DETERRENCE AND THE DEATH PENALTY: PARTIAL IDENTIFICATION ANALYSIS USING REPEATED CROSS SECTIONS

Charles F. Manski
Department of Economics and Institute for Policy Research,
Northwestern University

and

John V. Pepper
Department of Economics
University of Virginia

Ehrlich, *AER*, 1975, p. 398:

"In fact, the empirical analysis suggests that on the average the tradeoff between the execution of an offender and the lives of potential victims it might have saved was of the order of 1 for 8 for the period 1933–1967 in the United States."

National Research Council (Blumstein, Cohen, and Nagin, 1978, p. 62):

"The current evidence on the deterrent effect of capital punishment is inadequate for drawing any substantive conclusion."

Researchers have long used data on homicide rates to examine the deterrent effect of capital punishment.

A large body of work addresses the question, yet the literature has failed to achieve consensus.

Numerous shortcomings of early research were documented in the 1978 report of the National Research Council Panel on Research on Deterrent and Incapacitative Effects.

These problems persist in more recent work.

Donohue and Wolfers (2005) find that a seemingly trivial change to the model estimated by Dezhbakhsk, Rubin and Shepard (2003) "flips the sign of the original estimates: instead of saving eighteen lives, each execution leads to eighteen lives lost."

A fundamental difficulty is that the outcomes of counterfactual policies are unobservable.

Data alone cannot reveal what the homicide rate in a state without (with) a death penalty would have been had the state (not) adopted a death penalty statute.

Data must be combined with assumptions about treatment selection and/or treatment response.

It is tempting to impose assumptions strong enough to yield a definitive finding. Then a deterrent effect is *point-identified*.

However, the assumptions may be inaccurate, yielding flawed and conflicting conclusions.

We study inference under weaker assumptions. Such assumptions typically do not point-identify deterrent effects, but they may partially identify them, yielding bounds rather than point estimates.

To demonstrate, we consider the problem of drawing inferences on the deterrent effects of a death penalty statute using data from repeated cross sections of states.

We focus on the years following the 1972 Supreme Court case *Furman vs. Georgia*, which resulted in a moratorium on the application of the death penalty, and the 1976 case *Gregg vs. Georgia*, which ruled that the death penalty could be applied subject to certain criteria.

We examine the effect of death penalty statutes on murder rates in 1975 and 1977. In 1975 the death penalty was illegal throughout the country. In 1977 thirty-two states had death penalty statutes.

Table 1: Homicide Rates per 100,000 Residents by Year and Treatment Status in 1977

	Group		Total
Year	Untreated	Treated	
1975	8	10.3	9.6
1977	6.9	9.7	8.8
Total	7.5	10	9.2

Comparing 1975 and 1977, the treated states are those that legalized the death penalty after the *Gregg* decision and the untreated ones are those that did not.

Assuming random assignment in 1977 gives 2.8
(9.7 − 6.9).

A before-and-after comparison of the treated gives −0.6
(9.7 − 10.3).

The *difference-in-difference* (DID) estimate is 0.5
[(9.7 − 10.3) − (6.9 − 8.0)].

Average Treatment Effects and the Selection Problem

We consider inference on the average treatment effect of a death penalty statute on the homicide rate:

$$ATE_{dX} = E[Y_d(1)|X] - E[Y_d(0)| X].$$

t = 1 denotes a sanctions regime that includes the presence of a death penalty statute. t = 0 denotes one without such a statute.

$Y_d(1)$ and $Y_d(0)$ are the homicide rates if a state were or were not to have a death penalty statute.

X are observed covariates.

d indicates 1975 or 1977 (= 0 if 1975, = 1 if 1977).

$Z_{jd} = 1$ if state j has a death penalty statute in year d.
$Z_{jd} = 0$ otherwise.

The observed murder rate is $Y_{jd} = Y_{jd}(1)Z_{jd} + Y_{jd}(0)(1 - Z_{jd})$.

