Center for Demography and Ecology

University of Wisconsin-Madison

Bounding Disagreements About Treatment Effects: A Case Study of Sentencing and Recidivism

Charles F. Manski
Daniel S. Nagin

CDE Working Paper No. 95-29

BOUNDING DISAGREEMENTS ABOUT TREATMENT EFFECTS:
A CASE STUDY OF SENTENCING AND RECIDIVISM

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

and

Daniel S. Nagin
Heinz School of Public Policy
Carnegie-Mellon University

November 1995

This research is supported by National Science Foundation Grant SBR92-23220. We are grateful to the National Juvenile Court Data Archive for permission to use the Utah Juvenile Court Case Records. We thank John Pepper for superb research assistance.

Abstract

Empirical inference on treatment effects is a core objective of social science research. The conventional practice is to obtain point estimates of treatment effects using parametric models that make strong and thereby controversial identification assumptions. In this paper we obtain bounds under weaker nonparametric assumptions and explore the sensitivity of such bounds to alternative assumptions about treatment selection and response to treatment. This mode of analysis clarifies the source of common disagreements about the magnitudes and signs of treatment effects. We use a treatment question facing the juvenile justice system to showcase the value of the approach in empirical social science research. We compare the impacts on recidivism of the two main sentencing options available to judges: confinement in residential treatment facilities and diversion to nonresidential treatment.

1. Introduction

Empirical inference on treatment effects is a core objective of social science research. All efforts to infer treatment effects, whether from observational or experimental data, must confront the fact that the data are inherently incomplete. One wants to compare the outcomes of mutually exclusive treatments but each person being observed can experience at most one of the treatments being compared.

Identification of treatment effects necessarily requires assumptions about the process determining treatment selection and outcomes. The most longstanding practice, and still the most prevalent one, is to assume that among those persons with specified observable covariates, treatment selection is statistically independent of outcomes. This assumption is variously called *random*, *exogenous* or *ignorable* treatment selection. The specified covariates are often said to "control for" treatment assignment (see Maddala, 1983; and Rosenbaum and Rubin, 1983).

The assumption of random treatment selection is appropriate in the analysis of data from classical randomized experiments, but is suspect when observed treatments are self-selected or otherwise chosen purposefully. In the past twenty years, a variety of alternative assumptions have been proposed and applied to nonexperimental data. Parametric latent-variable models specify the joint distribution of treatments and outcomes up to a finite set of parameters, while parametric instrumental-variables approaches assume that treatment effects are constant across the population and that a specified covariate influences treatment selection but not outcomes (Heckman, 1976, 1978; Maddala, 1983; Winship and Mare, 1992). The development of latent-variable models and instrumental-variables methods was initially greeted with some enthusiasm as "solving" the problem of identifying treatment effects from nonexperimental data. It soon became apparent, however, that these approaches replace the suspect assumption of random treatment selection with alternative assumptions that are no less suspect (see Goldberger, 1983; LaLonde, 1986).

The methodological research program initiated in Manski (1989, 1990) and exposited most fully to date in Manski (1993, 1995) moves away from the conventional focus of social science research on parametric models that yield exact identification of treatment effects. This work shows that informative bounds may be obtained under weak nonparametric assumptions. An important objective is to bound disagreements among researchers. Inferences predicated on weak assumptions can achieve wide consensus. Inferences that require strong assumptions almost inevitably are subject to sharp disagreements.

This paper uses a treatment question facing the juvenile justice system to showcase the value of the new approach in empirical social science research. The question of how judges should sentence convicted juvenile offenders has long been of interest to policy makers, social scientists, and criminologists. Here we compare the impacts on recidivism of the two main sentencing options available to judges: confinement in residential treatment facilities and diversion to nonresidential treatment.

Our empirical analysis exploits the rich event-history data on juvenile offenders collected by the state of Utah (National Juvenile Court Data Archive, 1992). We present several sets of findings and show how conclusions about treatment effects vary depending on the assumptions made. First considering two extreme cases, we present estimates under the assumption of random treatment selection and bounds obtained without making any assumptions at all about the process determining treatment selection and outcomes.

We then present bounds obtained under two alternative models of judicial decision making. The *outcome optimization* model assumes judges make sentencing decisions that minimize the chance of recidivism. The *skimming* model assumes that judges classify offenders as "higher risk" or "lower risk," sentencing the former to residential treatment and the latter to nonresidential treatment. Both models express common hypotheses about judicial decision making without imposing the poorly motivated functional-form and distributional assumptions that mar parametric latent-variable models and instrumental variables methods. Both models have a wide range of social science applications. Our analysis of the outcome optimization model extends work described in Manski (1990, 1995). The identifying power of the skimming model has not previously been examined.

Finally, we bring to bear further prior information in the form of *exclusion restrictions*. These are assumptions positing that various subpopulations of offenders characterized by different covariate values have the same response to treatment but may face different treatment selection rules. The covariates defining these subpopulations are called *instrumental variables*.

Abstracting from the specifics of our juvenile-justice application, this paper demonstrates the value for informed policy and scientific debate of analyzing treatment effects under a sequence of progressively stronger assumptions. We believe that reporting "layered" empirical findings in this manner improves upon the conventional practice of reporting only point estimates based on assumptions strong enough to identify treatment effects. We also hope that this paper helps to demystify nonparametric statistical methods. Indeed we would argue that our nonparametric estimates of treatment effects are easier to interpret than conventional parametric estimates.

The paper is organized as follows. Section 2 provides background on the juvenile justice application. Section 3 uses this application to motivate our description of the basic ideas underlying nonparametric bounding of treatment effects. Section 4 presents the outcome optimization and skimming models and discusses their implications for identification of treatment effects. Section 5 explains the identifying power of exclusion restrictions. In each of Sections

3 through 5, the empirical application follows presentation of the relevant theory. Section 6 sums up.

2. Background on the Application

2.1. The Impact of the Juvenile Justice System on Delinquency

There is a longstanding debate among scholars of the juvenile justice system about the efficacy of formal processing compared to nonjudicial remediation in preventing future delinquency. The juvenile justice system is rooted in a tradition of using its power and resources for the purpose of remediating the causes of delinquency (Benard, 1992; Cullen, et al., 1988; Rothman, 1980). Accordingly, youth who are remanded to the system are commonly required to participate in counseling and educational programs intended to provide the individual with the capacity to resist the temptations of delinquency and to provide alternatives to delinquency. The vast literature reporting on evaluations of such programs (Andrews, 1990; Sechrist, et al.; Whitehead and Lab, 1989) is testimony to the search for effective programs of remediation.

In opposition to this tradition, labeling theorists (Lemert, 1951; Becker, 1963) argue that regardless of intentions, formal processing and treatment of juvenile offenders is socially stigmatizing and thereby hinders reintegration into conventional nondelinquent lifestyles. According to labeling theorists, formal processing has the reverse of the desired effect. It neither rehabilitates nor deters. Rather, it results in "deviance amplification."

There is no shortage of empirical research. The problem is the interpretation of the evidence. Researchers generally find that offenders who are subjected to formal judicial processing tend to have higher recidivism rates than ones who are treated informally (Smith and Paternoster, 1990). While this finding accords with the prediction of labeling theory, it may also be an artifact of treatment selection. Smith and Paternoster (1990, page 1111) observe: "....high risk youth are more likely to receive more severe dispositions. Thus, those individuals assigned more severe sanctions would be more likely to commit new offenses whether or not any relationship existed between juvenile court disposition and future offending." They go on to argue that it is implausible to assume that treatment selection is random conditional on the covariates that researchers typically can measure.

2.2. Data

Our analysis is based on data for male offenders born from 1970 through 1974, extracted from the state of Utah's extensive records on youth contacts with its juvenile justice system. The National Juvenile Court Data Archive has organized these records into files which, for each birth cohort from 1962 to 1974, record all delinquency, status offense, and abuse and neglect cases referred for court intake. The data for each case include the dates of referral and disposition and the reasons for and disposition of the referral. Also recorded are the age, race, sex, and number of prior referrals of the youth who is the subject of the case.

