Stanford Law Review

Uses and Abuses of Empirical Evidence in the Death Penalty Debate
Author(s): John J. Donohue and Justin Wolfers
Source: *Stanford Law Review*, Vol. 58, No. 3 (Dec., 2005), pp. 791-845
Published by: Stanford Law Review
Stable URL: http://www.jstor.org/stable/40040281
Accessed: 11-08-2016 18:21 UTC

Your use of the JSTOR archive indicates your acceptance of the Terms & Conditions of Use, available at http://about.jstor.org/terms

JSTOR is a not-for-profit service that helps scholars, researchers, and students discover, use, and build upon a wide range of content in a trusted digital archive. We use information technology and tools to increase productivity and facilitate new forms of scholarship. For more information about JSTOR, please contact support@jstor.org.

*Stanford Law Review* is collaborating with JSTOR to digitize, preserve and extend access to *Stanford Law Review*

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

# USES AND ABUSES OF EMPIRICAL EVIDENCE IN THE DEATH PENALTY DEBATE

John J. Donohue* and Justin Wolfers**

INTRODUCTION ..... 791
I. THEORY: WHAT ARE THE IMPLICATIONS OF THE DEATH PENALTY FOR HOMICIDE RATES? ..... 795
II. A CENTURY OF MURDERS AND EXECUTIONS ..... 796
III. THE IMPORTANCE OF COMPARISON GROUPS ..... 798
A. *Canada Versus the United States* ..... 798
B. *Non-Death Penalty States Versus Other States in the United States* ... 800
IV. PANEL DATA METHODS ..... 804
A. *Katz, Levitt, and Shustorovich* ..... 811
B. *Dezhbakhsh and Shepherd* ..... 813
C. *Mocan and Gittings* ..... 816
D. *Other Studies* ..... 818
V. INSTRUMENTAL VARIABLES ESTIMATES ..... 821
A. *Problems with Invalid Instruments* ..... 827
B. *Problems with Statistical Significance* ..... 833
VI. A PARTIAL RECONCILIATION: LACK OF STATISTICAL POWER AND REPORTING BIAS ..... 836
CONCLUSION ..... 841

## INTRODUCTION

Over much of the last half-century, the legal and political history of the

---

* Leighton Homer Surbeck Professor of Law, Yale Law School.

** Assistant Professor of Business and Public Policy, The Wharton School, University of Pennsylvania, and CEPR, IZA, and NBER. The authors wish to thank Sascha Becker, Chris Griffin, and Joe Masters for extremely valuable research assistance, and Dale Cloninger, Larry Katz, Naci Mocan, Joanna Shepherd, and Paul Zimmerman for generously sharing their data and code with us. We are grateful to Gerald Faulhaber, Andrew Leigh, David Rosen, Peter Siegelman, Carol Steiker, Betsey Stevenson, Joel Waldfogel, and Matthew White for useful discussions.



This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

death penalty in the United States has closely paralleled the debate within social science about its efficacy as a deterrent. Sociologist Thorsten Sellin's careful comparisons of the evolution of homicide rates in contiguous states from 1920 to 1963 led to doubts about the existence of a deterrent effect caused by the imposition of the death penalty.[1] This work likely contributed to the waning reliance on capital punishment, and executions virtually ceased in the late 1960s. In the 1972 *Furman* decision, the Supreme Court ruled that existing death penalty statutes were unconstitutional.[2] In 1975, Isaac Ehrlich's analysis of national time-series data led him to claim that each execution saved eight lives.[3] Solicitor General Robert Bork cited Ehrlich's work to the Supreme Court a year later, and the Court, while claiming not to have relied on the empirical evidence, ended the death penalty moratorium when it upheld various capital punishment statutes in *Gregg v. Georgia* and related cases.[4] The injection of Ehrlich's conclusions into the legal and public policy arenas, coupled with the academic debate over Ehrlich's methods, led the National Academy of Sciences to issue a 1978 report which argued that the existing evidence in support of a deterrent effect of capital punishment was unpersuasive.[5] Over the next two decades, as a series of academic papers continued to debate the deterrence question, the number of executions gradually increased, albeit to levels much lower than those seen in the first half of the twentieth century.

The current state of the political debate over capital punishment is one of disagreement, controversy, and division. Governor George Ryan of Illinois suspended executions in that state in 2000 and commuted the death sentences of all Illinois death row inmates in 2003.[6] As a number of other jurisdictions were considering similar moratoria, New York's highest court ruled in 2004

---

1. Thorsten Sellin, *Homicides in Retentionist and Abolitionist States*, *in* CAPITAL PUNISHMENT 135 (Thorsten Sellin ed., 1967) [hereinafter Sellin, *Homicides*]; *see also* Thorsten Sellin, *Experiments with Abolition*, *in* CAPITAL PUNISHMENT, *supra*, at 122.

2. Furman v. Georgia, 408 U.S. 238 (1972).

3. Isaac Ehrlich, *The Deterrent Effect of Capital Punishment: A Matter of Life and Death*, 65 AM. ECON. REV. 397 (1975).

4. *See* Gregg v. Georgia, 428 U.S. 153 (1976). In *Gregg*, Justice Stewart stated, "Although some of the studies suggest that the death penalty may not function as a significantly greater deterrent than lesser penalties, there is no convincing empirical evidence either supporting or refuting this view." *Id.* at 185. Yet, he then asserted: "We may nevertheless assume safely that there are murderers, such as those who act in passion, for whom the threat of death has little or no deterrent effect. But for many others, the death penalty undoubtedly is a significant deterrent." *Id.* Justice Stewart did not clarify whether he believed that murders would increase if convicted murderers who might otherwise be executed instead received sentences of life without parole and, if so, on what basis this might be safely assumed.

5. NAT'L ACADEMY OF SCIENCES, DETERRENCE AND INCAPACITATION: ESTIMATING THE EFFECTS OF CRIMINAL SANCTIONS ON CRIME RATES (Alfred Blumstein et al. eds., 1978) [hereinafter DETERRENCE AND INCAPACITATION].

6. John Biemer, *Death Penalty Reforms Lauded*, CHI. TRIB., Nov. 24, 2003, at M1.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

that the state's death penalty statute was unconstitutional.[7] Executions in California are virtually nonexistent, although the state continues to add prisoners to death row at a rapid pace.[8] Meanwhile, executions continue apace in Texas, which accounts for over one-third of all post-*Gregg* executions.[9]

A host of more recent academic studies has examined the death penalty over the last decade, with mixed results. While Lawrence Katz, Steven Levitt, and Ellen Shustorovich found no robust evidence of deterrence,[10] several researchers claim to have uncovered compelling evidence to the contrary.[11] This latter research appears to have found favor with Cass Sunstein and Adrian Vermeule, who describe it as "powerful"[12] and "impressive,"[13] and they refer to "many decades' worth of data about [capital punishment's] deterrent effects."[14] While Sunstein and Vermeule claim not to endorse any specific analysis, these "sophisticated multiple regression studies"[15] are "[t]he foundation for [their] argument,"[16] and they specifically rely on many of the recent studies that we will reexamine as buttressing their premise that "capital

---

7. William Glaberson, *4-3 Ruling Effectively Halts Death Penalty in New York*, N.Y. TIMES, June 25, 2004, at A1.

8. By the end of 2004, California's death row population was the highest in the country (637 inmates). *See* THOMAS P. BONCZAR & TRACY L. SNELL, BUREAU OF JUSTICE STATISTICS, CAPITAL PUNISHMENT, 2004, at 1 (2005).

9. *See id.* at 9.

10. Lawrence Katz, Steven D. Levitt & Ellen Shustorovich, *Prison Conditions, Capital Punishment, and Deterrence*, 5 AM. L. & ECON. REV. 318 (2003).

11. *See* Hashem Dezhbakhsh, Paul H. Rubin & Joanna M. Shepherd, *Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data*, 5 AM. L. & ECON. REV. 344 (2003); H. Naci Mocan & R. Kaj Gittings, *Getting Off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment*, 46 J.L. & ECON. 453, 453 (2003); Paul R. Zimmerman, *State Executions, Deterrence, and the Incidence of Murder*, 7 J. APPLIED ECON. 163, 163 (2004).

Joanna Shepherd, an author of several studies finding a deterrent effect, has recently argued before Congress that recent research has created a "strong consensus among economists that capital punishment deters crime," going so far as to claim that "[t]he studies are unanimous." *Terrorist Penalties Enhancement Act of 2003: Hearing on H.R. 2934 Before the Subcomm. on Crime, Terrorism, and Homeland Security of the H. Comm. on the Judiciary*, 108th Cong. 10-11 (2004), *available at* http://judiciary.house.gov/media/pdfs/printers/108th/93224.pdf. Upon further probing from the committee chairman about "the findings of anti-death penalty advocates that are 180 degrees from your conclusions," *id.* at 24, Shepherd responded:

> There may be people on the other side that rely on older papers and studies that use outdated statistical techniques or older data, but all of the modern economic studies in the past decade have found a deterrent effect. So I am not sure what the other people are relying on.

*Id.*

12. Cass R. Sunstein & Adrian Vermeule, *Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs*, 58 STAN. L. REV. 703, 706 (2005) (in this Issue).

13. *Id.* at 713.

14. *Id.* at 737.

15. *Id.* at 711.

16. *Id.* at 706.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

punishment powerfully deters killings."[17] This empirical evidence leads to the heart of their claim that it would be irresponsible for government to fail to act upon the studies and vigorously prosecute the death penalty. Carol Steiker has offered a considered response to this claim based on moral theory;[18] by contrast, we are interested in exploring its empirical premise.

Thus, our aim in this Article is to provide a thorough assessment of the statistical evidence on this important public policy issue and to understand better the conflicting evidence. We test the sensitivity of existing studies in a number of intuitively plausible ways—testing their robustness to alternative sample periods, comparison groups, control variables, functional forms, and estimators. We find that the existing evidence for deterrence is surprisingly fragile, and even small changes in specifications yield dramatically different results. Our key insight is that the death penalty—at least as it has been implemented in the United States since *Gregg* ended the moratorium on executions—is applied so rarely that the number of homicides it can plausibly have caused or deterred cannot be reliably disentangled from the large year-to-year changes in the homicide rate caused by other factors. Our estimates suggest not just "reasonable doubt" about whether there is any deterrent effect of the death penalty, but profound uncertainty. We are confident that the effects are not large, but we remain unsure even of whether they are positive or negative. The difficulty is not just one of statistical significance: whether one measures positive or negative effects of the death penalty is extremely sensitive to very small changes in econometric specifications. Moreover, we are pessimistic that existing data can resolve this uncertainty.

We begin in the next Part by sketching the relevant economic theories of crime and the difficulties in identifying their effects. We then begin our tour of the statistical evidence. Part II analyzes aggregate time-series evidence. Part III analyzes first differences—the change in homicide rates that occurs following death penalty reforms. In Part IV, we turn to panel data analysis, and Part V analyzes the key instrumental variables estimates. Part VI contains our attempt at reconciling the conflicting evidence, assessing the limited precision with which we might be able to pin down the deterrent effect of the death penalty with existing data. Our organizing theme involves an attempt to examine the evidence compiled by previous scholars with the aim of highlighting the ways in which this evidence can both provide insight but also potentially mislead policy analysts.

---

17. *Id.* at 738.