The Law of Iterated Expectations gives

$$E[Y_d(1)] = E[Y_d(1)|Z_d = 1]P(Z_d = 1) + E[Y_d(1)|Z_d = 0]P(Z_d = 0).$$

The data in Table 1 reveal that

$P(Z_1 = 1) = 0.70$. $P(Z_1 = 0) = 0.30$. $E[Y_1(1)|Z_1 = 1] = 9.7$.

The data do not reveal $E[Y_d(1)|Z_d = 0]$.

Linear Homogeneous Treatment Response

Assume that

$$Y_{jd}(t) = \alpha_j + \beta \cdot d + \gamma \cdot t + \delta_{jd}.$$

Evaluated at realized values of treatments and outcomes,

$$Y_{jd} = \alpha_j + \beta \cdot d + \gamma \cdot Z_{jd} + \delta_{jd}.$$

Let $X_j = 1$ if state j is treated ($Z_{j1} = 1$), $X_j = 0$ if untreated ($Z_{j1} = 0$).

Assume that $E(\delta_d | X, Z_d) = 0$ and $E(\alpha | X, Z_d) = E(\alpha | X)$. Then

$$E(Y_d | X, Z_d) = E(\alpha | X) + \beta \cdot d + \gamma \cdot Z_d.$$

It follows that γ has the DID form

$$\gamma = [E(Y_1 | X = 1, Z_1 = 1) - E(Y_0 | X = 1, Z_0 = 0)] - [E(Y_1 | X = 0, Z_1 = 0) - E(Y_0 | X = 0, Z_0 = 0)].$$

Partial Identification Assuming Bounded Outcomes

Recall that

$$E[Y_d(1)] = E[Y_d(1)|Z_d = 1]\cdot P(Z_d = 1) + E[Y_d(1)|Z_d = 0]\cdot P(Z_d = 0).$$

The homicide rate per 100,000 residents cannot be larger than 100,000. Thus, $E[Y_1(1)|\, Z_1 = 0] \in [0, 100{,}000]$.

To obtain a more reasonable bound, note that across all states and both years, the observed homicide rate always was in the range [0.8, 32.8]. Assume that $E[Y_1(1)|\, Z_1 = 0] \in [0, 35]$. Then

$$E[Y_d(1)] \in \{E[Y_d(1)|Z_d = 1]\cdot P(Z_d = 1) + 0\cdot P(Z_d = 0),$$
$$E[Y_d(1)|Z_d = 1]\cdot P(Z_d = 1) + 35\cdot P(Z_d = 0)\}.$$

Sharp bounds on the ATE can be found by taking the difference between the lower (upper) bound on $E[Y_d(1)]$ and the upper (lower) bound on $E[Y_d(0)]$.

Table 2: Partial Identification of the ATE Under the Bounded Outcomes Assumption

	1975	1977
Probability of Death Penalty Statute: $P(Z_d = 1)$	0	0.7
Murder Rate with Statute: $E[Y_d(1)\|Z_d = 1]$	N. A.	9.7
Murder Rate without Statute: $E[Y_d(0)\|Z_d = 0]$	9.6	6.9
Bounds:		
$E[Y_d(1)]$	[0, 35]	[6.8 , 17.3]
$E[Y_d(0)]$	9.6	[2.1 , 26.6]
ATE_d	[-9.6, 25.4]	[-19.8 ,15.2]

Middle-Ground Assumptions

We consider "middle-ground" assumptions that presume some commonality of deterrent effects across states or time, but not homogeneity.

We do not endorse any particular assumption. We demonstrate how the conclusions drawn depends on the assumptions imposed.

This provides a spectrum of possibilities to readers of research on deterrence.

Date-Invariant Treatment Effects

Assume that $ATE_1 = ATE_0$.

Then the date-invariant ATE must lie in the intersection of the two date-specific intervals shown in Table 1, these being [-9.6, 25.4] and [-19.8, 15.2]. The result is [−9.6, 15.2].

Here is a longer derivation that will have payoff later.

Assume that mean treatment response at date d has the form

$$E[Y_1(t)] = E[Y_0(t)] + \beta.$$

Then $E[Y_1(1)] - E[Y_1(0)] = E[Y_0(1)] - E[Y_0(0)]$.