Abuse and neglect cases are not relevant for our purpose. We restrict attention to referrals that are not ultimately dismissed and in which the most serious charge involves an act that would be criminal if committed by an adult. For this segment of the case population, two broad categories of disposition are available to juvenile court judges: placement in a residential facility or assignment to nonresidential treatment such as probation, restitution, or counseling.

We use the Utah data to bound the impact on recidivism of these two modalities of treatment. Recidivism is defined as the youth generating a subsequent referral that meets the above screening criteria for inclusion in the sample within the 24 month period following the date of treatment. Referral is a filtered indicator of a return to offending -- a referral occurs only when a youth is apprehended for an offense and the authorities deem it appropriate to refer the apprehended youth to the juvenile justice system. Our analysis assumes that the probability a new offense generates a referral does not vary with the offender's prior treatment by the juvenile justice system.

The date of treatment is defined as the date of disposition, and the date of recidivism as the date of referral of the subsequent qualifying case. Because the age of majority in Utah is 18, we consider only those cases in which the offender was less than age 16 at the date of treatment. After culling the data of cases in which it was not possible to determine an individual's age, treatment, offense, or exposure period following treatment, 13,197 cases were available for analysis.[1]

The 24-month exposure period includes the time that an offender sentenced to residential treatment is detained in the residential facility. In principle, residential treatment affects recidivism through two channels -- by incapacitating the offender while detained and by deterring or stigmatizing the offender after the sentence is completed. In practice, the incapacitative effect is limited by the non-secure nature of some residential facilities and by the short duration of juvenile residential sentences, which usually are less than six months in length (personal communication, National Juvenile Court Data Archive).

[1] The Utah data are remarkably free of the defects that commonly mar administrative data. Indeed the most common reason for a case being excluded from the analysis was not defective data but rather the ambiguity that arises when cases overlap in time. Specifically, there were instances in which a new case was filed following the referral but prior to the disposition of an earlier case. We exclude these overlapping cases because of the conceptual difficulty of defining treatment and time to failure.

3. Basic Analysis

3.1. The Classical and Status Quo Treatment Effects

There are many different ways to formalize the loose idea of a "treatment effect," and we shall be concerned with two of these. First, we examine the *classical treatment effect* that has long been the focus of empirical research, namely the expected value of the population difference in the outcomes of two mandated treatments. In our juvenile justice application, the classical treatment effect is the difference in the recidivism rate of juvenile offenders under two extreme treatment-selection rules. One extreme is residential treatment of all juvenile offenders with specified covariates, and the other is nonresidential treatment of all such offenders.

In practice, judges generally apply neither of these extreme treatment selection rules. The norm is for a judge to sentence some offenders to a residential program and others to nonresidential treatment. Policy makers may be more interested in assessing changes from this status quo than in the classical treatment effect. Many such changes are of potential interest. We examine one that is easily described and analyzed. This is the expected difference in recidivism if the treatment selection rules actually used by judges were replaced by one mandating residential treatment of all juvenile offenders. We refer to this as a *status quo treatment effect*.

For various reasons, one might be interested in treatment effects within an identifiable subset of the population, rather than in the entire population. For example, one might want to focus on juveniles convicted of committing a felony and ask how their recidivism would change if the status quo were replaced by mandatory residential treatment. Clearly, the problem of inferring treatment effects on identifiable subpopulations is no different conceptually from that of inferring treatment effects on the entire population. One simply redefines the "population" to include only the sub-population of interest. Indeed, as our empirical work demonstrates, a useful feature of the nonparametric methods applied here is that they make it simple to explore the possibility that treatment effects vary across segments of the population.

Throughout the paper, we reluctantly maintain the standard assumption of "individualistic treatment" made routinely, albeit often only implicitly, in analyses of treatment effects. Individualistic treatment means that each person's outcome may depend on the treatment he or she receives, but not on the treatments received by other persons. In our application, individualistic treatment means an absence of general deterrence effects of sentencing policy on recidivism and also an absence of social norm effects in which the stigma of being labelled a delinquent decreases with the prevalence of delinquency in the population. General deterrence, social norm effects, and other "macro effects" may well be important in practice, but their complexity has long led researchers analyzing treatment effects to abstract from them. See Garfinkel, Manski, and Michalopolous (1992).

3.2. The Problem of Identifying Treatment Effects

Estimation of treatment effects requires consideration of two distinct issues, identification and sampling variability. Identification concerns the conclusions that could be drawn if one could observe the treatments received and outcomes experienced by everyone in the population. Sampling variability arises when these data are available for only a sample of the population. Here our central concern is identification. To ease exposition, the discussion in this section makes no distinction between sample and population quantities. We shall, however, take due account of sampling variability when we present our empirical analysis.

Also to ease exposition, the present discussion supposes that the objective is to infer treatment effects on the entire population. Treatment effects on identifiable subpopulations will be estimated in our empirical application.

What quantities do the data reveal? To be concrete, consider the data set that we analyze, assembled from the records of the Utah juvenile justice system. For each juvenile offender, we know (1) the treatment received, residential or nonresidential, and (2) whether the offender recidivated and, if so, how long after treatment. The Utah data also reports various characteristics of each case (e.g., felony or not) and the offender (e.g., prior record) that may be used to define identifiable subsets of the population of offenders.

Abstracting from sampling variability, three features of the population (or of a subpopulation of interest) are revealed by the data. These are the probability that an offender is sentenced to residential treatment, the probability of recidivism among offenders who are residentially treated, and the probability of recidivism among offenders who are nonresidentially treated. Definition of recidivism requires specification of an exposure period (i.e., period at risk of failure) of interest. One might, for example, want to know about recidivism during the one year, two year, or five year period following conviction. Our notation leaves the exposure period implicit, in order to simplify the presentation of basic ideas.

The data reveal some but not all of the information necessary for estimating treatment effects. Two vital counterfactuals are missing. One is the probability of recidivism among offenders sentenced to residential treatment, had they instead been sentenced to nonresidential treatment. The other is the probability of recidivism among offenders sentenced to nonresidential treatment, had they instead been sentenced to residential treatment.

Let us formalize how these counterfactuals combine with the measurable quantities to determine treatment effects. Let $t = 1$ denote residential treatment and $t = 0$ denote nonresidential treatment. Let $y(t) = 1$ if an offender would recidivate during the exposure period if that offender were to receive treatment t, and let $y(t) = 0$ if the offender would not recidivate. Let $z = 1$ if an offender is actually sentenced to residential treatment and $z = 0$ if the offender is sentenced to nonresidential treatment. The three quantities revealed by the data are

$P(z = 1)$ -- This is the probability of residential treatment, which can be measured directly from the data by the fraction of offenders who are sentenced to residential facilities.

$P[y(1) = 1|z = 1]$ -- This is the probability of recidivism under treatment 1, among those offenders who actually receive treatment 1. This quantity can be measured directly from the data by the rate of recidivism among offenders sentenced to residential treatment.

$P[y(0) = 1|z = 0]$ -- This is the probability of recidivism under treatment 0, among those offenders who actually receive treatment 0. This quantity can be measured directly from the data by the rate of recidivism among offenders sentenced to nonresidential treatment.

The two counterfactual quantities not revealed by the data are

$P[y(1) = 1|z = 0]$ -- This is the probability of recidivism under treatment 1, among those offenders who actually receive treatment 0.

$P[y(0) = 1|z = 1]$ -- This is the probability of recidivism under treatment 0, among those offenders who actually receive treatment 1.