18. Carol S. Steiker, *No, Capital Punishment Is Not Morally Required: Deterrence, Deontology, and the Death Penalty*, 58 STAN. L. REV. 751 (2005) (in this Issue).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

## I. THEORY: WHAT ARE THE IMPLICATIONS OF THE DEATH PENALTY FOR HOMICIDE RATES?

The theoretical premise underlying the deterrence argument is simple: raise the price of murder for criminals, and you will get less of it. In general, the death penalty raises the price of homicide as long as execution is worse than life imprisonment for most potential murderers.[19]

While this argument is qualitatively reasonable, its quantitative significance may be minor. In 2003, there were 16,503 homicides (including nonnegligent manslaughter), but only 144 inmates were sentenced to death.[20] Moreover, of the 3374 inmates on death row at the beginning of the year, only 65 were executed.[21] Thus, not only did very few homicides lead to a death sentence, but the prospect of execution did not greatly affect the life expectancy of death row inmates. Indeed, Katz, Levitt, and Shustorovich have made this point quite directly, arguing that "the execution rate on death row is only twice the death rate from accidents and violence among all American men" and that the death rate on death row is plausibly lower than the death rate of violent criminals not on death row.[22] As such they conclude that "it is hard to believe that in modern America the fear of execution would be a driving force in a rational criminal's calculus."[23] Moreover, even if there were a deterrent effect, capital punishment is sufficiently expensive[24] that it may potentially divert

---

19. The general rule is subject to a caveat. Once a criminal has already committed enough murders to get the maximum penalty, marginal deterrence is lost by a death penalty regime. At that point, the cost of killing to avoid capture goes to zero, and the death penalty may increase incentives to kill to avoid execution.

20. FED. BUREAU OF INVESTIGATION, CRIME IN THE UNITED STATES 15 (2003), *available at* http://www.fbi.gov/ucr/03cius.htm; *see also* BONCZAR & SNELL, *supra* note 8, at 1.

21. BONCZAR & SNELL, *supra* note 8, at 1.

22. Katz, Levitt & Shustorovich, *supra* note 10, at 318-19.

23. *Id.* at 320. On the other hand, even if criminals are not effective calculators, the vivid character of the death penalty might give criminals pause to a greater degree than its likely risk of implementation alone would warrant. The recent literature suggests two possibilities: (1) many individuals treat events with small likelihoods of occurrence as having zero probability, which would mean that the highly unlikely event of execution would essentially have a zero possibility of deterring instead of just a very small likelihood of deterring; and (2) certain catastrophic events that occur with low frequency are given greater prominence in decisionmaking than their likelihood warrants if individuals are given frequent vivid reminders of these events, which could conceivably make the death penalty more of a deterrent than a rational calculation of the risk such as that offered by Katz, Levitt, and Shustorovich would suggest. *See* ROBERT COOTER & THOMAS ULEN, LAW AND ECONOMICS 351 (4th ed. 2004). Again, only empirical investigation can answer the question of which effect would be more dominant on potential murderers.

24. *Public Policy Choices and Deterrence and the Death Penalty: A Critical Review of New Evidence: Hearing on H. B. 3834 Before the Joint Comm. on Judiciary of Mass. Leg.* (July 14, 2005) (statement of Jeffrey Fagan) [hereinafter Fagan Statement] (citing an array of studies documenting the high cost of capital cases compared to a sentence of life without parole), *available at* http://www.deathpenaltyinfo.org/MassTestimonyFagan.pdf.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

resources away from more effective crime prevention strategies.

A more sociological approach notes that there may be social spillovers as state-sanctioned executions cheapen the value of life, potentially demonstrating that deadly retribution is socially acceptable. Thus, executions may actually stimulate more homicide through the so-called "brutalization effect."[25] With theory inconclusive, we now turn to examining the data.

## II. A CENTURY OF MURDERS AND EXECUTIONS

Several of the early studies of the death penalty were based on analysis of the aggregate U.S. time-series data. Figure 1 depicts the homicide and execution rates for the United States over the last century.[26] Because data issues can be a concern with crime data, we present two series for homicides—one from the Uniform Crime Reports and the other compiled from Vital Statistics sources, based on death certificates.[27]

No clear correlation between homicides and executions emerges from this long time series. In the first decade of the twentieth century, execution and homicide rates seemed roughly uncorrelated, followed by a decade of divergence as executions fell sharply and homicides trended up. Then for the next forty years, execution and homicide rates again tended to move together—first rising together during the 1920s and 1930s, and then falling together in the 1940s and 1950s. As the death penalty fell into disuse in the 1960s, the homicide rate rose sharply. The death penalty moratorium that began with *Furman* in 1972 and ended with *Gregg* in 1976 appears to have been a period in which the homicide rate rose. The homicide rate then remained high and variable through the 1980s while the rate of executions rose. Finally, homicides dropped dramatically during the 1990s. By any measure, the resumption of the death penalty in recent decades has been fairly minor, and both the level of the execution rate and its year-to-year changes are tiny: since 1960 the proportion of homicides resulting in execution ranged from 0% to 3%. By contrast, there was much greater variation in execution rates over the previous sixty years, when the execution rate ranged from 2.5% to 18%. This immediately hints that—even with modern econometric methods—it is unlikely that the last few

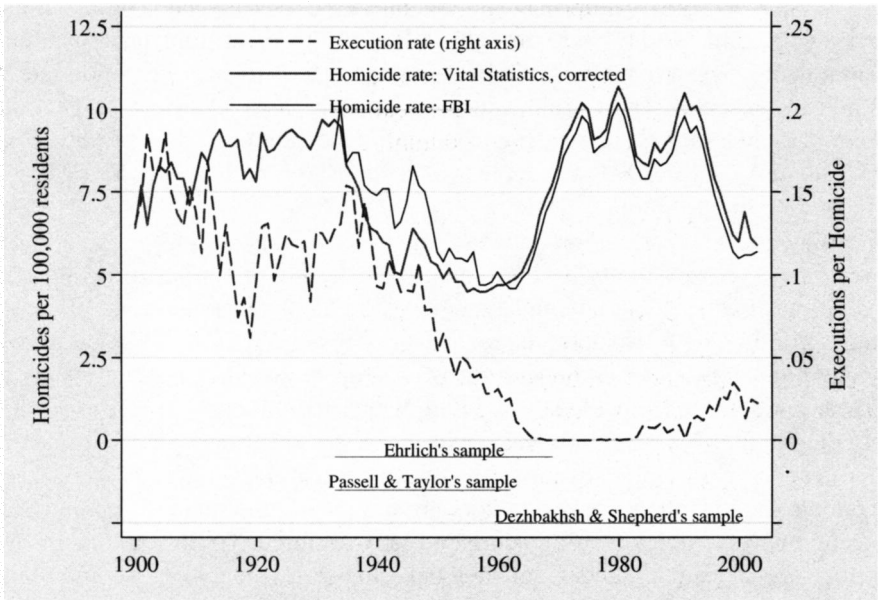
25. William J. Bowers & Glenn L. Pierce, *Deterrence or Brutalization? What Is the Effect of Executions?*, 26 CRIME & DELINQ. 453 (1980); *see also* Steiker, *supra* note 18, at 786-89 (discussing the "brutalization effect" as initially brought up in Sunstein and Vermeule's article (Sunstein & Vermeule, *supra* note 12, at 713 n.37, 745 & n.125)).

26. The execution data come from the Espy file. *See* M. WATT ESPY & JOHN ORTIZ SMYKLA, ICPSR STUDY NO. 8451, EXECUTIONS IN THE UNITED STATES, 1608-2002: THE ESPY FILE (2004) [hereinafter ESPY FILE], *available at* http://webapp.icpsr.umich.edu/cocoon/NACJD-STUDY/08451.xml.

27. Given the incomplete nature of Vital Statistics reporting in the first half of the century, we rely on Douglas Eckberg's estimates of the homicide rate. *See* Douglas Lee Eckberg, *Estimates of Early Twentieth-Century U.S. Homicide Rates: An Econometric Forecasting Approach*, 32 DEMOGRAPHY 1 (1995).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 1. Homicides and Execution in the United States




decades generated enough variation in execution rates to overturn earlier conclusions about the deterrent effect of capital punishment.

This simple chart reconciles many of the conflicting results from the death penalty literature. Ehrlich's provocative 1975 paper argued that he could isolate the movements in the homicide rate caused by changing execution policies, concluding that each execution deterred an average of eight homicides.[28] Passell and Taylor showed that Ehrlich's result relied heavily on movements from 1963 to 1969.[29] When they limited the Ehrlich model to the period from 1935 to 1962, they found no deterrent effect.[30] Indeed, this led the subsequent National Academy of Sciences report to argue that "the real contribution to the strength of Ehrlich's statistical findings lies in the simple graph of the upsurge of the homicide rate after 1962, coupled with the fall in the execution rate in the same period."[31] While Ehrlich's contribution involved a sophisticated econometric technique, the National Academy report went on to note that his "whole statistical story lies in this simple pairing of these observations and not in the theoretical utility model, the econometric type specification, or the use of best econometric method. Everything else is relatively superficial and

28. Ehrlich, *supra* note 3, at 414.

29. Peter Passell & John B. Taylor, *The Deterrent Effect of Capital Punishment: Another View*, 67 AM. ECON. REV. 445 (1977).

30. *Id.* at 447.

31. Lawrence R. Klein et al., *The Deterrent Effect of Capital Punishment: An Assessment of the Estimates*, *in* DETERRENCE AND INCAPACITATION, *supra* note 5, at 336, 344.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

dominated by this simple statistical observation."[32]

Most recently, Dezhbakhsh and Shepherd have analyzed national time-series data from 1960 to 2000. In light of Figure 1, it is not surprising that they find a strong negative relationship between executions and the homicide rate.[33] While they do not report their results in terms of lives saved per execution, their estimates suggest that each execution reduces the homicide rate by about 0.05 homicides per 100,000 people, which translates to around 150 (!) fewer homicides per execution.

Why does the correlation between executions and homicides vary so much over time? One possibility is simply that the deterrent effect has truly changed over time and that capital punishment has suddenly become very effective starting in the 1990s. If so, more recent estimates are obviously to be preferred. If anything, however, administration of the death penalty has become both slower and execution methods less vivid, which would lead one to expect that any deterrent effect would be weakened in this period. Alternatively it may be that despite efforts in all of these studies to control for a range of social and economic trends, other omitted factors are preventing the relationship between executions and homicides from being correctly captured. To illustrate that these factors are indeed omitted from national time-series analyses, we introduce comparison groups into the analysis.

## III. THE IMPORTANCE OF COMPARISON GROUPS

As economists have come to understand how difficult it is to control convincingly for all relevant factors, many have lost faith in the ability of pure time-series analysis to isolate causal relationships. An alternative approach borrows a page from medical studies, emphasizing the importance of comparing results among those groups or regions receiving the "treatment" of the death penalty with a comparison group that is untreated, but otherwise susceptible to similar influences (a "placebo" or "control group"). If the execution rate is driving the homicide rate, then one should not expect to see a similar pattern in the homicide time series for these comparison groups.