Let $E_t \equiv E[Y_0(t)]$. Then $E[Y_1(t)] = E_t + \beta$.

These equations reduce the number of unknown mean potential outcomes by one. Without the assumption, we do not know the four quantities $E[Y_d(t)]$, $d = 0, 1$; $t = 0, 1$. With the assumption, we do not know the three quantities (E_0, E_1, β).

To obtain the identifying power of the assumption, first consider each pair (d, t) and obtain the identification region for $E[Y_d(t)]$ using only the assumption of bounded outcomes. Let this interval be called $[L_d(t), U_d(t)]$. Combining this with date invariance yields

$$L_0(t) \le E_t \le U_0(t), \qquad t = 0, 1;$$

$$L_1(t) \le E_t + \beta \le U_1(t), \qquad t = 0, 1.$$

The feasible values of the three unknowns (β, E_0, E_1) are the triples that satisfy the four inequalities.

We find that $\beta \in [-7.5, 17.0]$, $E_o = 9.6$, $E_1 \in [0, 24.8]$.

Hence, ATE $\in [-9.6, 15.2]$.

Bounding Time-Series Variation in Mean Response Levels

The above analysis did not restrict time-series variation in treatment response levels. Some states could have become much more prone to homicide over time. Others could have become much less prone to homicide.

The assumption and data implied that $\beta \in [-7.5, 17.0]$.

One might not think it credible that large variations in potential homicide rates could have occurred over a short time period. Assuming that β lies in a narrower interval may imply a narrower bound on the ATE.

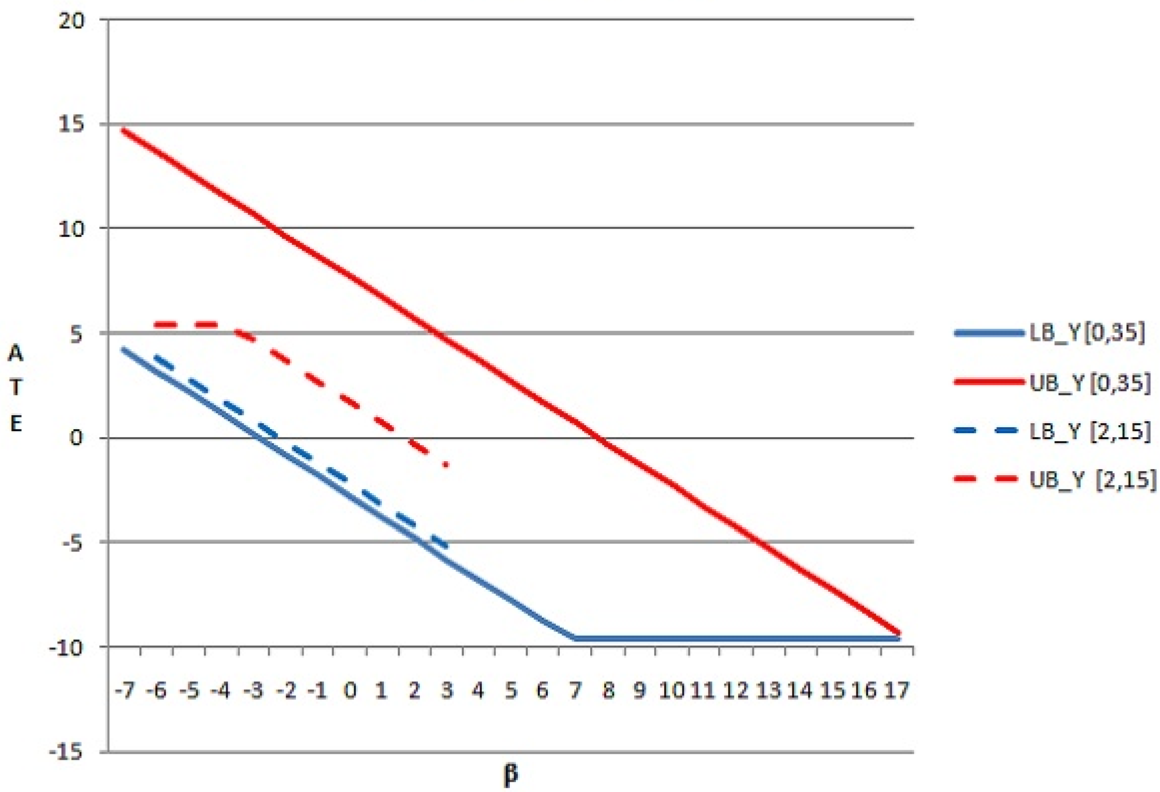
Figure 1 displays the bound on the ATE as a function of β.

