski, 1988) or some other measure of the central tendency of the π-distribution of F.

Imputation Rules

Bayesian decision theory at least faces up to the fact that the decision maker does not know the objective function $f(\cdot)$. 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 degenerate Bayes rules placing probability one on a single element of F.

2.2. 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 are no optimal actions and the decision maker is not indifferent among all undominated actions. There are some undominated actions that the decision maker cannot order.

This change of focus, albeit simple, has at least one striking 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. The Planner's Problem Under Uncertainty and Ambiguity

The study of decision under ambiguity can go only so far at the level of abstraction of Section 2. To develop further the themes introduced there, I now turn attention to the planner's problem of treatment choice.

3.1. The Choice Set and Objective Function

From here on I assume that each member j of a population J has some observable covariates $x_j \in X$ and an individual-specific *response function* $y_j(\cdot): T \rightarrow Y$ mapping the mutually exclusive and exhaustive *treatments* $t \in T$ into real-valued *outcomes* $y_j(t) \in Y$. I formalize the population as a probability space (J, Ω, P). Then $P[x, y(\cdot)]$ gives the population distribution of covariates and response functions.

A planner must choose a treatment for each member of the population. A *treatment rule* is a function $\tau(\cdot): J \rightarrow T$ specifying which treatment each person receives. Person j's outcome under rule $\tau(\cdot)$ is $y_j[\tau(j)]$. The population mean outcome under rule $\tau(\cdot)$ is

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

I assume that the planner wants to select a treatment rule to maximize $E\{y_j[\tau(j)]\}$. A planner could have other objectives, but maximization of mean outcome has long been the dominant concern of the literature on treatment choice.

Not all treatment rules are feasible to implement. The planner cannot distinguish among persons with the same observed covariates and so cannot implement treatment rules that systematically differentiate among such persons.[2] Thus the feasible treatment rules have the form

---

[2] The planner can randomly assign different treatments to persons with the same observed covariates. This possibility can be embraced by including in x a component whose value is randomly drawn by the planner from a specified distribution. The planner can make the chosen treatment vary with this covariate component. See Section 3.4 for further discussion.

$$(2)\ \tau(j) = z(x_j),$$

where $z(\cdot): X \rightarrow T$. Let Z denote the space of all functions mapping X into T. Then the planner wants to solve this optimization problem:[3]

$$(3) \quad \max_{z(\cdot) \in Z} E\{y[z(x)]\}.$$

Consider, for example, the problem of setting social policy directed at the population of unemployed persons. Each member of this population might receive one of three treatments: no public assistance ($t = 1$); publicly funded retraining ($t = 2$); or public assistance in job search ($t = 3$). The relevant outcome $y_j(t)$ might be life-cycle earned income net of treatment cost. The planner might observe each person j's age x_j. Then the feasible treatment rules are ones in which treatment may vary with age but not with other personal characteristics. The planner might want to choose a feasible rule to maximize mean net income.

3.2. Optimal Treatment Under Uncertainty

The planner is said to face a problem of decision under *uncertainty* if, in addition to observing each person's covariates, he knows the population distribution $P[x, y(\cdot)]$ of covariates and response functions. Observing each person's covariates implies that the planner knows the covariate distribution

---

[3] In practice, institutional or resource constraints may restrict the feasible treatment rules to a proper subset of Z. I abstract from this complication here. If problem (3) does not have a solution, the planner may have to suffice with selection of some "near-optimal" treatment rule. I abstract from this complication also.

P(x). So the essential new assumption is that the planner somehow knows the conditional response-function distributions $P[y(\cdot)|x]$, $x \in X$.

Knowledge of $P[x, y(\cdot)]$ makes problem (3) solvable. The optimal treatment rule is easily found. For each $z(\cdot) \in Z$, use the law of iterated expectations to write

$$(4)\quad E\{y[z(x)]\} = E\{E\{y[z(x)]|x\}\} = E\{\sum_{t \in T} E[y(t)|x]\cdot 1[z(x) = t]\}.$$

For each $x \in X$, the bracketted expression on the right side is maximized by choosing $z(x)$ to be a treatment that maximizes $E[y(t)|x]$ on $t \in T$. Hence the optimal treatment rule is[4]

$$(5)\quad z^*(x) = \underset{t \in T}{\operatorname{argmax}}\ E[y(t)|x], \quad x \in X$$

and the optimized population mean outcome is

$$(6)\quad V^* \equiv E\{\max_{t \in T} E[y(t)|x]\}.$$

3.3. Treatment Under Ambiguity

The planner may face a problem of decision under ambiguity if he has incomplete knowledge of $P[x, y(\cdot)]$. Suppose the planner knows only that

---

[4] If there are multiple maxima, $z^*(x)$ can be any selection from the maximizing set.