### A. *Canada Versus the United States*

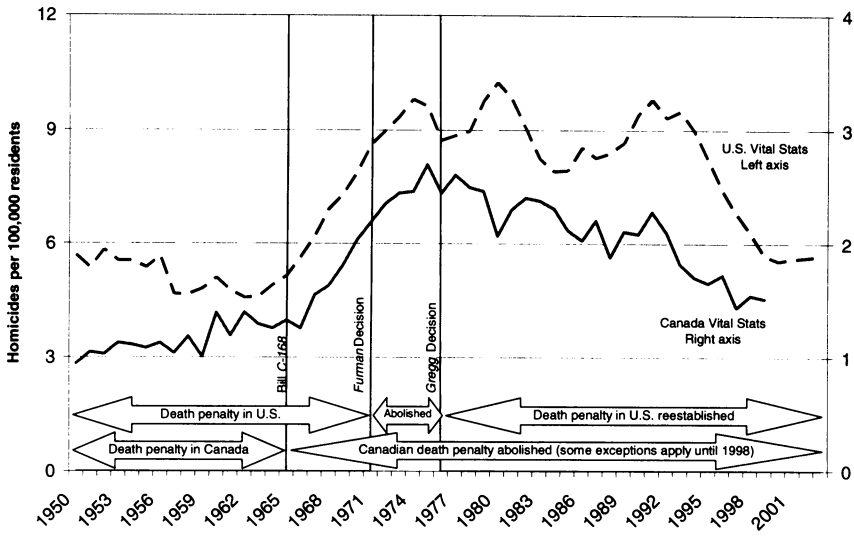
Given its proximity and different pattern of reliance on capital punishment, Canada presents an interesting comparison group for the United States, and Figure 2 compares the evolution of their homicide rates through time. The

32. *Id.*

33. Hashem Dezhbakhsh & Joanna M. Shepherd, *The Deterrent Effect of Capital Punishment: Evidence from a "Judicial Experiment"* tbls.3 & 4 (Am. Law & Econ. Ass'n Working Paper No. 18, 2004), *available at* http://law.bepress.com/cgi/viewcontent.cgi?article=1017&context=alea (last visited Dec. 4, 2005).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 2. Homicide Rates and the Death Penalty in the United States and Canada




Canadian homicide rate (right axis) is roughly one-third as high and one-third as variable as the rate in the United States (left axis).

The most striking finding is that the homicide rate in Canada has moved in virtual lockstep with the rate in the United States, while approaches to the death penalty have diverged sharply. Both countries employed the death penalty in the 1950s, and the homicide trends were largely similar. However, in 1961, Canada severely restricted its application of the death penalty (to those who committed premeditated murder and murder of a police officer only); in 1967, capital punishment was further restricted to apply only to the murder of on-duty law enforcement personnel.[34] As a result of these restrictions, no executions have occurred in Canada since 1962. Nonetheless, homicide rates in both the United States and Canada continued to move in lockstep. The *Furman* case in 1972 led to a death penalty moratorium in the United States. While many death penalty advocates attribute the subsequent sharp rise in homicides to this moratorium, a similar rise is equally evident in Canada, which was obviously unaffected by this U.S. Supreme Court decision. In 1976, the capital punishment policies of the two countries diverged even more sharply: the *Gregg* decision led to the reinstatement of the death penalty in the United States, while the death penalty was dropped from the Canadian criminal code.[35] Over the subsequent two decades, homicide rates remained high in the United

34. *See* DEP'T OF JUSTICE OF CANADA, FACT SHEET: CAPITAL PUNISHMENT IN CANADA (providing information on the history of the death penalty in Canada), *available at* http://canada.justice.gc.ca/en/news/fs/2003/doc_30896.html (last visited Nov. 21, 2005).

35. JOHN W. EKSTEDT & CURT T. GRIFFITHS, CORRECTIONS IN CANADA: POLICY AND PRACTICE 402 (2d ed. 1988).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

States while they fell in Canada. It is only over the last decade that homicide rates have started to decline in the United States, a fact that is difficult to attribute to reforms occurring decades earlier.

The Canadian move towards abolition is also interesting because it represented a major policy shock: prior to abolition, the proportion of murderers executed in Canada was considerably higher than that in the United States.[36] Of course, one might still be concerned that Canada is not quite an appropriate comparison group—perhaps Canada-specific factors were driving its homicide rate down following the abolition of its death penalty, back up during the U.S. moratorium, and back down over the ensuing period—effectively hiding the effects of execution-related changes. As such, it might be worth considering an alternative comparison group that is more clearly subject to the same set of economic and social trends.

### B. *Non-Death Penalty States Versus Other States in the United States*

Naturally, those states that have never had the death penalty should be unaffected by changes in death penalty policy throughout the rest of the country. Figure 3 facilitates the comparison of homicide rates across states that should be influenced by changes in death penalty law and practice from those that should not.

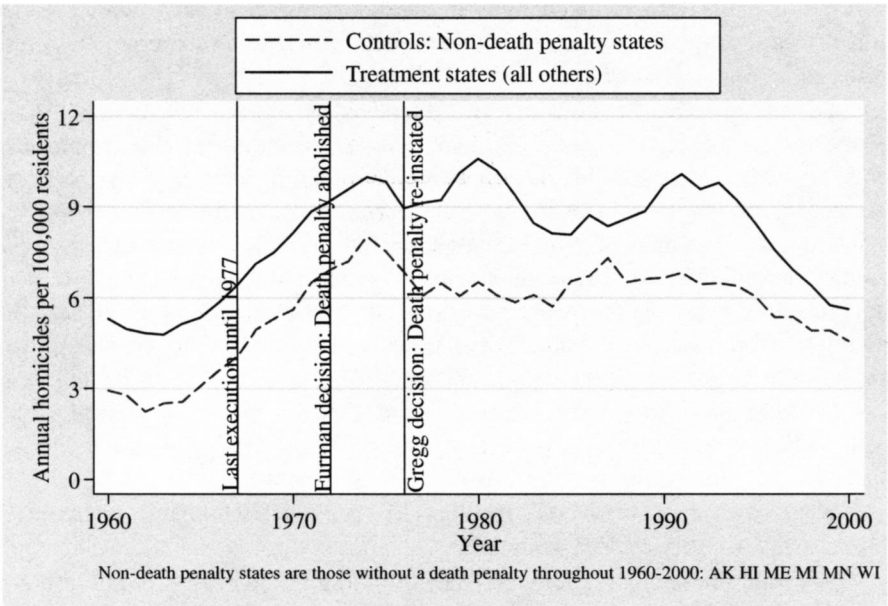
We begin by considering the cleanest comparison group: there are six states that have not had the death penalty on the books at any point in our 1960 to 2000 sample. Deterrence in these states was unaffected by either the *Gregg* or *Furman* decisions, and hence homicide rates in these states are a useful baseline for comparing the evolution of the homicide rates in other states. The remaining states are considered "treatment" states because either *Gregg* abolished their existing death penalties or *Furman* enabled their subsequent reinstatement (or, more commonly, both). Again, the most striking finding is

36. A comparison of the Canadian abolition experiment with the post-*Furman* Texas experiment is instructive. Over the two decades prior to abolition, the annual number of homicides in Canada fluctuated from around 150 to 250. *See Homicides*, DAILY, Oct. 1, 2003, *available at* http://www.statcan.ca/Daily/English/031001/d031001a.htm. From the 1970s to the 1990s, the number of murders in Texas was about ten times larger, fluctuating from 1200 to 2500 per year, despite having only half the population of Canada. *See* FED. BUREAU OF INVESTIGATION, *supra* note 20, at 15.

However, the number of executions was fairly similar: roughly seven per year in both Canada and Texas during the respective periods. Specifically, Canada had 148 executions for the years 1943 to 1962 (two decades before the policy change), or an average of 7.4 executions per year. *See* Richard Clark, Executions in Canada from Confederation to Abolition, *available at* http://www.geocities.com/richard.clark32@btinternet.com/canada.html (last visited Nov. 21, 2005). From 1977 to 1996 (two decades after the moratorium), Texas averaged seven executions per year. *See* ESPY FILE, *supra* note 26. As a result, the change in the likelihood that a homicide would result in execution caused by the Canadian death penalty abolition is an order of magnitude larger than that caused by Texas's reinstatement.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 3. Homicide Rates in the United States




the close co-movement of homicide rates in these two groups of states. Both sets of states experienced higher homicide rates during the death penalty moratorium than over the subsequent decade; the gap widened for the subsequent decade and narrowed only in the late 1990s. It is very difficult to find evidence of deterrence in these Supreme Court-mandated natural experiments that the death penalty has any causal effects at all on the homicide rate. Clearly, most of the action in homicide rates in the United States is unrelated to capital punishment.

The lesson from examining these time-series data is that it is crucial to take account of the fact that most of the variation in homicide rates is driven by factors that are common to both death penalty and non-death penalty states, and to both the United States and Canada. The empirical difficulty is that these factors may be spuriously correlated with executions, and hence the plausibility of any attempt to isolate the causal effect of executions rests heavily on either finding useful comparison groups or convincingly controlling for these other factors.

This issue is particularly relevant to Dezhbakhsh and Shepherd's analysis of changes in capital punishment laws. These authors present a series of before-and-after comparisons, focusing *only* on states that abolished the death penalty[37] or *only* on states adopting the death penalty.[38] Unfortunately, by

37. Dezhbakhsh & Shepherd, *supra* note 33, at tbl.5.
38. *Id.* at tbl.6.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

focusing only on the states experiencing these reforms, the authors risk confounding the effects of changes in capital punishment laws with broader forces that are equally evident in homicide data in states not experiencing these reforms.

The analysis by Dezhbakhsh and Shepherd is reproduced in Panel A of Table 1. The authors analyze each change in state laws during the sample. For each instance in which the death penalty was abolished, they compare the homicide rate one year prior to and one year after the abolition and report the average and median percentage change across all such abolitions. They also repeat this analysis for two- and three-year windows and for those times in which the death penalty was reinstated. Panel A exactly reproduces the numbers from their study, while Panel B shows our attempt at replicating their analysis.[39] In each case, they find that the *abolition* of the death penalty was associated with *rising* homicide rates, and the *reinstatement* of the death penalty was associated with *falling* homicide rates. Our replication largely succeeds in generating similar estimates: abolition of the death penalty is associated with a 10% to 20% increase in homicide, while reinstatement is associated with a 5% to 10% decrease.

However, these calculations may be confounding the effects of abolition or reinstatement of the death penalty with other broader trends. To test for this, we provide a comparison group for the abolition states in Panels A and B: we collect data on the change in homicide rates in all states that *did not* abolish the death penalty in that year.[40] These states did not experience any reform and so constitute a natural control group. Comparing Panel B with Panel C shows that the measured "effects" in states that changed their death penalty laws are similar to those in states that did not. Indeed, some of the "effects" in the comparison states are larger than those in the treatment states.

Panel D in Table 1 shows this formally, computing the difference between means (or medians) in treatment and control states—effectively a difference-in-differences approach. In no case do the figures in Panel D provide statistically or economically significant evidence for or against the deterrent effect. Half of the six estimates of the effects of abolition are positive and half are negative; the same is true for the effects of reinstating the death penalty. None of the estimates in Panel D are statistically significant. In sum, this analysis provides no evidence that the death penalty affects homicide rates and does not even paint a consistent picture of whether it is more likely to raise or lower rates.

39. They drop outliers from their calculation of the means, and we follow them in doing so; the medians are obviously more robust to such outliers. We were best able to match their numbers by assuming that North Dakota had capital punishment until *Furman*, although this seems a questionable judgment. Unfortunately, we cannot be confident of their coding because the authors were unwilling to share their data with us.