$\beta \leq -3 \;\Rightarrow\; \text{ATE} > 0.$

$\beta = 0 \;\Rightarrow\; \text{ATE} \in [-2.8, 7.7]$

$\beta \geq 8 \;\Rightarrow\; \text{ATE} < 0.$

Figure 1: Bounds on the ATE as a Function of β



Tighter Bounds on Counterfactual Mean Response Levels

We have thus far placed only a weak bound on the counterfactual homicide rates $E[Y_d(t) \mid Z_d \neq t]$, supposing that they lie in [0, 35].

Of the 110 state-specific homicide rates in 1975 and 1977, the central ninety percent fall in the interval [2, 15]. Suppose that one uses this interval as a bound on $E[Y_d(t) \mid Z_d \neq t]$. Then

$$\in [-6.1, 3.0], \quad \text{ATE} \in [-5.2, 5.4].$$

Figure 1 displays how the bound on the ATE varies with .

$\beta \leq -2 \Rightarrow \text{ATE} > 0.$

$\beta = 0 \Rightarrow \text{ATE} \in [-2.2, 1.7]$

$\beta \geq 2 \Rightarrow \text{ATE} < 0.$

Date-Invariant Treatment Effects Combined with Covariate-Invariant Date Intercepts

Let X separate states into K groups.

Let $E_{t|X} \equiv E[Y_0(t)|X]$. Assume that $E[Y_1(t)|X] = E_{t|X} + \beta$.

Let $[L_d(t|x), U_d(t|x)]$ be the bound on $E[Y_d(t)|X]$ obtained using only the assumption that outcomes lie in [0, 35]. Then

$$L_0(t|X) \le E_{t|X} \le U_0(t|X), \qquad t = 0, 1; \text{ all } X$$
$$L_1(t|X) \le E_{t|X} + \beta \le U_1(t|X), \qquad t = 0, 1; \text{ all } X.$$

The feasible values of the (2K + 1) unknowns (β, $E_{0|X}$, $E_{1|X}$, all X) satisfy these 4K inequalities.

To illustrate, we evaluate the ATE with two definitions of X.

First, X indicates whether a state does or does not have a death penalty statute in 1977; that is, whether it is treated or untreated.

Second, we let X indicate the location of a state in one of four mutually exclusive and exhaustive census regions.

Treatment Group as the Covariate

Let $X_j = 1$ if $Z_{j1} = 1$ and $X_j = 0$ if $Z_{j1} = 0$. The former are the treated states and the latter are the untreated ones. With this definition of X, the eight inequalities become

$$E_{0|0} = E(Y_0|X = 0), \qquad E_{0|0} + \beta = E(Y_1|X = 0),$$
$$0 \le E_{1|0} \le 35, \qquad 0 \le E_{1|0} + \beta \le 35,$$
$$E_{0|1} = E(Y_0|X = 1), \qquad 0 \le E_{0|1} + \beta \le 35,$$
$$0 \le E_{1|1} \le 35, \qquad E_{1|1} + \beta = E(Y_1|X = 1).$$

The equalities in the first row point-identify the date intercept

$$\beta = E(Y_1|X = 0) - E(Y_0|X = 0).$$

We find $\beta = -1.1$.