With this background, we can now define formally the classical and status quo treatment effects and pinpoint the problem of identifying each. The classical treatment effect, or CTE, is the difference between the recidivism probabilities that would occur under mandatory residential treatment of all juvenile offenders and mandatory nonresidential treatment of all offenders. Thus

$$(1) \quad CTE \equiv P[y(1) = 1] - P[y(0) = 1].$$

The first element of the CTE measures the probability of recidivism if all offenders were sentenced to residential treatment, and the second element measures the probability of recidivism if all were sentenced to nonresidential treatment.

The status quo treatment effect, or STE, is the difference between the recidivism probability that would occur under mandatory residential treatment of all juvenile offenders and the one that is observed under the actual sentencing rules used by judges. Thus

$$(2) \quad STE \equiv P[y(1) = 1] - P[y(z) = 1].$$

Observe that the STE has the same first element as the CTE but a different second element. The quantity $y(z)$ is the recidivism outcome experienced by an offender under the treatment that this offender actually receives. So $P[y(z) = 1]$ measures the probability of recidivism under the status quo sentencing rules used by judges.

The identification problem is that, of all the elements of the CTE and STE, the only one fully revealed by the data is P[y(z) = 1]. To see this, we use the law of total probability to write the three recidivism probabilities as follows:

Recidivism Probability Under Mandatory Residential Treatment

(3) $P[y(1) = 1] = P[y(1) = 1|z = 1]\cdot P(z = 1) + P[y(1) = 1|z = 0]\cdot P(z = 0)$

Recidivism Probability Under Mandatory Nonresidential Treatment

(4) $P[y(0) = 1] = P[y(0) = 1|z = 1]\cdot P(z = 1) + P[y(0) = 1|z = 0]\cdot P(z = 0)$

Recidivism Probability Under the Status Quo

(5) $P[y(z) = 1] = P[y(1) = 1|z = 1]\cdot P(z = 1) + P[y(0) = 1|z = 0]\cdot P(z = 0)$.

Consider equation (3). The missing quantity needed to identify the recidivism probability under mandatory residential treatment is the counterfactual probability $P[y(1) = 1|z = 0]$; all other quantities are revealed by the data. Similarly, in equation (4), the missing quantity is the counterfactual probability $P[y(0) = 1|z = 1]$. Only in equation (5) do the data reveal all the quantities needed to identify the recidivism probability of interest.

3.3. Bounds on Treatment Effects Using the Data Alone

If our data were obtained from a classical randomized experiment, or if we could argue convincingly that treatment selection was effectively random, then it would still be the case that the counterfactual probabilities in equations (3) and (4) would not be revealed by the data. But random selection of treatment implies that the counterfactual probability $P[y(1) = 1|z = 0]$ equals the measured recidivism probability $P[y(1) = 1|z = 1]$ of those offenders who actually receive residential treatment. Similarly, randomization implies that $P[y(0) = 1|z = 1] = P[y(0) = 1|z = 0]$. Hence, the classical and status quo treatment effects under the assumption of random treatment selection are

$$(6) \quad CTE \equiv P[y(1) = 1|z = 1] - P[y(0) = 1|z = 0]$$

and

$$(7) \quad STE \equiv P[y(1) = 1|z = 1] - P[y(z) = 1].$$

Our data is neither the product of a randomized experiment nor is it possible to argue convincingly that judges effectively sentence in a random manner. How then should we proceed?

A logical starting point is to assume nothing at all about the counterfactual probabilities. Even if no assumptions are made, we can still make progress toward identifying treatment effects. We know that each counterfactual probability must be no smaller than zero and no larger than one. This simple fact implies bounds on the recidivism probability under mandatory residential treatment.

The lower bound on $P[y(1) = 1]$ is derived by assuming that none of the offenders who receive nonresidential treatment would have recidivated had they received residential treatment. Thus, we set $P[y(1) = 1|z = 0] = 0$ in equation (3) and obtain $P[y(1) = 1|z = 1] \cdot P(z = 1)$ as the lower bound. The upper bound is derived by assuming that all of the offenders who receive nonresidential treatment would have recidivated had they received residential treatment. Thus, we set $P[y(1) = 1|z = 0] = 1$ in (3) and obtain $P[y(1) = 1|z = 1] \cdot P(z = 1) + P(z = 0)$ as the upper bound. Analogous reasoning applies to the recidivism probability under mandatory nonresidential treatment. To summarize, we have

No-Assumptions Bound on Recidivism Probability Under Mandatory Residential Treatment[2]

$$(8) \quad P[y(1) = 1|z = 1] \cdot P(z = 1) \leq P[y(1) = 1] \leq P[y(1) = 1|z = 1] \cdot P(z = 1) + P(z = 0).$$

No-Assumptions Bound on Recidivism Probability Under Mandatory Nonresidential Treatment

$$(9) \quad P[y(0) = 1|z = 0] \cdot P(z = 0) \leq P[y(0) = 1] \leq P[y(0) = 1|z = 0] \cdot P(z = 0) + P(z = 1).$$

To illustrate, suppose that 90 percent of offenders actually receive residential treatment and that 50 percent of these individuals recidivate. Because the actual treatment rate is so high, the data provide a good fix on the impact of pushing the treatment rate up to 100 percent. We can say, without any assumptions whatsoever about the treatment selection rules actually used by judges, that the recidivism rate under mandatory residential treatment would be no lower than $0.45 = (0.5)(0.9)$ and no higher than $0.55 = (0.5)(0.9) + (1)(0.1)$. The lower bound (upper bound) holds if none (all) of the 10 percent of offenders

---

[2] As used here and elsewhere in the paper, the phrase "no-assumptions bound" should not be taken to mean that our analysis makes no assumptions at all. We do, after all, maintain various assumptions -- treatment is individualistic, outcomes and treatments are correctly measured, and so on. The phrase means only that no assumptions are made that restrict the values of the counterfactual probabilities.

who actually receive nonresidential treatment would recidivate if they were instead to receive residential treatment.

These bounds on recidivism probabilities imply bounds on the classical and status quo treatment effects. Consider the classical treatment effect. The lower bound on this quantity is the lower bound on $P[y(1) = 1]$ minus the upper bound on $P[y(0) = 1]$. The upper bound on the CTE is the upper bound on $P[y(1) = 1]$ minus the lower bound on $P[y(0) = 1]$. Hence we have

No-Assumptions Bound on the Classical Treatment Effect

$$(10)\quad \{P[y(1) = 1|z = 1]\cdot P(z = 1)\} - \{P[y(0) = 1|z = 0]\cdot P(z = 0) + P(z = 1)\}$$
$$\leq \text{CTE}$$
$$\leq \{P[y(1) = 1|z = 1]\cdot P(z = 1) + P(z = 0)\} - \{P[y(0) = 1|z = 0]\cdot P(z = 0)\}.$$

To illustrate, consider again the numerical example. As previously shown, the recidivism probability under mandatory residential treatment can be bounded within a fairly tight interval, namely 0.45 to 0.55. Now consider the bound under the alternative of mandatory nonresidential treatment. Suppose that the observed recidivism probability among those receiving nonresidential treatment is 0.40. Then the lower bound on $P[y(0) = 1]$ is $.04 = (0.4)(0.1)$ and the upper bound is $0.94 = (0.4)(0.1) + (1)(0.9)$. This bound is very broad, 0.04 to 0.94, because only 10 percent of offenders actually receive nonresidential treatment. So the data only reveal the impact of this treatment for a small fraction of the population.

In this example, the classical treatment effect can take any value between -0.49 (i.e., 0.45 - 0.94) and 0.51 (i.e., 0.55 - 0.04). Observe that the width of the bound on the CTE is 1. This is not a happenstance. Inspection of (10) shows that, in the absence of assumptions that restrict the values of the counterfactual probabilities, the width of the bound on the CTE always is 1.