$P[x, y(\cdot)] \in \Phi$, where Φ is a specified set of (covariate, response function) distributions. Now problem (3) may not be solvable. I say "may not" because determination of an optimal treatment rule does not require complete knowledge of $P[x, y(\cdot)]$. It requires only that the planner know, for almost every $x \in X$, a treatment that maximizes $E[y(t)|x]$.

Under ambiguity, the planner can still partition the feasible treatment rules into dominated and undominated subclasses. A feasible treatment rule $z'(\cdot)$ is dominated if there exists another feasible rule, say $z''(\cdot)$, such that

$$(7) \quad \int y[z'(x)]d\phi \leq \int y[z''(x)]d\phi, \quad \text{all } \phi \in \Phi,$$

the inequality being strict for some $\phi \in \Phi$. Henceforth Z^* denotes the undominated subset of Z.

Observe that the sets Φ and Z^* are inversely related. As Φ expands to include more distributions, fewer treatment rules $z'(\cdot)$ satisfy (7). Thus the worse the problem of ambiguity, the smaller the set of treatment rules that the planner can eliminate as dominated.

3.4. Refining the Observed Covariates Under Uncertainty and Ambiguity

In Section 2.2 I called attention to the abstract possibility that expansion of the choice set may decrease welfare in decisions under ambiguity. I now show how this may happen in the treatment-choice setting.

The planner's choice set is the set of all functions mapping covariates into treatments. Thus the choice set expands if the planner observes some additional covariates, say $w_j \in W$, for each person j. Whereas previously the set of feasible treatment rules was the space of all functions mapping X into T, now the set of feasible rules is the space of all functions mapping $(X \times W)$ into T.

Under uncertainty, observation of additional covariates cannot decrease the optimized mean outcome. With x observed, the optimal treatment rule was (5) and the optimized mean outcome was (6). With (x, w) observed, the optimal rule is $\{\text{argmax}_{t \in T} E[y(t)|x, w], (x, w) \in X \times W\}$ and the optimized mean outcome is $E\{\max_{t \in T} E[y(t)|x, w]\}$. The new optimized mean outcome is necessarily at least as large as the previous one.

Under ambiguity, observation of additional covariates may decrease the mean outcome realized by the planner. This is particularly easy to see in some extreme cases. Suppose that the planner knows nothing about the distribution of response. Then all feasible rules are undominated. Also suppose that x is null. Then the only feasible treatment rules when w is unobserved are ones that give the same treatment to every person, which yield mean outcomes $E[y(t)]$, $t \in T$.

Now consider two polar possibilities for the additional covariates w. In the case of *conditionally homogeneous response*, all persons with covariates w have the same response function, say $y_w(\cdot)$. In the case of *independent response*, w is statistically independent of $y(\cdot)$.

Conditionally Homogeneous Response

Suppose that all persons with covariates w have the same response function $y_w(\cdot)$. Observation of w allows the planner to choose a treatment specific to each response function appearing in the population, but the planner does not know what these response functions are. If the planner happens to choose the worst treatment specific to each response function, the realized mean outcome is $E[\min_{t \in T} y_w(t)]$. If the planner happens to choose the best treatment specific to each response function, the realized mean outcome is $E[\max_{t \epsilon T} y_w(t)]$. By Jensen's inequality,

$$(8a) \quad E[\min_{t \epsilon T} y_w(t)] \leq \min_{t \epsilon T} E[y_w(t)]$$

$$(8b) \quad E[\max_{t \epsilon T} y_w(t)] \geq \max_{t \epsilon T} E[y_w(t)],$$

the inequalities being strict unless almost all members of the population have the same response function. Hence observation of w lowers the worst mean outcome and raises the best mean outcome realizable if w is not observed.

Independent Response

If w is independent of $y(\cdot)$, using w to choose treatments effectively randomizes treatment selection within the population. Let $\zeta(\cdot): W \rightarrow T$ be any treatment rule using w to assign treatments. Statistical independence of w and $y(\cdot)$ implies that the realized mean outcome is

$$(9) \quad E\{y[\zeta(w)]\} = E\{\sum_{t \epsilon T} E[y(t)|w]\cdot 1[\zeta(w) = t]\} = \sum_{t \epsilon T} E[y(t)]\cdot P[\zeta(w) = t].$$

Hence the mean outcome using w to assign treatments falls in between the worst and best mean outcomes realizable if w is unobserved.

4. The Structure of Response Ambiguity

In Section 3, we found that the planner's knowledge of the response-function distributions $P[y(\cdot)|x]$, $x \in X$ determines the form of the decision problem that he faces. Knowledge of response depends on the available empirical evidence and on the assumptions that the planner maintains. In this section I show that there are limits to the knowledge that planners may possess. These limits determine the structure of the ambiguity that planners face.