40. *See infra* Table 1, Panel C (the "control" states). Similarly, we collect the appropriate comparison groups for the states that reinstated the death penalty.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Table 1: Estimating How Changes in Death Penalty Laws Effect Murder: Selected Before and After Comparisons: 1960-2000

| **Dependent Variable: *% Change in State Murder Rates Around Regime Changes*** | | | | | | |
|---|---|---|---|---|---|---|
| | Death Penalty Abolition | | | Death Penalty Reinstatement | | |
| | 1-Year Window (1) | 2-Year Window (2) | 3-Year Window (3) | 1-Year Window (4) | 2-Year Window (5) | 3-Year Window (6) |
| | **Panel A: Reproducing Dezhbakhsh and Shepherd Tables 5, 6** | | | | | |
| Mean Change | 10.1%*** (2.8) | 16.3%*** (2.2) | 21.9%*** (2.5) | -6.3%** (3.4) | -6.4%** (2.9) | -4.1% (2.9) |
| Median Change | 8.3% | 14.9% | 18.4% | -9.3% | -6.8% | -7.5% |
| Number of States Where Homicide Increased | 33/45 | 39/45 | 41/45 | 12/41 | 16/39 | 13/39 |
| | **Panel B: Our Replication: Changes Around Death Penalty Shifts (Treatment)** | | | | | |
| Mean Change | 10.1%*** (2.9) | 16.0%*** (2.3) | 21.5%*** (2.6) | -6.3%* (3.4) | -7.0%** (2.9) | -3.8% (2.9) |
| Median Change | 8.5% | 13.8% | 18.5% | -9.3% | -8.5% | -7.4% |
| Number of States Where Homicide Increased | 35/46 | 39/46 | 41/46 | 12/41 | 15/39 | 14/39 |
| | **Panel C: Our Innovation: Changes in Comparison States (Control)** | | | | | |
| Mean Change | 8.7%*** (0.5) | 16.0%*** (0.8) | 20.6%*** (1.1) | -7.5%*** (1.5) | -6.6%*** (1.5) | -3.7%*** (1.3) |
| Median Change | 8.5% | 16.1% | 20.9% | -11.5% | -9.8% | -5.2% |
| Number of States Where Homicide Increased | 44/46 | 44/46 | 44/46 | 7/41 | 8/39 | 8/39 |
| | **Panel D: Difference-in-Difference Estimates (Treatment-Control)** | | | | | |
| Mean Change | 1.4% (2.9) | -0.1% (2.4) | 0.9% (2.8) | 1.2% (3.7) | -0.5% (3.2) | -0.1% (3.2) |
| Median Change | <0.001% (2.7) | -2.3% (2.5) | -2.4% (3.6) | 2.2% (3.5) | 1.3% (4.5) | -2.2% (2.0) |

Notes: Sources, data, and specification are as described in Dezhbakhsh & Shepherd, *supra* note 33, at tbls.5-6. Standard errors are in parentheses, and standard errors on median change are estimated by bootstrap. ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively. Panel A

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

estimates are evaluated using a one-tailed test, which makes it easier to find statistically significant evidence of deterrence. The rest of Table 1 follows our more conventional assumption that death penalty effects should be evaluated using a two-tailed test (thereby testing for either deterrence or antideterrence).

Each cell reports the mean or median percentage change in homicide rates in states that either abolished or reinstated the death penalty. The one-year window reports how murder rates changed from one year before abolition or reinstatement to one year after; the two-year window is the change in the homicide rate over the two years subsequent to reform compared to the two years before, with similar calculations for the three-year window. Panel A and our replication in Panel B might seem to suggest that crime rises when the death penalty is abolished and falls when it is reinstated, but Panel C shows that the same changes in murder rates also occur in the states that do not alter their death penalty laws (the control group). Panel D shows no differential change in murder rates between the treatment (change in death penalty law) and control groups (no change in death penalty law).

The estimates in Table 1 involve direct comparison of treatment and control states, but they do not account for other factors that may have affected the homicide rate differently in each state. This suggests that a panel data analysis may provide more reliable estimates. Sunstein and Vermeule argue that "a significant body of recent evidence [shows] that capital punishment may well have a deterrent effect, possibly a quite powerful one" and that "[a] wave of sophisticated multiple regression studies have exploited a newly available form of data, so-called 'panel data,' that uses all information from a set of units (states or counties) and follows that data over an extended period of time."[41] With this motivation, we now turn to expanding the above analysis into a formal panel structure.

## IV. PANEL DATA METHODS

The simplest panel data extension to the previous analysis above involves running the regression:

$$\frac{Murders_{s,t}}{(Population_{s,t}/100{,}000)} = \beta_1 Death\ Penalty\ Law_{s,t} + \sum_s State\ Effects_s + \sum_t Time\ Effects_t + \lambda Controls_{s,t} + \varepsilon_{s,t}$$

where the dependent variable is the homicide rate in a given state and year, and the variable of interest is an indicator set equal to one when a state has an active death penalty law. As such, $\beta_1$ measures the effects on the homicide rate of a state having a death penalty law in place. The inclusion of state fixed effects controls for persistent differences across states, the time fixed effects control for national time trends that are common across states, and control variables include indicators of state economic conditions, demographics, and law enforcement variables. Following Dezhbakhsh and Shepherd, we restrict our

41. Sunstein & Vermeule, *supra* note 12, at 706, 711.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

sample to the period from 1960 to 2000 and run a weighted least squares regression, clustering standard errors at the state level.

In Column 1 of Table 2, we report the results from Dezhbakhsh and Shepherd's estimation, in which they estimate the above equation without year fixed effects, but controlling for decade fixed effects.[42] Column 2 shows our replication attempt based on independently collected data (but using the same sources).[43] While our coefficient estimates do not precisely match theirs, the difference is tolerable. The real difference comes in the estimate of the standard error (which speaks to the persuasiveness of the data): we report a standard error nearly three times larger than theirs, and hence our coefficient is statistically insignificant. We do not know for certain the source of this divergence, and the authors provided no useful guidance. Thus, despite their claims that their estimates of "standard errors are further corrected for possible clustering effects—dependence within clusters (groups),"[44] our best guess is that they report simple ordinary least squares (OLS) standard errors. As Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan show, using OLS standard errors in panel estimation involving autocorrelated data may severely understate the standard deviation of the estimators (and hence exaggerate claims of statistical significance).[45]

Given the importance of not confounding overall crime trends in the 1970s with changes in death penalty laws (a lesson illustrated sharply in Table 1), we add controls for year fixed effects in Column 3. Indeed, in failing to control for year fixed effects, Dezhbakhsh and Shepherd's study is a clear outlier in the literature.[46] This is important: as Figure 2 shows, homicide rates were higher during the death penalty moratorium than during the early or late 1970s, and so simply controlling for the average crime rate in the 1970s would lead the

---

42. It is easy to lose this point: Dezhbakhsh and Shepherd refer only to controlling for "time-specific binary variables," and it was only through corresponding with the authors that we understood this to mean decade rather than year fixed effects. Dezhbakhsh & Shepherd, *supra* note 33, at 18. Indeed, they never use the term "decade" in connection with their econometric specification.

43. While Dezhbakhsh and Shepherd were unwilling to share their data for this Article, we have reconstructed it as closely as possible using the sources noted in their data appendix.

44. Dezhbakhsh & Shepherd, *supra* note 33, at 17.

45. Marianne Bertrand, Esther Duflo & Sendhil Mullainathan, *How Much Should We Trust Differences-in-Differences Estimates?*, 119 Q. J. ECON. 249 (2004).

46. Papers using year fixed effects include: Dezhbakhsh, Rubin & Shepherd, *supra* note 11; Joanna M. Shepherd, *Deterrence Versus Brutalization: Capital Punishment's Differing Impacts Among States*, 104 MICH. L. REV. 203 (2005) [hereinafter Shepherd, *Deterrence Versus Brutalization*]; Joanna M. Shepherd, *Murders of Passion, Execution Delays, and the Deterrence of Capital Punishment*, 33 J. LEGAL STUD. 283 (2004) [hereinafter Shepherd, *Murders of Passion*]; Zimmerman, *supra* note 11. Mocan & Gittings, *supra* note 11, both include year fixed effects and control for state-specific time trends. Katz, Levitt & Shustorovich, *supra* note 10, control for year fixed effects and, in various specifications, also control for state-specific trends, state-decade interactions, and separate time fixed effects by region.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

regression to find a deterrent effect, even though the same pattern was observed in states that experienced no change to their death penalty laws. It turns out that controlling for these confounding trends cuts the coefficient on the death penalty in half and makes the coefficient clearly statistically insignificant.

One possible objection to this analysis is that there are many states that are de jure death penalty states but de facto nonexecuting, and hence, the binary legal classification is inadequate. Thus, in Column 4 we make a distinction between those states that actively apply their death penalty statutes and those that do not. We define a death penalty statute as inactive if that state had no executions over the preceding ten years, an admittedly crude approach. In each case, we find no statistically significant effects of the death penalty. Moreover, the data suggest that active death penalty statutes are neither more nor less (in)effective than inactive death penalty statutes.

Table 2: Panel Data Estimates of the Effects of Death Penalty Laws on Murder Rates: 1960-2000

| **Dependent Variable:** ***Annual Homicides Per 100,000 Residents***$_{s,t}$ | | | | |
|---|---|---|---|---|
| | Dezhbakhsh and Shepherd (1) | Our Replication (2) | Controlling for Year Fixed Effects (3) | De Facto Versus De Jure Laws (4) |
| Death Penalty Law | -0.87*** | -0.95 | -0.47 | |
| | (.21) | (.57) | (.74) | |
| Active Death Penalty Law (≥ *1 Execution in Previous Decade*) | | | | -0.57 |
| | | | | (.63) |
| Inactive Death Penalty Law (*No Executions in Previous Decade*) | | | | -0.45 |
| | | | | (.77) |
| State Fixed Effects | Yes | Yes | Yes | Yes |
| Decade Fixed Effects | Yes | Yes | Yes | Yes |
| Year Fixed Effects | No | No | Yes | Yes |
| Adjusted $R^2$ | .804 | .791 | .834 | .834 |
| Sample Size (Excludes DC, HI) | (unknown) | 2009 | 2009 | 2009 |

Notes: Sources and data are as described in Dezhbakhsh & Shepherd, *supra* note 33, at tbl.7. Population-weighted least squares regression also includes controls for state per capita real income, the unemployment rate, police employment, proportions of the population nonwhite, aged 15-19, and aged 20-24. ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively.

Dezhbakhsh and Shepherd find that a death penalty law is associated with less crime, but our replication in Column 2, as well as other plausible estimates in Columns 3 and 4, show no significant effect.

The most important finding in Table 2 is simply how difficult it is to isolate

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

any causal effects with confidence. The standard errors in our preferred estimates suggest that even if death penalty laws deterred 15% of all homicides (or caused 15% more homicides), the data speak so unclearly that they could not rule out the possibility of no effect.

These data also allow us to extend the analysis of the distribution of estimates across death penalty experiments. Specifically, we extend our panel data approach, but rather than analyzing a single variable describing whether a state has a death penalty law, we estimate separate effects for each experiment.[47] That is, for each of the forty-five death penalty abolitions in the sample, we analyze its effects by including a separate dummy variable set equal to one for that state subsequent to the law change. We also include forty-one further dummy variables for each death penalty adoption in the sample. In all other respects, the specification remains the same as in Dezhbakhsh and Shepherd, although we continue to control for year fixed effects. Table 3 reports these results.