If no data were available, we would only be able to make the trivial statement that the CTE must be no smaller than -1 and no larger than 1. The data allow us to narrow this interval of width 2 to one of width 1. In a very real sense then, the data move us half way toward identification of the CTE. Unfortunately, the data alone are not sufficiently informative to identify the sign of the CTE. Having width 1, the bound on the CTE always contains the value zero. Thus, identification of the sign of the CTE necessarily requires assumptions about the treatment selection and outcome process.

It remains to consider the status quo treatment effect. The lower bound on this quantity is the lower bound on $P[y(1) = 1]$ minus the known value of $P[y(z) = 1]$, given in (5). The upper bound on the STE is the upper bound on $P[y(1) = 1]$ minus the known value of $P[y(z) = 1]$. Hence we have

No-Assumptions Bound on the Status Quo Treatment Effect

$$(11) \quad -P[y(0) = 1 | z = 0]\cdot P(z = 0) \leq STE \leq P[y(0) = 0 | z = 0]\cdot P(z = 0).$$

The width of this bound is $P(z = 0)$, which is less than the width of the bound on the CTE. Nevertheless this bound, like the earlier one, always contains the value zero. So the sign of the STE is not identified using the data alone. In our numerical example, the status quo recidivism probability is 0.49 = (0.5)(0.9) + (0.4)(0.1). Thus, the STE is no smaller than -0.04 (i.e., 0.45 - 0.49) and no larger than 0.06 (i.e., 0.55 - 0.49).

3.4. Application

Table 1 derives the no-assumptions bounds on the CTE and STE for the combined 1970 to 1974 Utah cohorts. As described in Section 2.2, these individuals generated 13,197 cases meeting our screening criteria. The bounds are computed for an exposure period of 24 months.

The data reveal that 11 percent of the offenders are sentenced to residential treatment, the remaining 89 percent being sentenced to nonresidential treatment. Recidivism rates are high regardless of treatment, with 77 percent of the offenders sentenced to residential treatment recidivating within 24 months, and 59 percent of those sentenced to nonresidential treatment. Thus, abstracting from sampling variability, $P(z = 1) = 0.11$, $P[y(1) = 1|z = 1] = 0.77$, and $P[y(0) = 1|z = 0] = 0.59$.

Because only 11 percent of offenders are sentenced to residential treatment, the bound on the recidivism probability under mandatory residential treatment, $P[y(1)=1]$, is very wide -- at least 0.08 and no more than 0.97. With 89 percent of the population receiving nonresidential sentences, the bound under mandatory nonresidential treatment is comparatively narrow -- $P[y(0) = 1]$ must lie between 0.53 and 0.64. These bounds on the recidivism probabilities under the two mandatory sentencing regimes imply that the classical treatment effect is at least –0.56 but no more than 0.44. The status quo treatment effect is at least -0.53 but no more than 0.36.

The bounds in Table 1 pertain to the population of all cases. It is of interest to examine whether treatment effects vary with case characteristics. Labeling theory, for example, suggests that the adverse impact of contact with the juvenile justice system may diminish with the number of contacts because the stigmatization that triggers "secondary deviance" may be completed within the first few contacts. Indeed, once the initial stigmatization occurs, subsequent contact might have a deterrent effect.

Table 1: Basic Analysis (All Cases, 24 Month Exposure Period)

Probability of Residential Treatment: $P(z = 1) = 0.11$

Status Quo Recidivism Probability: $P[y(z) = 1] = 0.61$

Recidivism Probability in Subpopulation Receiving Residential Treatment:

$$P[y(1) = 1 | z = 1] = 0.77$$

Recidivism Probability in Subpopulation Receiving Nonresidential Treatment:

$$P[y(0) = 1 | z = 0] = 0.59$$

Classical Treatment Effect Assuming Random Treatment Selection:

$$0.77 - 0.59 = 0.18$$

Bound on Recidivism Probability Under Mandatory Residential Treatment:

$$0.08 = (0.77) \cdot (0.11) \leq P[y(1) = 1] \leq (0.77) \cdot (0.11) + 0.89 = 0.97$$

Bound on Recidivism Probability Under Mandatory Nonresidential Treatment:

$$0.53 = (0.59) \cdot (0.89) \leq P[y(0) = 1] \leq (0.59) \cdot (0.89) + 0.11 = 0.64$$

No-Assumptions Bound on Classical Treatment Effect:

$$-0.56 = 0.08 - 0.64 \leq CTE \leq 0.97 - 0.53 = 0.44$$

No-Assumptions Bound on the Status Quo Treatment Effect:

$$-0.53 = 0.08 - 0.61 \leq STE \leq 0.97 - 0.61 = 0.36$$

Sample Size: N = 13,197

Table 2 reports no-assumptions bounds on the CTE conditioned on the number of prior referrals of the offender, and 90 percent confidence intervals around the estimated bounds. The confidence intervals are computed using the bootstrap method described in Manski, Sandefur, McLanahan, and Powers (1992). Observe that the widths of the confidence intervals are only slightly larger than the widths of the estimated bounds. Thus, identification is the dominant problem in inference on these treatment effects; sampling variation is a second-order concern. This finding should not be surprising given the large size of the Utah dataset. Indeed, the same finding holds in all of our empirical analysis. Hereafter, we focus attention on the estimated bounds rather than on their confidence intervals.

Table 2: No-Assumptions Bound on Classical Treatment Effect, by Priors

Priors	Sample Size	Fraction Residential Treatment	CTE Assuming Random Treatment Selection	CTE		90 Percent Confidence Interval	
				L.B.	U.B.	L.B.	U.B.
0	7402	0.04	0.09	-0.48	0.52	-0.49	0.53
1	2719	0.11	0.07	-0.65	0.35	-0.67	0.36
2+	3072	0.27	0.02	-0.65	0.35	-0.66	0.37

Table 2 also reports point estimates of the CTE obtained under the assumption that treatment selection is random. These estimates are always positive, implying that residential treatment increases recidivism probability at all levels of priors, but the magnitude of the estimates declines with the number of priors. This pattern is consistent with the aforementioned prediction of labeling theory. Of course, this interpretation of the pattern of estimates falls apart if one is unwilling to maintain the assumption that treatment selection is random.

4. Two Models of Treatment Selection

Although the logical starting point for the analysis of treatment effects is to assume nothing about the counterfactual probabilities, researchers ordinarily want to make sharper inferences than those implied by the data alone. The price of sharper inferences is the loss of credibility associated with the imposition of untestable assumptions. Methodological research cannot determine when this price is worth paying, but it can assist empirical research by clarifying the identifying power of alternative assumptions.

In this section, we examine the identifying power of two treatment-selection assumptions that might perhaps describe actual judicial decision making. These are the *outcome optimization* model and the *skimming* model. In posing these models, our intention is not to endorse one or the other. It is rather to juxtapose the two and compare their implications for the counterfactual recidivism probabilities.

4.1. Outcome Optimization

The outcome optimization model operationalizes the economist's standard assumption that decisions under uncertainty are made to minimize expected loss or, equivalently, to maximize expected utility. In particular, this model assumes that, given the available information, the decision maker chooses the treatment that yields the better expected outcome.

The outcome optimization model obviously has an enormous range of possible applications, from selection of educational and training treatments, to family planning, to disease prevention. In the context of our juvenile justice application, the model assumes that the judge selects the sentencing alternative, residential or nonresidential treatment, that minimizes recidivism probability. A sizable contingent of scholars and policy makers have long held that the objective of the juvenile justice system should be to reform juvenile offenders, not punish them (Bernard, 1992; Cullen et al., 1988). The outcome optimization model assumes that judges actually sentence in accord with this normative view.

Formally, let s denote all the characteristics of the offender and environment that may be associated with treatment selection. That is, s is a vector of variables measuring the information available to the judge at the time of sentencing. Because s includes all the information available to the judge, the treatment z chosen by the judge necessarily is a function of s.