With knowledge of β,

$$E_{0|0} = E(Y_0|X = 0), \qquad E_{0|1} = E(Y_0|X = 1),$$

$$E_{1|0} \in [\max(0, -\beta), \min(35, 35 - \beta)]; \quad E_{1|1} = E(Y_1|X = 1) - \beta.$$

The effect of treatment on the treated is point-identified, with

$$ETT \equiv E_{1|1} - E_{0|1}$$

$$= [E(Y_1|X = 1) - E(Y_0|X = 1)] - [E(Y_1|X = 0) - E(Y_0|X = 0)].$$

The effect of treatment on the untreated is partially identified.

$$ETU \equiv E_{1|0} - E_{0|0}$$

$$\in [\max(0, -\beta) - E(Y_0|X = 0), \ \min(35, 35 - \beta) - E(Y_0|X = 0)].$$

Table 3: Treatment Effects with Date-Invariant Treatment Effects, with and without Covariate-Invariant Date Intercepts			
Assumption	ETT	ETU	ATE
Linear Response	0.5	0.5	0.5
Bounded Outcomes, 1977	[−25.3, 9.7]	[−6.9, 28.1]	[−19.8,15.2]
Bounded Outcomes, 1975	[−35.0, 35.0]	[−9.6, 25.4]	[−9.6, 25.4]
Date-Invariant Treatment Effects			
No Covariate			[−9.6, 15.2]
Region as Covariate			[−9.0, 10.1]
Treatment Group as Covariate	0.3	[−6.9, 27.0]	[−1.9 , 8.3]

Bounded Instrumental Variables

Traditional instrumental variables (IVs) assume that specified groups of treatment units have the same mean treatment response or the same average treatment effects.

It often is difficult to motivate IV assumptions, but it may be easier to motivate weaker ones asserting that mean response or average treatment effects do not differ too much across groups.

Such assumptions uses *bounded instrumental variables*.

To demonstrate, consider identification of the ATE when the researcher selects a $\Delta \geq 0$ and assumes that

$$| ATE_{dx} - ATE_{dx'} | \leq \qquad \text{for all x and x}'.$$

= 0 gives a traditional IV assumption asserting that groups of states with different covariates have the same ATE.

$\Delta > 0$ supposes that the ATE may differ across groups by no more than Δ.

Figure 2 maintains the earlier assumptions, adds the new one, and displays the bound on the ATE as a function of Δ.

Figure 3 sets $\Delta = 0$ and displays the bound as a function of β.