Table 3: Estimating the Individual Effects of Death Penalty Reform on the Homicide Rate for 41 Reinstatements and 45 Abolitions: 1960-2000

| **Dependent Variable:** ***Annual Homicides per 100,000 Residents***$_{s,t}$ | | | | | | |
|---|---|---|---|---|---|---|
| State | Death Penalty Reinstatement | | | Death Penalty Abolition | | |
| | Year | Estimated Effect | 95% Confidence Interval | Year | Estimated Effect | 95% Confidence Interval |
| Alabama | 1976 | -3.2 | (-4.1, -2.4) | 1972 | -1.2 | (-2.8, 0.5) |
| Arizona | 1976 | 1.1 | (0.2, 1.9) | 1972 | -1.5 | (-3.2, 0.2) |
| Arkansas | 1976 | -0.5 | (-1.4, 0.3) | 1972 | -2.4 | (-4.1, -0.8) |
| California | 1977 | 2.3 | (1.3, 3.2) | 1972 | 1.1 | (-0.8, 2.9) |
| Colorado | 1976 | -0.8 | (-1.9, 0.3) | 1972 | -1.7 | (-3.7, 0.2) |
| Connecticut | 1976 | 0.6 | (-0.8, 2.0) | 1972 | -2.5 | (-4.4, -0.6) |
| Delaware | 1976 | -2.2 | (-3.1, -1.4) | 1972 | -2.7 | (-4.6, -0.7) |
| | 1961 | -1.6 | (-2.2, -1.0) | | | |
| Florida | 1976 | -3.4 | (-4.2, -2.6) | 1972 | -0.2 | (-2.0, 1.5) |
| Georgia | 1976 | -5.1 | (-6.0, -4.3) | 1972 | 1.0 | (-0.6, 2.7) |
| Idaho | 1976 | 0.2 | (-0.6, 1.0) | 1972 | -2.8 | (-4.6, -1.0) |
| Illinois | 1976 | 0.3 | (-0.7, 1.2) | 1972 | -0.3 | (-2.2, 1.6) |
| Indiana | 1976 | 0.2 | (-0.5, 1.0) | 1972 | -0.4 | (-2.2, 1.4) |
| Iowa | | | | 1965 | -3.2 | (-4.7, -1.6) |
| Kansas | 1994 | 3.1 | (1.8, 4.4) | 1972 | -2.2 | (-4.1, -0.3) |
| Kentucky | 1976 | -1.6 | (-2.5, -0.8) | 1972 | -1.6 | (-3.3, 0.0) |
| Louisiana | 1976 | 1.4 | (0.7, 2.1) | 1972 | 1.5 | (-0.2, 3.2) |
| Maryland | 1976 | -0.6 | (-1.6, 0.4) | 1972 | -0.1 | (-2.1, 1.9) |
| Massachusetts | 1982 | -0.3 | (-1.2, 0.7) | 1972 | -2.8 | (-4.6, -0.9) |
| | | | | 1984 | -0.3 | (-1.0, 0.5) |
| Mississippi | 1976 | -1.9 | (-2.9, -0.9) | 1972 | 0.6 | (-1.1, 2.3) |

47. As such, this approach is a natural extension of the analysis in Table 1, with the advantage that panel analysis allows for regression-adjusted comparisons and takes account of the full time series, rather than an arbitrary comparison window. Note that while Table 1 included Washington, D.C., missing police data force us to drop it from this analysis.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

| | | | | | | |
|---|---|---|---|---|---|---|
| Missouri | 1976 | 0.3 | (-0.5, 1.0) | 1972 | -1.4 | (-3.1, 0.4) |
| Montana | 1976 | 0.6 | (-0.5, 1.8) | 1972 | -2.6 | (-4.5, -0.7) |
| Nebraska | 1976 | 0.3 | (-0.5, 1.1) | 1972 | -2.9 | (-4.8, -0.9) |
| Nevada | 1976 | -0.8 | (-1.8, 0.3) | 1972 | 1.2 | (-0.5, 2.9) |
| New Hampshire | 1991 | 0.1 | (-0.7, 1.0) | 1972 | -3.5 | (-5.4, -1.6) |
| New Jersey | 1982 | -1.3 | (-2.3, -0.2) | 1972 | -1.3 | (-3.3, 0.7) |
| New Mexico | 1979 | 0.3 | (-0.5, 1.1) | 1969 | 0.5 | (-0.9, 1.8) |
| New York | 1995 | -2.9 | (-4.4, -1.5) | 1965 | 2.9 | (1.0, 4.7) |
| North Carolina | 1977 | -2.4 | (-3.4, -1.5) | 1972 | -1.3 | (-3.0, 0.3) |
| North Dakota | | | | 1972 | -3.8 | (-5.6, -2.0) |
| Ohio | 1976 | -1.2 | (-1.9, -0.5) | 1972 | -0.4 | (-2.2, 1.3) |
| Oklahoma | 1976 | 1.1 | (0.3, 1.8) | 1972 | -1.8 | (-3.5, -0.1) |
| Oregon | 1978 | -0.6 | (-1.6, 0.4) | 1964 | -1.8 | (-2.8, -0.7) |
| Pennsylvania | 1976 | -0.1 | (-0.9, 0.7) | 1972 | -0.9 | (-2.6, 0.8) |
| Rhode Island | 1977 | -1.1 | (-2.4, 0.2) | 1984 | 0.6 | (0.1, 1.0) |
| South Carolina | 1976 | -4.8 | (-5.6, -3.8) | 1972 | -0.5 | (-2.2, 1.2) |
| South Dakota | 1979 | 0.5 | (-0.1, 1.1) | 1972 | -4.4 | (-6.3, -2.6) |
| Tennessee | 1976 | -2.1 | (-2.9, -1.3) | 1972 | -0.1 | (-1.8, 1.7) |
| Texas | 1976 | -0.1 | (-1.1, 0.9) | 1972 | -0.1 | (-1.7, 1.6) |
| Utah | 1976 | 0.8 | (-0.1, 1.6) | 1972 | -3.1 | (-4.8, -1.4) |
| Vermont | | | | 1965 | -2.9 | (-4.4, -1.4) |
| Virginia | 1976 | -2.7 | (-3.6, -1.7) | 1972 | -2.0 | (-3.8, -0.3) |
| Washington | 1976 | 0.7 | (-0.5, 1.9) | 1972 | -1.8 | (-3.6, -0.0) |
| West Virginia | | | | 1965 | -2.8 | (-4.5, -1.0) |
| Wyoming | 1977 | -0.9 | (-1.5, -0.2) | 1972 | -3.4 | (-5.3, -1.4) |
| Simple Average | | -0.70 | | | -1.32 | |
| Precision-Weighted Average | | -0.67 | | | -0.86 | |
| Population-weighted Average | | -0.72 | | | -0.39 | |

Notes: This table shows the effect on murder rates of forty-one reinstatements of death penalty laws and forty-five abolitions of such laws. It is derived from the same data and models that were used to estimate aggregated effects of such legal changes averaged over all switching states (in Table 2, *infra*). Alaska, Hawaii, Maine Michigan, Minnesota, and Wisconsin are not shown because they never had the death penalty throughout the sample period (and there is some debate over North Dakota). The District of Columbia and Hawaii were dropped from the sample because of missing police data. Sources, data, and specification follow Dezhbakhsh & Shepherd, *supra* note 33, at tbl.7, as described in Table 2, except that we add year fixed effects and include forty-one death penalty reinstatements and forty-five death penalty abolition dummy variables (set equal to zero before the change and one subsequently), rather than a single binary variable covering all eighty-six experiments. Controls include per capita real income; the unemployment rate; police employment; proportions of the population nonwhite, aged 15-19, and aged 20-24; and state and year fixed effects. Standard errors are in parentheses, clustered at the state level. The precision-weighted average is generated by weighting by the inverse of the squared standard error.

For neither death penalty abolitions nor reinstatements do we see a

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

particularly coherent picture. Estimates of the "effect" of death penalty abolition on the homicide rate (conditional on the control variables) are positive in eight cases and negative in thirty-seven cases. Likewise, reinstatement of the death penalty was subsequently associated with a higher homicide rate in seventeen states and a lower rate in twenty-four states. On average, the homicide rate appears to be lower than otherwise suggested by developments in the control variables following *either* abolition or reinstatement of the death penalty. That said, these differences are not statistically significant, and these comparisons merely point to the difficulty in discerning *any* causal effect of death penalty laws.

Figure 4 shows the distribution of before-and-after comparisons across states, using the data in Table 3. These distributions highlight the problem of getting these data to speak clearly: the variance of individual state homicide rates is so great that it is difficult to discern the average effects of these changes with any precision, even with eighty-six "experiments" to analyze. Shepherd has performed a related reanalysis of three papers that examine the effects of executions (rather than the presence of a death penalty law), and she also finds that there are about as many states whose experiences are consistent with the deterrence hypothesis as with the antideterrence one.[48]

It is worth noting that Mocan and Gittings also include an analysis of the efficacy of death penalty laws over a sample running from 1977 to 1997, although their regressions only include data from 1980 to 1997.[49] Despite their professed confidence in their results, Mocan and Gittings's analysis includes only six policy change experiments. We have reanalyzed their data following a similar design to that above: we follow their data and programs (which they graciously shared) but analyze the death penalty "effects" separately for each state, making sure to control for the same variables as in their main specification. For the four states adopting the death penalty, their specification suggests that homicide rates were subsequently higher in Kansas and New Hampshire and lower in New Jersey and New York. In their sample, only Massachusetts and Rhode Island abolished the death penalty, and in both cases homicide rates fell following the law change (relative to the baseline established by their regression). These facts make it difficult to conclude with any confidence that the death penalty raises or lowers homicide rates.[50]

---

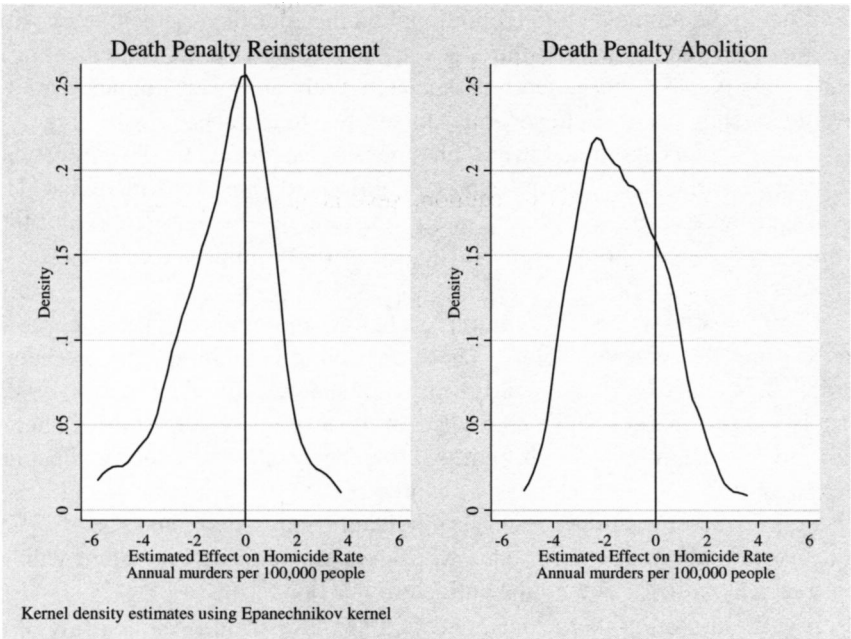
48. *See* Shepherd, *Deterrence Versus Brutalization*, *supra* note 46 (reanalyzing data from Dezhbakhsh & Shepherd, *supra* note 33, Dezhbakhsh, Rubin & Shepherd, *supra* note 11, and Shepherd, *Murders of Passion*, *supra* note 46). Shepherd argues that antideterrence is evident in some states because they do not execute sufficient convicts to reach a "threshold effect" required for deterrence.

49. Mocan & Gittings, *supra* note 11, at 478.

50. That Mocan and Gittings obtain statistically significant estimates reflects the fact that New York and New Jersey were the two states consistent with deterrence, and their influence in a population-weighted regression dwarfs that of the four states inconsistent with deterrence. *Id.*

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 4. Distribution of Regression-Estimated Effects Across States




Given the demonstrated difficulties in linking the presence of death penalty laws with homicide rates, several authors also have tried to exploit variation in the intensity with which death penalty laws have been applied. Consequently, the variable of interest in these studies does not describe the presence of a death penalty law but rather a variable measuring the propensity to invoke the death penalty. The intensity with which a state pursues death penalty prosecutions may be highly politicized, raising the possibility that such estimates may reflect omitted factors related to the political economy of punishment. On the demand side, variation in crime rates may change the political pressure for executions. Equally on the supply side, it seems plausible that more vigorous deployment of the death penalty might occur at the same time that the government elects to "get tough on crime" along a range of other dimensions, including sentencing, prison conditions, arrests, police harassment, and so on. As these studies move beyond the sharp judicial or legislative experiments analyzed above, the issues involved in distinguishing correlation from causation may become even more salient.