The outcome optimization model makes two assumptions. First, the judge knows the probabilities $P[y(1) = 1|s]$ and $P[y(0) = 1|s]$ of recidivism under the two possible treatments. Economists refer to this as the *rational expectations* assumption. Second, the judge selects the sentence yielding the smaller probability of recidivism. Thus,

$$(12a) \quad P[y(1) = 1|s] < P[y(0) = 1|s] \Rightarrow z = 1$$

$$(12b) \quad P[y(0) = 1|s] < P[y(1) = 1|s] \Rightarrow z = 0.$$

Now consider the situation of a researcher who assumes that sentencing decisions accord with (12). When the researcher observes that a judge sentences an offender to residential treatment, she can infer that $P[y(1) = 1|s]$ is no larger than $P[y(0) = 1|s]$. Likewise, when she observes that a judge sentences an offender to nonresidential treatment, she can infer that $P[y(0) = 1|s]$ is no larger than $P[y(1) = 1|s]$. Summarizing,

$$(13a) \quad z = 1 \quad \Rightarrow \quad P[y(1) = 1|s] \leq P[y(0) = 1|s]$$

$$(13b) \quad z = 0 \quad \Rightarrow \quad P[y(0) = 1|s] \leq P[y(1) = 1|s].$$

The treatment z is necessarily a function of s, so (13a) and (13b) are equivalent to these inequalities:

$$(14a) \quad P[y(1) = 1|s, z = 1] \leq P[y(0) = 1|s, z = 1]$$

$$(14b) \quad P[y(0) = 1|s, z = 0] \leq P[y(1) = 1|s, z = 0].$$

By the law of iterated expectations, it follows that

$$(15a) \quad P[y(1) = 1|z = 1] \leq P[y(0) = 1|z = 1]$$

$$(15b) \quad P[y(0) = 1|z = 0] \leq P[y(1) = 1|z = 0].$$

Let us inspect inequality (15a). The left side is the observed recidivism probability of the offenders who are actually sentenced to residential treatment. The right side is the counterfactual recidivism probability for these offenders if they were to be sentenced to nonresidential treatment. Thus, inequality (15a) has identifying power. Using the data alone, the researcher can say only that $P[y(0) = 1|z = 1]$ lies between 0 and 1. Assuming the outcome optimization model, she can conclude that $P[y(0) = 1|z = 1]$ lies between $P[y(1) = 1|z = 1]$ and 1. Similarly, inequality (15b) shows that the counterfactual probability $P[y(1) = 1|z = 0]$ is at least as large as the observed recidivism probability $P[y(0) = 1|z = 0]$.

Inequalities (15a) and (15b) allow us to tighten the no-assumptions lower bounds on $P[y(1) = 1]$ and $P[y(0) = 1]$ developed in Section 3.3 using the data alone. The new lower bound on $P[y(1) = 1]$, obtained by setting $P[y(1) = 1|z = 0] = P[y(0) = 1|z = 0]$ in equation (3), is

$$P[y(1) = 1 | z = 1] \cdot P(z = 1) + P[y(0) = 1 | z = 0] \cdot P(z = 0).$$

Equation (5) shows that this quantity equals the status quo recidivism probability $P[y(z) = 1]$. The same reasoning shows that the new lower bound on $P[y(0) = 1]$ is $P[y(z) = 1]$ too. Thus we have

Bound on Recidivism Probability Under Mandatory Residential Treatment, Assuming Outcome Optimization[3]

$$(16) \quad P[y(z) = 1] \leq P[y(1) = 1] \leq P[y(1) = 1 | z = 1] \cdot P(z = 1) + P(z = 0).$$

Bound on Recidivism Probability Under Mandatory Nonresidential Treatment, Assuming Outcome Optimization

$$(17) \quad P[y(z) = 1] \leq P[y(0) = 1] \leq P[y(0) = 1 | z = 0] \cdot P(z = 0) + P(z = 1).$$

Bounds on the classical and status quo treatment effects are computed in the same fashion as in the no-assumptions case. The results are

Bound on the Classical Treatment Effect, Assuming Outcome Optimization

$$(18) \quad - P[y(1) = 0 | z = 1] \cdot P(z = 1) \leq CTE \leq P[y(0) = 0 | z = 0] \cdot P(z = 0).$$

---

[3] Earlier, in Manski (1990, 1995a), the bounds (16) and (17) were shown to hold under a deterministic version of the outcome optimization model, wherein the person selecting treatments knows the outcomes $[y(1), y(0)]$ and selects the treatment yielding the better outcome. The deterministic model is a special case of the model examined here, namely the case in which the information s is so extensive that the probabilities $P[y(1) = 1 | s]$ and $P[y(0) = 1 | s]$ take the extreme values zero or one.

Bound on the Status Quo Treatment Effect, Assuming Outcome Optimization

$$(19)\quad 0 \leq STE \leq P[y(0) = 0 | z = 0] \cdot P(z = 0).$$

The new bound on the classical treatment effect is narrower than the no-assumptions bound derived earlier, but the new lower bound is still non-positive and the new upper bound is still non-negative. Thus, combining the data with the outcome optimization assumption still does not suffice to identify the sign of the CTE.

The new lower bound on the status quo treatment effect is zero. This provides an ironic reminder that under the outcome optimization model, judges are already acting to minimize recidivism. Under this model, changing from the status quo to mandatory residential treatment cannot possibly lower the recidivism rate.

4.2. Skimming

The skimming model operationalizes the idea that program administrators may act to maximize the apparent effectiveness of their programs rather than the actual effectiveness. In the simple form discussed here, the model assumes that the population is composed of two types of persons, labelled A and B. Type A persons tend to have worse outcomes than type B persons under both treatments. Administrators assign all type A persons to one treatment and all type B persons to the other.

In educational and training programs, for example, it is sometimes hypothesized that administrators select participants who they expect will be most successful following completion of the program, regardless of how this success is influenced by the program. By "skimming" or "creaming" the applicants with the best prospects, administrators seek to maximize the apparent effectiveness of their programs.

In our juvenile justice application, the skimming model assumes that, conditioning on the information available to the judge, the population of offenders can be decomposed into "higher risk" and "lower risk" individuals. Higher risk individuals (type A) have higher probabilities of recidivism under both residential and nonresidential treatment than do lower risk individuals (type B). The individuals of each type are not necessarily homogeneous -- some offenders of each type may tend to recidivate less under residential treatment than under nonresidential treatment, and others may tend to recidivate more. Nevertheless, the model assumes that judges sentence all type A persons to residential treatment and all type B ones to nonresidential treatment.

Why might judges sentence in accord with the skimming model? One possibility is that judges, like the administrators of educational and training programs, act to maximize the apparent effectiveness of their sentencing decisions rather than the actual effectiveness. If the public blames the judicial system when offenders given "light" sentences commit subsequent crimes, then judges have an incentive to give light sentences only to type B persons, regardless of how sentencing actually affects recidivism. Another possibility is that judges act in accord with the normative view that the objective of the juvenile justice system should be to "punish the bad and forgive the good," where bad and good are interpreted as higher-risk and lower-risk respectively.

To formalize the model, again let s denote all the characteristics of the offender and environment that may be associated with treatment selection. Again assume that the judge knows the probabilities $P[y(1) = 1|s]$ and $P[y(0) = 1|s]$ of recidivism under treatments 1 and 0. Type A and type B offenders are distinguished by their values of these probabilities. In particular, it is assumed that there exists thresholds π_1 and π_0 such that

(20a) Type A offenders: $P[y(1) = 1|s] \geq \pi_1$ and $P[y(0) = 1|s] \geq \pi_0$

(20b) Type B offenders: $P[y(1) = 1|s] < \pi_1$ and $P[y(0) = 1|s] < \pi_0$.

Thus, regardless of treatment choice, type A offenders have higher recidivism probabilities than do type B offenders.