4.1. The Observability of Response Functions

It is generally thought, by scientists and planners alike, that empirical evidence is preferable to maintained assumptions as a basis for drawing conclusions. Unfortunately, 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 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 this person actually receives.[5]

Even the realized outcome is observable only retrospectively, after a

---

[5] 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), rather than 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 of consumer and firm behavior, where it is sometimes assumed that treatments are not mutually exclusive. Varian (1982, 1984), for example, supposes that an analyst observes multiple realized (treatment, outcome) pairs for a given individual j and uses these observations to learn about j's response function $y_j(\cdot)$.

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. 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 $s(\cdot)$: $K \rightarrow T$ denote the "status quo" treatment rule applied in the treated population. Then the realized (covariate, treatment, outcome) triples $\{x_k,\ s(k),\ y[s(k)];\ k \in K\}$ are observable in principle. Under the maintained 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.

I have just said that the treated population is observable "in principle." In practice, researchers often observe only a sample of the treated population, perhaps a random sample, from which $P[x, s, y(s)]$ may be estimated. To keep attention focused on the fundamental problem of mutual exclusivity of treatments, I shall abstract from the statistical issues that arise in finite-sample inference. The reader should keep in mind that a planner who can only estimate $P[x, s, y(s)]$ faces ambiguity beyond what is examined here.

4.2. Treatment Choice Using The Empirical Evidence Alone

What is the set of undominated treatment rules given empirical knowledge of P[x, s, y(s)] but no maintained assumptions about the distribution of response? I showed in Manski (1990) that this question has a simple but unpleasant answer. That is, all feasible treatment rules are undominated.

Let Y_0 and Y_1 denote the lower and upper endpoints of the logical range of the response functions. If outcomes are binary, for example, then $Y_0 = 0$ and $Y_1 = 1$. If outcomes can take any non-negative value, then $Y_0 = 0$ and $Y_1 = \infty$. For each $t \in T$ and $x \in X$, a sharp bound on the mean outcome $E[y(t)|x]$ is obtained by using the law of iterated expectations to write

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

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 $[Y_0, Y_1]$. Hence $E[y(t)|x]$ lies within this sharp bound:

$$(11)\quad E[y(t)|x, s = t]\cdot P(s = t|x) + Y_0\cdot P(s \neq t|x) \leq E[y(t)|x] \\ \leq E[y(t)|x, s = t]\cdot P(s = t|x) + Y_1\cdot P(s \neq t|x).$$

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 $[Y_0 - Y_1, Y_1 - Y_0]$. With the available empirical evidence, (11) yields the sharp bound on $E[y(t'')|x] - E[y(t')|x]$. The lower (upper) bound is the lower (upper) bound on $E[y(t'')|x]$ minus the upper (lower) bound on $E[y(t')|x]$. Thus

$$(12)\quad \begin{aligned} & E[y(t'')|x,\ s = t'']\cdot P(s = t''|x) + Y_0\cdot P(s \neq t''|x) \\ & \quad - E[y(t')|x,\ s = t']\cdot P(s = t'|x) - Y_1\cdot P(s \neq t'|x) \\ & \leq E[y(t'')|x] - E[y(t')|x] \\ & \leq E[y(t'')|x,\ s = t'']\cdot P(s = t''|x) + Y_1\cdot P(s \neq t''|x) \\ & \quad - E[y(t')|x,\ s = t']\cdot P(s = t'|x) - Y_0\cdot P(s \neq t'|x). \end{aligned}$$

This bound is a subset of the interval $[Y_0 - Y_1, Y_1 - Y_0]$. Its width is $(Y_1 - Y_0)\cdot[P(s \neq t''|x) + P(s \neq t'|x)]$, which can be no smaller than $(Y_1 - Y_0)$. 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 applied to persons with covariates x 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). Thus

$$(13)\quad z_{maximin}(x) = \underset{t \in T}{\operatorname{argmax}}\ E[y(t)|x,\ s = t]\cdot P(s = t|x) + Y_0\cdot P(s \neq t|x), \qquad x \in X.$$

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 Outcomes

Although there are fundamental limits to the observability of response functions, there are no limits other than internal consistency to the assumptions about treatment response that researchers can impose. The mean outcomes $E[y(t)|x]$, $t \in T$, $x \in X$ can be deduced, and ambiguity thus eliminated, if empirical knowledge of $P[x, s, y(s)]$ is combined with sufficiently strong maintained assumptions. Econometricians and other methodologists have developed an extensive body of such results. Examination of three leading cases indicates the range of approaches taken.

Exogenous Treatment Selection

Certainly the most common and longstanding practice is to assume that the mean of $y(t)$ among those persons who actually receive treatment t equals the mean of $y(t)$ among all persons with covariates x. That is,

$$(14)\quad E[y(t)|x] = E[y(t)|x,\ s = t].$$

This empirically nontestable assumption is variously called *exogenous* or *random*