However as Katz, Levitt, and Shustorovich emphasize, beyond the usual difficulties in establishing a causal relationship, there is a much simpler statistical dilemma: the annual number of executions fluctuates very little while the number of homicides varies dramatically. Under these conditions, it is "a difficult challenge to extract the execution-related signal from the noise in

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

homicide rates."[51] Indeed, following their own empirical investigation for the years 1950 to 1990, Katz, Levitt, and Shustorovich conclude that "[e]ven if a substantial deterrent effect does exist, the amount of crime rate variation induced by executions may simply be too small to be detected"[52] and that "[t]here simply does not appear to be enough information in the data on capital punishment to reliably estimate a deterrent effect."[53]

Countering these words of caution, several recent studies claim to have compiled robust evidence of the deterrent effect of capital punishment. We begin by updating Katz, Levitt, and Shustorovich's study to incorporate data revisions and add data from 1991 to 2000, before turning to these alternative studies.

### A. *Katz, Levitt, and Shustorovich*

Katz, Levitt, and Shustorvich generously provided us with their 1950 to 1990 dataset, so we were easily able to replicate their results. These authors regressed state homicide rates on the number of executions per 1000 prisoners (with a rich set of controls), concluding that "the execution rate coefficient is extremely sensitive to the choice of specification . . . ."[54] Panel A of Table 4 shows our replication of their original estimates over the 1950 to 1990 sample using revised data; these estimates are very close to those reported in their paper.[55] Panel B reports results over our updated 1950 to 2000 sample, while Panel C analyzes the largest possible sample, extending back as far as 1934 and forward through to 2000.

Reading across each row, estimates of the effects of executions on the homicide rate appear quite inconsistent across specifications, with point estimates ranging from positive to negative in Panels A and B. Reading down each column, we see that this inconsistency holds across time periods as well; while several specifications are consistent with deterrence for the 1950 to 1990 sample, these results largely disappear if the models are estimated over the slightly longer period from 1950 to 2000 (Panel B). Indeed, Panel C reveals that when the models are estimated over the longest period (1934 to 2000), the signs reverse, and executions are associated with *higher* rates of murder. In sum, the alternative samples continue to point to the difficulty in pinning down robust estimates of the deterrent effect of the death penalty suggested by Katz, Levitt, and Shustorovich.

---

51. Katz, Levitt & Shustorovich, *supra* note 10, at 319.

52. *Id.*

53. *Id.* at 321-22.

54. *Id.* at 330.

55. Note that we report standard errors clustered at the state level, although this makes little practical difference because Katz, Levitt, and Shustorovich reported standard errors clustered at the state-decade level.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Table 4: Estimating the Effect of Executions on Murder Rates Using the Katz, Levitt, and Shustorovich Model for Three Time Periods: 1934-2000

| **Dependent Variable:** ***Homicides per 100,000 Residents$_{s,t}$*** | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
|---|---|---|---|---|---|---|---|---|
| | **Panel A: Replication for 1950-1990 Sample** | | | | | | | |
| Executions$_{s,t}$ per 1000 Prisoners$_{s,t}$ | 0.32 (.38) | -0.67** (.33) | -0.31 (.31) | -0.56** (.30) | 0.01 (.20) | -0.07 (.14) | -0.08 (.14) | -0.22* (.14) |
| | **Panel B: Augmented Sample—1950-2000** | | | | | | | |
| Executions$_{s,t}$ per 1000 Prisoners$_{s,t}$ | 0.48 (.45) | -0.58 (.38) | -0.20 (.37) | -0.39 (.40) | -0.14 (.22) | -0.29 (.20) | -0.07 (.14) | -0.23 (.14) |
| | **Panel C: Maximum Sample—1934-2000** | | | | | | | |
| Executions$_{s,t}$ per 1000 Prisoners$_{s,t}$ | 1.54*** (.34) | 0.19 (.27) | 0.48 (.30) | 0.20 (.26) | 0.67*** (.24) | 0.31 (.19) | 0.06 (.12) | -0.02 (.12) |
| | **Implied Life-Life Tradeoff[a] [95% Confidence Interval]** | | | | | | | |
| Panel A: 1950-1990 | -1.8 [-3.6,-0.1] | 0.6 [-0.9,2.2] | -0.2 [-1.7,1.3] | 0.4 [-1.1,1.8] | -1.0 [-2.0,-0.1] | -0.8 [-1.5,-0.1] | -0.8 [-1.5,-0.2] | -0.5 [-1.1,0.2] |
| Panel B: 1950-2000 | -2.2 [-4.3,0.0] | 0.4 [-1.4,2.2] | -0.5 [-2.3,1.2] | -0.1 [-2.0,1.9] | -0.7 [-1.7,0.4] | -0.3 [-1.2,0.7] | -0.8 [-1.5,-0.2] | -0.4 [-1.1,0.2] |
| Panel C: 1934-2000 | -4.7 [-6.4,-3.1] | -1.5 [-2.7,-0.2] | -2.2 [-3.6,-0.7] | -1.5 [-2.7,-0.2] | -2.6 [-3.8,-1.5] | -1.7 [-2.6,-0.9] | -1.1 [-1.7,-0.6] | -1.0 [-1.5,-0.4] |
| | **Further Controls** | | | | | | | |
| Crime, Economic & Demographic Controls | No | Yes | No | Yes | No | Yes | No | Yes |
| State Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes | Yes | Yes |
| Region*Year Effects | No | No | Yes | Yes | No | No | No | No |
| State Time Trends | No | No | No | No | Yes | Yes | No | No |
| State*Decade Effects | No | No | No | No | No | No | Yes | Yes |

Notes: Panel A shows the Katz, Levitt, and Shustorovich estimates of the impact of executions on murder rates (using revised data). Panels B and C show how those estimates change using longer time periods, with all estimated effects showing increased execution rates correlated with increased murder rates for the full sample. The bottom half of the table shows the corresponding life-life tradeoff numbers, where negative numbers mean that net lives are lost for each execution. Note that in order to obtain the long samples in Panel C, we drop the infant mortality and unemployment rates as controls; this longer sample also introduces a few more missing data cells.

The eight specifications and data sources are as in Katz, Levitt & Shustorovich, *supra* note 10, at 327 tbl.2. Crime controls include prisoner

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

death rate, prisoners per crime, and prisoners per capita. Economic controls include the real per capita income, insured unemployment rate, and the infant mortality rate. The latter two are not included in Panel C. Demographic controls are the proportion of the population: black, urban, aged 0-24, and aged 25-44. Sample sizes are 1908, 2414, and 2954 for state-year observations in Panels A, B, and C, respectively, and all panels omit 1971 due to missing data on prison deaths. Population-weighted least squares regression is used, and standard errors are clustered at the state level. ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively.

*(a)* Implied life-life tradeoff reflects net lives saved when evaluated for a state with the characteristics of the average death penalty state in 1996.

In order to remain consistent with the debate about what Sunstein and Vermeule refer to as the "life-life tradeoff,"[56] we also compute the implied number of lives saved per execution. In order to fix a particular set of parameters (and to maintain continuity with the numbers reported by Dezhbakhsh, Rubin, and Shepherd), we report the implied net number of lives saved by an execution for a state with the characteristics of the average death penalty state in 1996 (holding all other factors constant).[57] Given that Table 4 involves the largest sample of data in our analysis, it is not surprising that the 95% confidence intervals surrounding these estimates, while wide, imply these estimates are notably more precise than we obtain with other specifications in Tables 5 through 9.

### B. *Dezhbakhsh and Shepherd*

The Dezhbakhsh and Shepherd study covers data from 1960 to 2000, and their analysis of the effects of executions largely shadows their analysis of the effects of death penalty laws.[58] That is, they run the same regression as described in Table 2, but replace the death penalty binary variable with a variable intended to capture the propensity to invoke the death penalty. The first column of Table 5 shows their reported results, while the next column shows the same regression, controlling for year fixed effects. As before, we

56. *See* Sunstein & Vermeule, *supra* note 12, at 708 (introducing the concept of a "life-life tradeoff" in the capital punishment debate).

57. To compute this, note that executing one more death row inmate raises the execution rate from $X/P$ to $(X+1)/P$, where $X$ is the number of executions, and $P$ is the denominator of the execution rate, which in this instance is the number of prisoners. The effect of the execution rate on the homicide rate is mediated by the estimated coefficient, $\beta$, yielding a decline in the homicide rate of $-\beta/P$. To determine the number of lives saved, we need to multiply the decline in the homicide rate (homicides per 100,000 people) by the population/100,000, and subtract one to take account of the executed convict. Thus a tradeoff of zero implies that each execution kills one convict and saves one homicide victim; a positive number implies that more than one homicide victim is saved, and a negative number suggests that each execution results in a greater number of total deaths.

58. Dezhbakhsh & Shepherd, *supra* note 33.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

continue to report standard errors clustered at the state level. Superficially, these results suggest extremely significant evidence in favor of deterrence.

Table 5: Estimating the Impact of Executions on Murder Rates, Testing the Sensitivity of Dezhbakhsh and Shepherd's Results: 1960-2000

| **Dependent Variable:** ***Annual Homicides per 100,000 Residents***$_{s,t}$ | | | | | | |
|---|---|---|---|---|---|---|
| | Published | Adding Year Fixed Effects | Omitting Texas | Alternative Definitions of Execution Risk | | |
| | (1) | (2) | (3) | (4) | (5) | (6) |
| Executions$_{s,t}$ | -0.145*** | -0.138*** | -0.137* | | | |
| | (.013) | (.013) | (.070) | | | |
| Executions$_{s,t}$ per 100,000 Residents$_{s,t}$ | | | | -8.36 | | |
| | | | | (5.84) | | |
| Executions$_{s,t}$ per 1000 Prisoners$_{s,t}$ | | | | | -0.38 | |
| | | | | | (0.47) | |
| Executions$_{s,t}$ per Homicide$_{s,t-1}$ | | | | | | -50.7 |
| | | | | | | (31.7) |
| *N* | (unknown) | 2009 | 1968 | 2009 | 2009 | 2009 |
| | **Implied Life-Life Tradeoff**[(a)] **[95% Confidence Interval]** | | | | | |
| Net Lives Saved per Execution | 7.8 | 7.3 | 7.3 | 7.4 | -0.1 | 5.0 |
| | [6.1, 9.4] | [5.8., 8.9] | [-1.0, 15.5] | [-4.1,18.8] | [-2.4, 2.2] | [-2.3, 12.3] |
| | **Further Controls** | | | | | |
| State Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Decade Fixed Effects | Yes | Yes | Yes | Yes | Yes | Yes |
| Year Fixed Effects | No | Yes | Yes | Yes | Yes | Yes |

Notes: Column 1 shows the results for the estimated impact of executions on murder rates reported in Dezhbakhsh & Shepherd, *supra* note 33, tbl.7 (and the basic specification and data sources are as described therein). Controls include per capita real income, the unemployment rate, police employment, proportions of the population nonwhite, aged 15-19, and aged 20-24. Column 2 begins by adding in year fixed effects, and Column 3 shows the estimated effects of executions on murder rates become much less precisely estimated when Texas is omitted. Columns 4-6 also show that the estimated effect of executions becomes insignificant when various measures of the execution rate are analyzed, instead of the raw number of executions. The bottom portion of Table 5 shows the corresponding life-life tradeoff numbers, where negative numbers mean that, on net, lives are lost for each execution.