We assume that all members of the population are either type A or type B -- the population does not contain persons who are higher-risk under one treatment but lower-risk under the other. We assume that the thresholds π_1 and π_0 do not vary across judges. We do not, however, require that the two thresholds be equal. It could be, for example, that type A offenders have recidivism probabilities above 0.6 under residential treatment and above 0.8 under nonresidential treatment. Then type B offenders would have recidivism probabilities below 0.6 and 0.8 under the two treatments.

The skimming model assumes that judges assign residential treatment to the higher-risk, type A offenders and nonresidential treatment to the lower-risk, type B offenders. Thus,

(21a) $P[y(1) = 1|s] \geq \pi_1$ and $P[y(0) = 1|s] \geq \pi_0 \Rightarrow z = 1$

(21b) $P[y(1) = 1|s] < \pi_1$ and $P[y(0) = 1|s] < \pi_0 \Rightarrow z = 0$.

In principle, the skimming rule can result in exactly the same or entirely opposite treatment assignments as the outcome optimization rule. To illustrate, let the two thresholds be the same, with $\pi_1 = \pi_0 = \pi$, and consider two special cases of (20). In one case, the recidivism probabilities satisfy

Type A offenders: $P[y(0) = 1|s] > P[y(1) = 1|s] \geq \pi$

Type B offenders: $P[y(0) = 1|s] < P[y(1) = 1|s] < \pi$.

In the other case, they satisfy

Type A offenders: $P[y(1) = 1|s] > P[y(0) = 1|s] \geq \pi$

Type B offenders: $P[y(1) = 1|s] < P[y(0) = 1|s] < \pi$.

In the first case, outcome minimization and skimming yield the same treatment decisions -- type A offenders are sentenced to residential treatment and type B ones to nonresidential treatment. In the second case, the two models imply contrary decisions. Under the outcome optimization rule, all of the type Bs are sentenced to residential treatment and all of the type As to nonresidential treatment. Under the skimming rule, the opposite occurs.

Now consider the situation of a researcher who assumes that sentencing decisions accord with the skimming model, but who does not know the thresholds π_1 and π_0. When the researcher observes that a judge sentences an offender to residential treatment, he can infer that $P[y(1) = 1|s] \geq \pi_1$ and $P[y(0) = 1|s] \geq \pi_0$. Likewise, when he observes that a judge sentences an offender to nonresidential treatment, he can infer that $P[y(1) = 1|s] < \pi_1$ and $P[y(0) = 1|s] < \pi_0$. Summarizing,

(22a) $z = 1 \Rightarrow P[y(1) = 1|s] \geq \pi_1$ and $P[y(0) = 1|s] \geq \pi_0$.

(22b) $z = 0 \Rightarrow P[y(1) = 1|s] < \pi_1$ and $P[y(0) = 1|s] < \pi_0$.

The treatment z is necessarily a function of s, so (22a) and (22b) are equivalent to these inequalities:

(23a) $P[y(1) = 1|s, z = 1] \geq \pi_1$ and $P[y(0) = 1|s, z = 1] \geq \pi_0$.

(23b) $P[y(1) = 1|s, z = 0] < \pi_1$ and $P[y(0) = 1|s, z = 0] < \pi_0$.

By the law of iterated expectations, it follows that

(24a) $P[y(1) = 1|z = 1] \geq \pi_1$ and $P[y(0) = 1|z = 1] \geq \pi_0$.

(24b) $P[y(1) = 1|z = 0] < \pi_1$ and $P[y(0) = 1|z = 0] < \pi_0$.

Finally, it follows from (24) that

(25a) $P[y(1) = 1|z = 1] > P[y(1) = 1|z = 0]$

(25b) $P[y(0) = 1|z = 0] < P[y(0) = 1|z = 1]$.

Let us inspect inequality (25a). The left side is the observed recidivism probability of the offenders who are sentenced to residential treatment. The right side is the counterfactual recidivism probability that offenders sentenced to nonresidential treatment would have if they were sentenced to residential treatment. Thus, (25a) has identifying power. Using the data alone, the researcher can say only that $P[y(1) = 1|z = 0]$ lies between 0 and 1. Assuming the skimming model, he can conclude that $P[y(1) = 1|z = 0]$ lies between 0 and $P[y(1) = 1|z = 1]$. Similarly, inequality (25b) shows that the counterfactual probability $P[y(0) = 1|z = 1]$ lies between the observed recidivism probability $P[y(0) = 1|z = 0]$ and 1.

Inequality (25a) allows us to tighten the no-assumptions upper bound on $P[y(1) = 1]$ developed in Section 3.3 using the data alone. The new upper bound on $P[y(1) = 1]$, obtained by setting $P[y(1) = 1|z = 0] = P[y(1) = 1|z = 1]$ in equation (3), is

$$P[y(1) = 1|z = 1] \cdot P(z = 1) + P[y(1) = 1|z = 1] \cdot P(z = 0),$$

which reduces to $P[y(1) = 1|z = 1]$, the recidivism probability of the offenders receiving residential treatment. The same reasoning applied to (25b) shows that the new lower bound on $P[y(0) = 1]$ is $P[y(0) = 1|z = 0]$, the recidivism probability of the offenders receiving nonresidential treatment. Thus we have

Bound on Recidivism Probability Under Mandatory Residential Treatment, Assuming Skimming

$$(26) \quad P[y(1) = 1|z = 1] \cdot P(z = 1) \leq P[y(1) = 1] \leq P[y(1) = 1|z = 1]$$

Bound on Recidivism Probability Under Mandatory Nonresidential Treatment, Assuming Skimming

$$(27) \quad P[y(0) = 1|z = 0] \leq P[y(0) = 1] \leq P[y(0) = 1|z = 0] \cdot P(z = 0) + P(z = 1).$$

These results have simple interpretations. If judges are skimming, then the type A offenders are being sentenced to residential treatment and the type Bs, who have lower recidivism rates, are not. Hence the observed recidivism rate among the offenders who are sentenced to residential treatment forms an upper bound on the recidivism rate that would occur if residential treatment were made mandatory for offenders of both types. Analogous reasoning shows that the observed recidivism rate among the type B offenders sentenced to nonresidential treatment forms a lower bound on the recidivism rate that would occur if nonresidential treatment were made mandatory for offenders of both types.

Bounds on the classical and status quo treatment effects are computed in the same fashion as in the no-assumptions case. The results are

Bound on the Classical Treatment Effect, Assuming Skimming

$$(28) \quad P[y(1) = 1|z = 1] \cdot P(z = 1) - \{P[y(0) = 1|z = 0] \cdot P(z = 0) + P(z = 1)\}$$

$\leq \text{CTE} \leq P[y(1) = 1 | z = 1] - P[y(0) = 1 | z = 0]$

Bound on the Status Quo Treatment Effect, Assuming Skimming

$$(29) \quad - P[y(0) = 1 | z = 0] \cdot P(z = 0) \leq \text{STE}$$
$$\leq \{P[y(1) = 1 | z = 1] - P[y(0) = 1 | z = 0]\} \cdot P(z = 0).$$

The new lower bounds on both treatment effects are the same as their respective no-assumptions lower bounds, and so are necessarily non-positive. The new upper bounds are both less than their respective no-assumptions upper bounds and can, in principle, be either negative or positive. If the quantity $P[y(1) = 1 | z = 1] - P[y(0) = 1 | z = 0]$ appearing in both upper bounds is negative, then the skimming model implies that the classical and the status quo treatment effects are both negative.[4] If this quantity is positive, then the signs of these treatment effects are not identified.

In Section 3.3, we showed that $P[y(1) = 1 | z = 1] - P[y(0) = 1 | z = 0]$ is the classical treatment effect under the assumption that treatment selection is random. Here we find that this quantity has an alternative interpretation as the upper bound on the CTE under the skimming model.