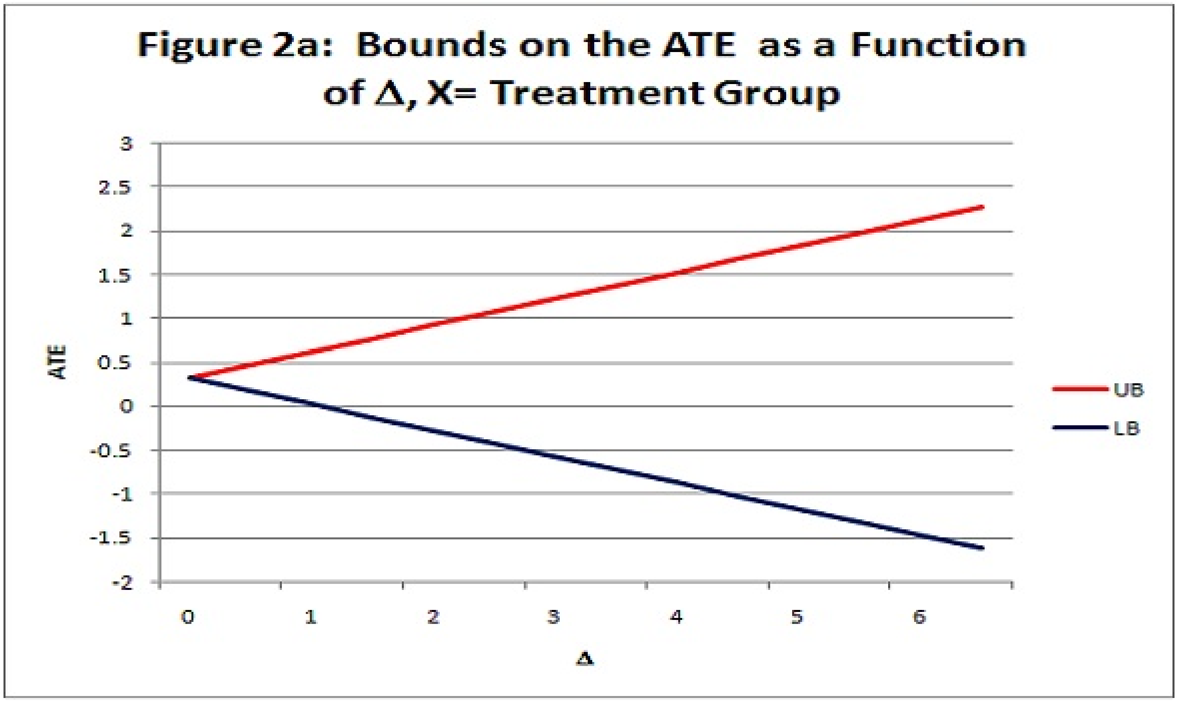


Figure 2a: Bounds on the ATE as a Function of Δ, X= Treatment Group
3
2.5
2
1.5
1
0.5
0
-0.5
-1
-1.5
-2
ATE
0
1
2
3
4
5
6
Δ
UB
LB

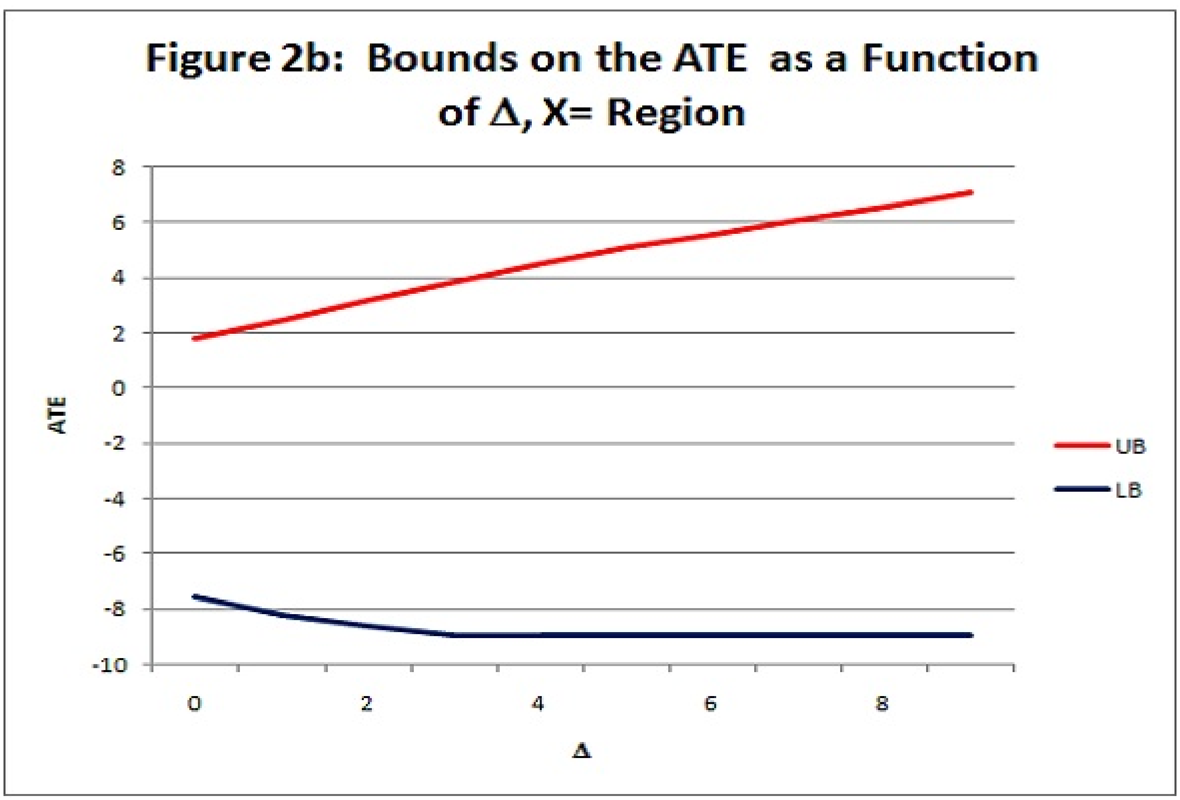
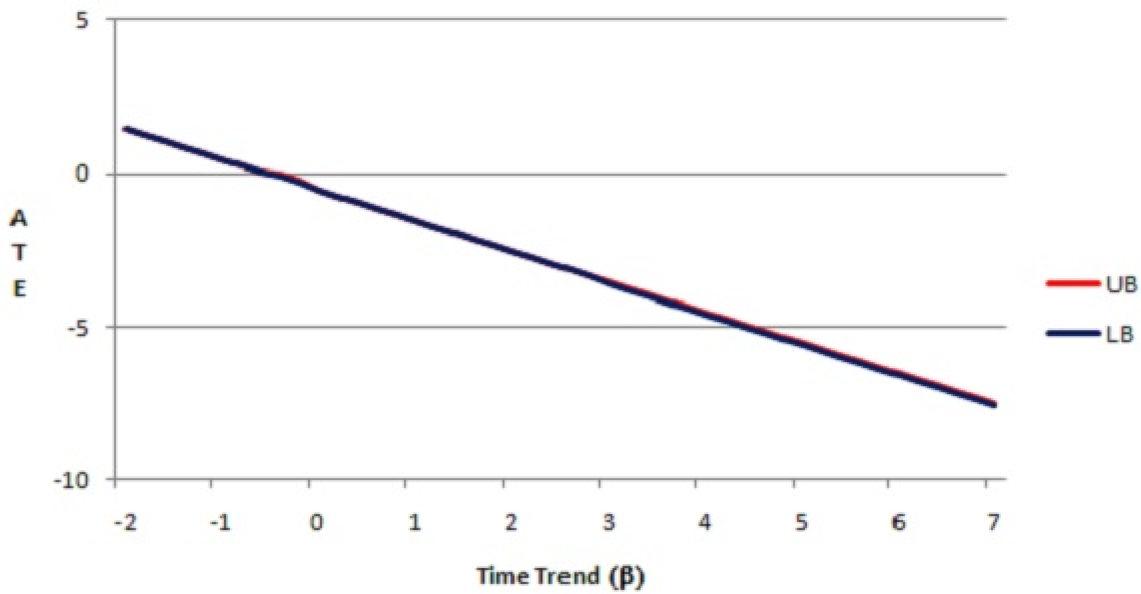