Population-weighted least squares regression is used. Standard errors are clustered at the state level. ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively.

*(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996.

However, as Richard Berk has noted, the distribution of executions across states is extraordinarily skewed.[59] Through 2004, Texas has executed 336

59. Richard Berk, *New Claims About Executions and General Deterrence: Déjà Vu All*

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

convicts since the *Gregg* decision. The next closest state is Virginia at 94 executions, while only ten other states have recorded more than twenty executions and seventeen states have recorded no executions.[60] As a result, it seems useful to test the sensitivity of the baseline equation to the omission of Texas. While the effect on the coefficient reported in Column 3 of Table 5 is rather small, the effect on the estimated standard error is dramatic, and the estimated impact of executions becomes statistically insignificant. Similarly, Shepherd has shown that the evidence for deterrence in these data rests critically on variation arising from a few states, and the vast majority of states experienced either no deterrence or antideterrence.[61] The implication of our Table 5, however, is not that Texas is an outlier (indeed, given the constancy of the coefficient, it probably lies along the regression line), but rather that in its absence, there is just too little variation in executions to discern an effect with any confidence.

A more direct difficulty with Dezhbakhsh and Shepherd's specification is that the independent variable is simply the number of executions in that state each year. Not only does this exaggerate the problem of Texas (the large number of executions partly reflects the fact that there are more people and more murders in Texas than in many other states), but it also is a somewhat bizarre choice. For example, this specification implies that one more execution in Wyoming would deter three-fourths of a homicide, while in California it would deter fifty homicides.

A very simple alternative that avoids this scaling issue is measuring executions per 100,000 residents. These results are reported in Column 4, and this regression suggests that the relationship between homicides and executions per capita is statistically insignificant.

An alternative scaling comes from Katz, Levitt, and Shustorovich, who define their executions variable as executions per 1000 prisoners.[62] This regression, shown in Column 5, again fails to find a significant relationship between homicide and execution rates, with the point estimate suggesting that each execution deters 0.9 homicides for a net loss of 0.1 life. Another alternative scaling—and perhaps the one most directly suggested by the economic model of crime—is to analyze the ratio of the number of executions to the (lagged) homicide rate.[63] Once again, this regression, shown in

---

*over Again?*, 2 J. EMPIRICAL LEGAL STUD. 303, 305 (2005).

60. BONCZAR & SNELL, *supra* note 8, at 9 tbl.9.

61. Shepherd, *Deterrence Versus Brutalization*, *supra* note 46.

62. This alternative scaling yields a slightly smaller sample because data on the number of prisoners in Alaska are not available until 1972. For other missing values of the prisoner variable, we simply use linear interpolation.

63. We use the lagged homicide rate so that the number of homicides does not appear in the construction of both the independent and dependent variables. Specifically, if there were measurement errors in the number of homicides, this would cause the dependent variable to increase (decrease) and the independent variable to decrease (increase), creating

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Column 6, fails to find any significant relationship.

### C. *Mocan and Gittings*

Mocan and Gittings examine state homicide rates over the 1984 to 1997 (post-moratorium) period,[64] running the following regression:

$$\frac{Murders_{s,t}}{(Population_{s,t}/100{,}000)} = \beta_1 \frac{Executions_{s,t-1}}{DeathSentences_{s,t-7}} + \beta_2 \frac{Pardons_{s,t-1}}{DeathSentences_{s,t-7}} + \beta_3 \frac{Removals_{s,t-1}}{DeathSentences_{s,t-6}}$$
$$+\beta_4 \frac{DeathSentences_{s,t-1}}{Arrests_{s,t-3}} + \beta_5 \frac{HomicideArrests_{s,t-1}}{Murders_{s,t-1}} + \gamma Controls_{s,t} + \sum_s State_s * Trend_t + \sum_t Time_t + \varepsilon_{s,t}$$

The authors provided us with their data, and Panel A of Table 6 shows that we were able to replicate their results. In the process of doing so, we found a number of coding errors, and a set of corrected estimates is given in Panel B.[65] These estimates are reasonably similar to those found in Panel A, although in no case are any of the estimates of the effects of executions statistically significant. Equally, the effects of death row removals appear somewhat stronger in these numbers.

One feature that is immediately obvious from inspecting their model is that it has a rather complex temporal structure: the variables of interest are constructed as ratios to the number of death sentences imposed six or seven years earlier or the number of arrests three years earlier. While the authors choose this functional form to maintain continuity with Dezhbakhsh, Rubin, and Shepherd, this rather contrived structure comes at a significant price. Their data only runs from 1977 to 1997, and hence this lag specification costs them one-third of their sample since their deterrence variables are only defined over the 1984 to 1997 period. Moreover, given that the authors are attempting to represent the probability of execution as perceived by potential murderers, and given the paucity of evidence on how these expectations are formed, there seems little reason to strongly prefer one specification over the other. Thus, in Panel C, we rerun their regressions but note Zimmerman's argument that "any truly meaningful (subjective) assessment a potential murderer makes . . . is likely to be based upon the most recent information available to him/her."[66]

---

an artificial negative correlation between execution and homicide rates.

64. Mocan & Gittings, *supra* note 11, at 478. While their data runs from 1977 to 1997, their complicated lag structure means that they can only estimate effects from 1984 onward.

65. Two types of coding errors were discovered. First, the authors attempted to drop all observations where the explanatory variable was the ratio of a positive value to zero but ended up both dropping the prior observation and including the variable they intended to drop, coded as the ratio of the numerator to 0.99. Second, in Models 3, 5, and 6, the execution rate was defined relative to the number of death sentences six years prior instead of seven years prior, as they did in their other specifications (and described in their text).

66. Zimmerman, *supra* note 11, at 170.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Table 6: Estimating the Impact of Executions on Murder Rates: Reanalyzing Mocan and Gittings: 1977-1997

| | **Dependent Variable:** | | | | | | |
|---|---|---|---|---|---|---|---|
| | ***Annual Homicides per 100,000 Residents$_{s,t}$*** | | | | | ***Log Homicide Rate$_{s,t}$*** | |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) |
| | **Panel A: Mocan and Gittings Results: Replication** | | | | | | |
| Executions$_{t-1}$ per Death Sentence$_{t-7}$ | -0.60$^{*}$ (.35) | | | -0.63$^{*}$ (0.34) | -0.63$^{**}$ (.29) | -0.05$^{*}$ (.03) | -0.05$^{*}$ (.03) |
| Pardons$_{t-1}$ per Death Sentence$_{t-7}$ | | 0.69$^{**}$ (.32) | | 0.73$^{**}$ (.30) | | 0.11$^{***}$ (.03) | |
| Death Row Removals$_{t-1}$ per Death Sentence$_{t-6}$ | | | 0.17$^{**}$ (.07) | | 0.18$^{**}$ (.07) | | 0.02$^{**}$ (0.01) |
| Sample (1984-1997) | 680 | 693 | 695 | 679 | 690 | 679 | 690 |
| | **Panel B: Correcting Programming Errors** | | | | | | |
| Executions$_{t-1}$ per Death Sentence$_{t-7}$ | -0.50 (.34) | | | -0.52 (.33) | -0.59 (0.39) | -0.01 (0.03) | -0.02 (0.02) |
| Pardons$_{t-1}$ per Death Sentence$_{t-7}$ | | 0.63$^{*}$ (.34) | | 0.71$^{**}$ (.30) | | 0.09$^{***}$ (0.03) | |
| Death Row Removals$_{t-1}$ per Death Sentence$_{t-6}$ | | | 0.24$^{***}$ (.08) | | 0.17$^{*}$ (0.09) | | 0.01 (0.01) |
| Sample (1984-1997) | 679 | 692 | 691 | 677 | 636 | 677 | 636 |
| | **Panel C: Measuring Deterrence Variables with a One-Year Lag on Full Sample** | | | | | | |
| Executions$_{t-1}$ per Death Sentence$_{t-1}$ | 0.03 (0.14) | | | 0.01 (0.13) | 0.01 (0.14) | 0.01 (0.01) | 0.01 (0.01) |
| Pardons$_{t-1}$ per Death Sentence$_{t-1}$ | | 0.41$^{***}$ (.13) | | 0.41$^{***}$ (0.13) | | 0.05$^{***}$ (0.01) | |
| Death Row Removals$_{t-1}$ per Death Sentence$_{t-1}$ | | | 0.02 (0.03) | | 0.02 (0.03) | | 0.002 (.002) |
| Sample (1978-1997) | 986 | 984 | 921 | 977 | 918 | 977 | 918 |
| | **Implied Life-Life Tradeoff for Executions$^{(a)}$ [95% Confidence Interval]** | | | | | | |
| Panel A: Replication | 4.4 [-1.8, 10.5] | | | 4.6 [-1.4, 10.6] | 4.6 [-0.5, 9.7] | 2.2 [-1.2, 5.7] | 2.3 [-1.3, 6.0] |
| Panel B: Corrected | 3.4 [-2.6, 9.4] | | | 3.6 [-2.2, 9.5] | 4.2 [-2.6,11.1] | -0.2 [-3.7, 3.4] | 0.5 [-2.7, 3.7] |
| Panel C: Full Sample | -1.2 [-3.1, 0.7] | | | -1.1 [-2.8, 0.7] | -1.1 [-3.0, 0.8] | -1.6 [-2.7, -0.5] | -1.6 [-2.8, -0.4] |

Notes: Panel A shows the estimated effect of executions on the homicide rate, where the specification and data are from Mocan & Gittings, *supra* note 11, at 464 tbl.2. Panel B corrects some programming errors, and the resulting estimated effects of execution on murder rates are no longer significant. Panel C alters the measure of the deterrence variables and uses the full sample period from 1978 to 1997, which leads to a positive correlation between

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

execution and homicide rates. The bottom portion of the table shows the corresponding life-life tradeoff numbers, where negative numbers mean that, on net, lives are lost for each execution.

Controls include lags of the homicide arrest rate, death sentence rate (conditional on arrest), prisoners per violent crime, prison death rate, as well as contemporary values of real per capita income, the unemployment rate, infant mortality rate, shares of the population who are: urban, black, aged 20-34, 35-44, 45-54, 55+, and dummy variables for whether a state has a Republican governor, whether the state drinking age is 18, 19, or 20, and the 1995 Oklahoma City bombing. Population-weighted least squares regression is used with standard errors clustered at the state level. ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively.

*(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996.

In Panel C, we construct each of the deterrence variables as ratios of variables lagged one year (instead of seven).[67] This relatively small change yields positive, albeit insignificant, coefficients. The difficulty in obtaining any consistent results is once again evident. Not only do the estimates of the effects of the execution rate vary significantly with only minor changes in specification, but the two related measures of the porosity of the death sentence now yield sharply different results, with the pardon rate robustly and positively associated with homicide, but the coefficient on the broader death row removal rate small and insignificant.

### D. *Other Studies*

At least four other studies are worthy of brief discussion. First, Zimmerman analyzes a state panel of homicide rates over the period from 1978 to 1997, and his OLS regressions suggest no relationship between homicide rates and the execution rate. (We comment on his instrumental variables results in the next Part.) This is consistent with our reanalysis of Mocan and Gittings's data over the same time period.