4.3. Application

Table 3 reports bounds on the CTE under the two alternative models of treatment selection. Inspection of the table reveals that these two models imply quite different conclusions. Assuming that judges select treatments to minimize the probability of recidivism, the results point strongly, although not

[4] If $P[y(1) = 1 | z = 1] - P[y(0) = 1 | z = 0]$ is negative, the researcher can also conclude that the two thresholds π_1 and π_0 differ, with $\pi_1 < \pi_0$. This follows from (24a) and (24b).

definitively, to the conclusion that the CTE is positive -- that is, mandatory residential treatment increases recidivism compared to mandatory nonresidential treatment. Consider, for example, the subpopulation of cases where the offender has one prior. Here we find that $-0.03 \le CTE \le 0.26$. Thus, mandatory residential treatment can at most be marginally beneficial (the lower bound of −0.03) but might be substantially harmful (the upper bound of 0.26).

Table 3: Bound on Classical Treatment Effect Under Alternative Models of Treatment Selection, by Priors

Priors	Treatment Selection Model					
	No Assumptions		Outcome Opt.		Skimming	
	L.B.	U.B.	L.B.	U.B.	L.B.	U.B.
0	-0.48	0.52	-0.02	0.50	-0.48	0.09
1	-0.65	0.35	-0.03	0.26	-0.65	0.07
2+	-0.65	0.35	-0.04	0.13	-0.65	0.02

Assuming that judges use skimming rules to select treatments, the evidence points more to the efficacy of residential treatment. Consider again the subpopulation of cases where the offender has one prior. Now we find that $-0.65 \le CTE \le 0.07$. Thus, mandatory residential treatment might be substantially beneficial (the lower bound of -0.65) but can be no more than moderately harmful (the upper bound of 0.07).

Table 4 reports bounds on the status quo treatment effect, which measures the change in recidivism probability that would occur if mandatory residential treatment were to replace the existing treatment rules used by judges. The STE may be of greater interest to policy makers than the CTE. We focus on the

Table 4: Bound on Status Quo Treatment Effect Under Alternative Models of Treatment Selection, Offenders with 2 or More Priors

Treatment Selection Model					
No Assumptions		Outcome Opt.		Skimming	
L.B.	U.B.	L.B.	U.B.	L.B.	U.B.
-0.61	0.13	0	0.13	-0.61	0.02

supopulation of cases in which the offender has two or more priors, because these offenders are the most likely target of a policy of mandatory residential treatment. The results offer a potent reminder of the sensitivity of conclusions to the assumed model of treatment assignment. Under the skimming model the results imply that mandatory residential treatment at most marginally increases recidivism (the upper bound of 0.02) and may greatly decrease recidivism (the lower bound of -0.61). Under the outcome optimization model, we reach the opposite conclusion. Here the best possible outcome of mandatory residential treatment is no increase in recidivism. This conclusion follows directly from the model's defining supposition that juvenile court judges are already assigning the recidivism-minimizing treatment.

5. Exclusion Restrictions

5.1. Theory

The analysis of Sections 3 and 4 may be applied to any identifiable subpopulation of offenders. Our notation has not explicitly denoted the subpopulation of interest, but this is easily done. A subpopulation is composed of those offenders with specified covariates x. The subpopulation is identifiable if x is observed by the researcher.

With this notation, we can explicitly make our analysis subpopulation-specific by everywhere conditioning on x. Thus, $P[y(1) = 1|x]$ and $P[y(0) = 1|x]$ are the recidivism probabilities in the subpopulation with covariates x under mandatory residential and nonresidential treatment respectively, $P[y(z) = 1|x]$ is the status quo recidivism probability, and $P(z = 1|x)$ is the probability of receiving treatment 1 in this subpopulation. In the outcome optimization and the skimming models, we need to assume that the information s available to judges includes knowledge of the covariates x. In the skimming model, the thresholds used to define type A and B offenders may now vary with x.

In principle, each identifiable subpopulation may have a distinct treatment selection and outcome process, implying distinct values of the treatment effects. An exclusion restriction is an assumption positing that the recidivism probabilities under mandatory treatments do not vary with x. Formally, let $x = 1,\ldots,K$ denote K different subpopulations comprising a larger population of interest. Then x is "excluded" from the determination of recidivism under mandatory treatments if

(30a) $P[y(1) = 1|x] = P[y(1) = 1], \quad \text{all } x = 1,\ldots,K$

(30b) $P[y(0) = 1|x] = P[y(0) = 1], \quad \text{all } x = 1,\ldots,K.$

An exclusion restriction does not state that the subpopulations are identical in all respects. In particular, the treatment selection process may vary across subpopulations. It may be that skimming takes place in each subpopulation but that the thresholds π_{1x} and π_{0x} vary with x. It may be that skimming takes place in some subpopulations, outcome optimization in others, and yet other treatment rules are applied elsewhere. It may be that the same treatment rule is applied in each subpopulation but that the information s available to judges differs across subpopulations. For all of these reasons and others, the treatment selection probability $P(z = 1|x)$ and the status quo recidivism probability $P[y(z) = 1|x]$ may vary with x.

An exclusion restriction can be imposed alone or can be layered on top of other assumptions (e.g., outcome optimization or skimming) that restrict the values of the recidivism probabilities in each subpopulation separately. Whatever other assumptions may be imposed, an exclusion restriction is used in the same way, which we now describe.

Consider the problem of inference on $P[y(1) = 1]$; the case of $P[y(0) = 1]$ is analogous. We first ignore the exclusion restriction and determine what can be learned about $P[y(1) = 1|x]$ in each subpopulation x using the other assumptions imposed. If no other assumptions are imposed, this means computing the no-assumptions bound in each subpopulation. If outcome optimization or skimming is assumed in a given subpopulation, we compute the bound given in Section 4.1 or 4.2, as appropriate.

Let B_{1x} denote the bound on $P[y(1) = 1|x]$ thus obtained in subpopulation x. The exclusion restriction implies, by (30a), that the population-wide recidivism probability $P[y(1) = 1]$ must lie within all of the bounds B_{1x}, $x = 1,\ldots,K$. That is,

(31) $P[y(1) = 1] \in B_{1x}$, all $x = 1,\ldots,K$.

Equation (31) expresses the identifying power of an exclusion restriction. Whereas the analysis of Sections 3 and 4 placed $P[y(1) = 1]$ within one population-wide bound, imposition of an exclusion restriction places $P[y(1) = 1]$ within the intersection of K subpopulation-specific bounds.

5.2. Application

The bounds obtained under the outcome optimization model and the skimming model are much narrower than the no-assumptions bounds, but do not sign the CTE. We next examine whether the addition of an exclusion restriction narrows the bounds sufficiently to identify the sign of the effect.

The Utah juvenile justice system is divided into eight districts composed of geographically contiguous counties. Four districts are relatively small and so we aggregate them into a single "mega-district." Table 5 reports the recidivism rate of those sentenced to residential treatment, $P[y(1) = 1|x, z = 1]$, and the rate of those sentenced to nonresidential treatment, $P[y(0) = 1|x, z = 0]$, by district and statewide for offenders with 2 or more priors. Observe that the rates vary across district. Assuming that equations (30a) and (30b) hold, the cross-district variation is not attributable to differences in response to treatment across districts (i.e., district is "excluded" from the determination of the response to treatment) but rather reflects cross-district variation in treatment selection.

Table 5: Treatment and Recidivism Rates for Offenders with 2 or More Priors, by District

District	P(z = 1)	P[y(1) = 1|z = 1]	P[y(0) = 1|z = 0]
1	0.30	0.75	0.79
2	0.27	0.85	0.84
3	0.26	0.86	0.82
4	0.26	0.88	0.86
Statewide	0.27	0.85	0.83

To illustrate how such variation can be used to narrow the bounds on the CTE, suppose that treatment selection in all districts is based on the skimming rule but the thresholds for assignment to residential treatment vary by district. Recall that the upper bound on the CTE under the skimming model is the treatment effect under the assumption of random treatment selection. For the state as a whole, this is 0.02 (= 0.85 − 0.83). To use inter-district variation in recidivism to maximally "push-down" this upper bound, we find the district with the smallest recidivism rate among those sentenced to residential treatment, which is District 1 with a 0.75 recidivism rate, and the district with the largest recidivism rate among those sentenced to nonresidential treatment, which is District 4 with a 0.86 rate. The difference between these two rates, namely -0.11, becomes the new upper bound on the CTE.