Figure 2b: Bounds on the ATE as a Function of Δ, X= Region
8
6
4
2
0
-2
-4
-6
-8
-10
ATE
0
2
4
6
8
Δ
UB
LB

Figure 3: Bounds on the ATE as a Function of β, Region as an IV



Conclusion

Researchers have had persistent problems providing credible inference on the deterrent effect of the death penalty.

The 1978 NRC report warned of the fundamental shortcomings of the data and methods, and questioned whether empirical research could provide useful information at all.

Nevertheless, researchers have continued to examine the same or more recent data using the same or similar methods.

To yield point identification, research continues to combine data with untenable assumptions.

The results have been highly sensitive to these assumptions and no consensus has emerged. As we see it, the research has failed to provide meaningful answers.

In this paper, we demonstrate what can be learned under relatively weak assumptions.

In particular, we study the identifying power of assumptions restricting variation in treatment response across places and time.

The results are findings of partial identification that bound the deterrent effect of capital punishment.

By successively adding stronger identifying assumptions, the analysis makes transparent how assumptions shape inferences.

If one assumes only that outcomes are bounded, one cannot identify the sign of the ATE and one can only draw weak conclusions about its magnitude.

Those who find it credible to make further assumptions can obtain more informative findings.

Society at large can draw strong conclusions only if there is a consensus favoring particular assumptions.

Without such a consensus, data on sanctions and murder rates cannot settle the debate about deterrence.

However, data combined with weak assumptions can bound and focus the debate.