Table 6 reports bounds on the CTE using the district exclusion restriction. Panel A reports the estimated bounds and Panel B reports companion 90 percent confidence intervals. The results are striking. Under the outcome optimization model, the estimated bounds are strictly positive for all levels of priors, which implies that residential treatment exacerbates recidivism. Under the skimming model, the bounds are strictly negative for offenders with one or more priors, which implies that residential treatment has an ameliorative effect. Finally, if no assumptions are made about the treatment selection process, imposition of the district exclusion does little to narrow the no-assumptions bounds reported in Section 3.

Table 6: Bound on Classical Treatment Effects Under Alternative Models of Treatment Selection, by Priors, Using District as Exclusion Restriction

A. Estimated Bounds

Priors	Treatment Selection Model					
	No Assumptions		Outcome Opt.		Skimming	
	L.B.	U.B.	L.B.	U.B.	L.B.	U.B.
0	-0.43	0.48	0.06	0.46	-0.43	0.03
1	-0.59	0.27	0.07	0.19	-0.59	-0.08
2+	-0.62	0.28	0.02	0.06	-0.62	-0.11

B: 90 Percent Confidence Intervals

Priors	Treatment Selection Model					
	No Assumptions		Outcome Opt.		Skimming	
	L.B.	U.B.	L.B.	U.B.	L.B.	U.B.
0	-0.45	0.50	0.04	0.47	-0.45	0.05
1	-0.62	0.31	0.03	0.20	-0.62	-0.00
2+	-0.64	0.32	-0.01	0.07	-0.64	-0.04

6. Conclusion

This paper has sought to achieve two related objectives. One is to demonstrate the value of nonparametric analysis -- which avoids unmotivated distributional and functional form assumptions -- in focusing attention on the substantive issues that commonly cloud the interpretation of empirical findings in the social sciences. The other is to demonstrate the usefulness of a research approach that differs from the conventional practice of the social sciences. It is conventional to make assumptions strong enough to yield point identification and then argue the merits of those assumptions relative to alternatives. We seek instead to bound disagreements among researchers by examining what can be learned about treatment effects under alternative models of treatment selection, and by layering progressively stronger assumptions about which there is likely to be less consensus.

In this spirit, what conclusions can be drawn from our empirical findings? The answer plainly depends on the assumptions one is willing to make. If no assumptions are made about the treatment selection process, the results mostly provide guidance in the negative -- the bounds on the treatment effects do not identify their signs and place relatively weak restrictions on their magnitudes. However, for those who have a favored model of treatment selection, the results are far more informative. If one's favored model of judicial sentencing behavior is outcome optimization, the results point to the conclusion that residential treatment exacerbates recidivism. If one favors the skimming model, the results provide little support for this conclusion -- instead they suggest an ameliorative effect. Imposition of the district exclusion restriction strengthens each of these opposing conclusions.

The opposing conclusions arising from the two models of treatment selection illustrates the value for informed scientific discourse of identifying zones of agreement and disagreement among researchers, as well as the reasons for such agreements and disagreements. In this application, unless researchers can agree on the treatment selection process, analysis can do little to settle the question of the sign of treatment effects. However, for those who can agree on the selection process, empirical analysis can be extremely informative.

The results also illustrate the value of layering assumptions based on consensus about their validity. We suspect that many observers of the juvenile justice system would concur that the skimming model is a highly plausible depiction of the actual treatment selection process. Concurrence on this model alone, however, does not provide a firm basis for signing treatment effects. Only with the additional imposition of the district exclusion restriction is it possible, in this application, to sign treatment effects with full confidence.

References

Andrews, D., Zinger, I., Hoge, R., Bonta, J., Gendreau, P., and Cullen, F. (1990), "Does Correctional Treatment Work? A Clinically Relevant and Psychologically Informed Meta-analysis," Criminology, 28, 369-404.

Becker, H. (1963), Outsiders, New York: Free Press.

Bernard, T. (1992), Cycles of Juvenile Justice, New York: Oxford University Press.

Cullen, F., Cullen, J., and Woznick, J. (1988), "Is Rehabilitation Dead? The Myth of the Punitive Public," Journal of Criminal Justice, 16, 303-317.

Garfinkel, I., Manski, C., and Michalopolous, C. (1992), "Micro-Experiments and Macro Effects," in C. Manski and I. Garfinkel (editors), Evaluating Welfare and Training Programs, Cambridge, Mass.: Harvard University Press.

Goldberger, A. (1983), "Abnormal Selection Bias," In S. Karlin, T. Amemiya, and L. Goodman (editors), Studies in Econometrics, Time Series, and Multivariate Statistics. New York: Academic Press.

Heckman, J. (1976), "The Common Structure of Statistical Models of Truncation, Sample Selection, and Limited Dependent Variables and a Simple Estimator for Such Models." Annals of Economic and Social Measurement, 5, 479-492.

Heckman, J. (1978), "Dummy Endogenous Variables in a Simultaneous Equation System." Econometrica, 46, 931-959.

LaLonde, R. (1986), "Evaluating the Econometric Evaluations of Training Programs with Experimental Data," American Economic Review, 76, 604-620.

Lemert, E. (1951), Social Pathology, New York: McGraw-Hill.

Maddala, G. S. (1983), Qualitative and Limited Dependent Variable Models in Econometrics, Cambridge, England: Cambridge University Press.

Manski, C. (1989), "Anatomy of the Selection Problem," Journal of Human Resources, 24, 343-360.

Manski, C. (1990), "Nonparametric Bounds on Treatment Effects," American Economic Review Papers and Proceedings, 80, 319-323.

Manski, C. (1993), "Identification Problems in the Social Sciences," Sociological Methodology, 23, 1-56.

Manski, C. (1995), Identification Problems in the Social Sciences, Cambridge, Mass.: Harvard University Press.

Manski, C., Sandefur, G., McLanahan, S. and Power, D. (1992), "Alternative Estimates of the Effect of Family Structure During Adolescence on High School Graduation," Journal of the American Statistical Association, 87, 25-37.

National Juvenile Court Data Archive (1992), "Users Guide to Utah Juvenile Court

Case Records, 1962-1972 Birth Cohort," National Center for Juvenile Justice, Pittsburgh, PA.

Rosenbaum, P., and Rubin, D. (1983), "The Central Role of the Propensity Score in Observational Studies for Causal Effects," Biometrika, 70, 41-55.

Rothman, D. (1980), Conscience and Convenience: The Asylum and Its Alternative in Progressive America, Boston: Little, Brown.

Sechrist, L., White, S., and Brown, E. (1979), The Rehabilitation of Criminal Offenders: Problems and Prospects, Washington, D.C.: National Academy Press.

Smith, D. and Paternoster, R. (1990), "Formal Processing and Future Delinquency: Deviance Amplification as Selection Artifact," Law and Society Review, 24, 1109-1131.

Whitehead, J. and Lab. S. (1989), "A Meta-analysis of Juvenile Correctional Treatment," Journal of Research in Crime and Delinquency, 26, 276-295.

Winship, C. and Mare, R. (1992), "Models for Sample Selection Bias," Annual Review of Sociology 18, 327-350.

Center for Demography and Ecology
University of Wisconsin
1180 Observatory Drive Rm. 4412
Madison, WI 53706-1393
U.S.A.
608/262-2182
FAX 608/262-8400
email: name@ssc.wisc.edu