Second, Dezhbakhsh, Rubin, and Shepherd analyze a quite impressive county-level dataset covering the period from 1977 to 1996. While their paper only reports instrumental variables results (more on these below), the authors have generously shared their data with us, and we have computed simple panel OLS results, borrowing all other aspects of their specification. Again, we find wildly inconsistent results across specifications, ranging from statistically

67. The immediate advantage of using the one-year lag is that the sample size increases by fifty percent from what Mocan and Gittings present. We remain unsure whether Mocan and Gittings or Zimmerman (or neither) is correct on the appropriate lag structure because there is little evidence on how criminals form their expectations. Even so, if a small change among reasonable choices makes a large difference in the estimation, then the results are too fragile to warrant reliance.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

significant antideterrent effects to statistically significant deterrent effects. Disaggregating to the county level does not alleviate the problems we have seen with state-level analyses. This should not be surprising because the study's key explanatory variable, the execution rate, is still measured at the state level.[68]

Third, Dale Cloninger and Roberto Marchesini have analyzed data on the recent Illinois moratorium experiment.[69] Governor Ryan issued a moratorium on executions in January 2000 and subsequently commuted all death sentences in January 2003. It seems useful to compare the evolution of homicides in Illinois subsequent to January 2000 with the same evolution in the rest of the country. The methods employed by Cloninger and Marchesini reflect the authors' backgrounds as financial economists: they apply an event study methodology, examining the usual co-movement of the number of homicides in Illinois with the number of homicides nationally and then asking whether this relationship changed following the Illinois moratorium.

The main difficulty with their analysis is that they follow finance methods a little too closely. In finance, the variable of interest is usually a stock return, so it is standard practice to take a stock index and analyze its percentage change over some period. As such, Cloninger and Marchesini analyze the relationship between twelve-month-ended *growth* in the homicide rate in Illinois and their comparison sample. However, the debate over the efficacy of capital punishment is usually posed as asking whether it leads to lower *levels* of homicide, rather than a differential growth rate.[70] Moreover, differential growth rates—if interpreted literally—would lead to predictions that homicide rates may head to 0% or 100%.

Cloninger and Marchesini generously shared their monthly data (covering January 1994 to December 2003) with us, and Figure 5 shows the seasonally adjusted number of homicides in Illinois and in the rest of the United States through this time period. The close relationship between the two again supports the contention that *levels* of homicide provide a useful baseline against which to compare the subsequent experience in Illinois. Figure 5 also shows a dashed line: the projected number of homicides in Illinois if the relationship between

---

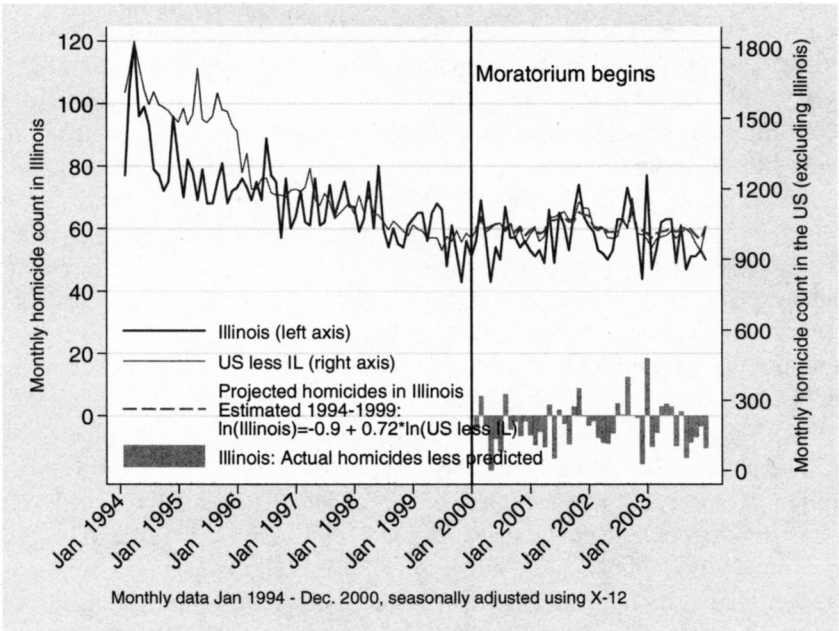
68. There are potentially further issues arising from the unreliability of county-level data. *See* Michael D. Maltz & Joseph Targonski, *A Note on the Use of County-Level UCR Data*, 18 J. QUANTITATIVE CRIMINOLOGY 297, 298 (2002); *see also* Ian Ayres & John J. Donohue, III, *Shooting Down the "More Guns, Less Crime" Hypothesis*, 55 STAN. L. REV. 1193 (2003).

69. Dale O. Cloninger & Roberto Marchesini, *Execution Moratoriums, Commutations and Deterrence: The Case of Illinois* (Econ. Working Paper Archive, Working Paper No. 0507002, 2005), *available at* http://econwpa.wustl.edu:80/eps/le/papers/0507/0507002.pdf (last visited Dec. 4, 2005).

70. The homicide rate is probably preferable to the homicide count, although we analyze the latter here to maintain continuity with Cloninger and Marchesini, noting that population growth is unlikely to have driven much of a gap between movements in homicide rates and levels over such a short time horizon.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 5. Homicides Before and After the Illinois Moratorium




the series for Illinois and the United States over the period from 1994 to 1999 had continued over the next four years. In an event study, one compares the subsequent evolution of the variable of interest with this projection, and the bars show the gap between Illinois homicides and the projected number of homicides.

It should be clear from inspecting the graph that the relationship between homicides in Illinois and the rest of the country is roughly unchanged since the moratorium. If anything, the bars appear persistently negative, suggesting that Illinois experienced about three fewer homicides per month than one would have expected based upon its previous relationship with the rest of the country.[71]

Finally, Cloninger and Marchesini[72] applied similar methods to analyze another quasi-experiment: a period from 1996 to early 1997 in which executions ground to a halt until the Texas Court of Criminal Appeals ruled on

71. The post-moratorium decline in homicides is actually statistically significant, although given how sparse this specification is, we do not want to overstate this point. Over the full sample, we estimated:

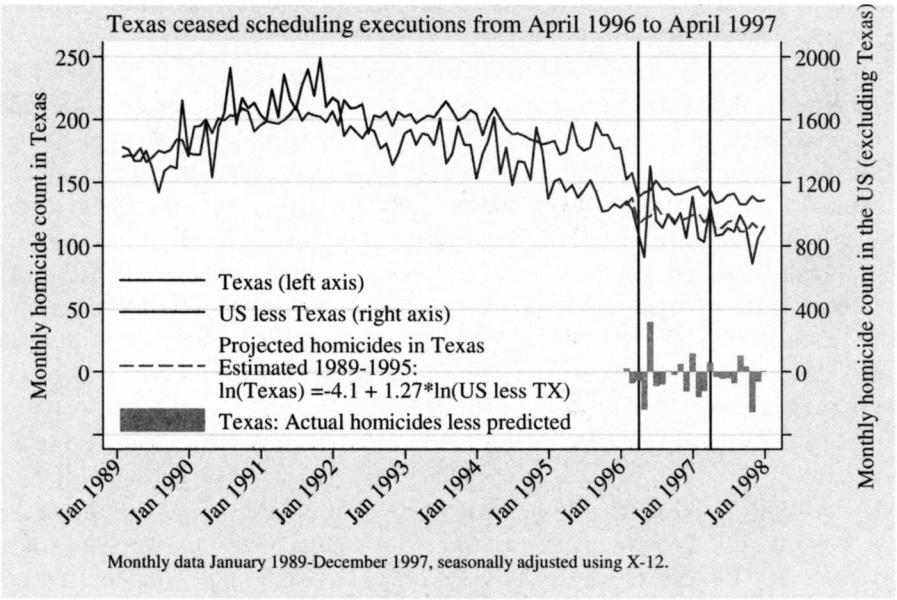
$$ln(Illinois\ homicides)_t = \underset{(0.90)}{-1.04} + \underset{(0.13)}{0.74}*ln(US\text{-}IL\ homicides)_t - \underset{(0.03)}{0.06}\ Post\ 2000_t$$

where we report Newey-West standard errors to account for up to sixth-order autocorrelation. Using this full-sample estimate, murders were six percent lower during the moratorium.

72. Dale O. Cloninger & Roberto Marchesini, *Execution and Deterrence: A Quasi-Controlled Group Experiment*, 33 APPLIED ECON. 569 (2001).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 6. Homicides Before and After the Texas Stay




the legality of new legislation limiting state habeas corpus petitions.[73] Figure 6 shows our reanalysis of these data, focusing again on the number of homicides (rather than their rate of change), and once again we find no evidence of an abnormal rise (or fall) in Texas homicides during this period.

## V. INSTRUMENTAL VARIABLES ESTIMATES

The studies that we have examined so far simply highlight the *correlation* between execution and homicide rates while controlling for other factors. Although their authors typically have premised their analyses on the assumption that changes in execution policy *cause* changes in crime rates, there are other possibilities that might explain this correlation.

First, a "get tough on crime" attitude might lead to longer jail sentences,[74] increased use of life without parole,[75] harsher prison conditions,[76] as well as increased use of the death penalty. It might be that criminals are responding to these other changes in deterrence, and given that the existing estimates contain no (or inadequate) controls for these factors, they may be driving the correlation between homicides and executions. There are good reasons to be

73. *Ex parte* Davis, 947 S.W.2d 216 (Tex. Crim. App. 1996); *see also* Kate Thomas, *Texas Executions Take a Sabbatical*, NAT'L L.J., Aug. 26, 1996, at A8.

74. Passell & Taylor, *supra* note 29.

75. Fagan Statement, *supra* note 24.

76. Katz, Levitt & Shustorovich, *supra* note 10.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

concerned by this possibility, as very few criminals are potentially affected by the death penalty, while many inmates are likely to be affected by these broader changes in deterrence policies.

Second, public support for the death penalty may be a function of current crime rates, and as such, causation may run from homicides to executions. This could go in either direction: a high homicide rate might make the public frustrated enough to increase use of the death penalty; alternatively if a higher homicide rate leads to more executions (for a fixed execution rate), this might undermine support for the death penalty.

Finally, and more generally, there may be a large number of unobservable factors changing through time that are correlated with death penalty usage and that also affect homicide. In the absence of a comprehensive set of control variables, these unobserveable factors might be driving a spurious correlation between executions and the death penalty.

The only way to resolve clearly the issue of causation would be to run an experiment in which we would implement the death penalty more (or less) vigorously in some states and in some years than in others, and then compare the outcomes. Of course experimenting with capital punishment laws in this manner does not seem particularly feasible, but one might imagine quasi-experiments: perhaps there are some factors that might change death penalty policy but do not otherwise affect homicide rates. These factors are called "instrumental variables" and can be used to analyze the effects of such quasi-experiments. Naturally, the credibility of such an exercise depends critically on whether the instrumental variables really do generate useful experiments that change the death penalty rates but do not affect other factors.

Given the promise that the instrumental variables approach holds for resolving questions of causality, it is not surprising that Sunstein and Vermeule seem to repose the greatest confidence in a recent application of this method by Dezhbakhsh, Rubin, and Shepherd.

To briefly review that study, Dezhbakhsh, Rubin, and Shepherd analyze county data from 1977 to 1996, using data provided by John Lott and David Mustard.[77] Following Ehrlich, their paper posits that homicide rates are a function of three primary deterrence variables: homicide arrest rates, the probability of a death sentence conditional on arrest, and the probability of execution conditional on a death sentence. Lott and Mustard's data allow the authors to account for a range of other factors, so they also add controls for the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; and NRA membership. While they have county-level data for their dependent variable (the homicide rate), the

77. *See* Dezhbakhsh, Rubin & Shepherd, *supra* note 11.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

homicide arrest rate, and the control variables, they only have state-level data on the variables of interest (the "deterrence" explanatory variables). Thus, to be somewhat more specific, their main regression is:[78]

$$\frac{Murders_{c,s,t}}{(Population_{c,s,t}/100000)} = \beta_1 \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} + \beta_2 \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} + \beta_3 \frac{Executions_{s,t}}{DeathSentences_{s,t-6}}$$
$$+\gamma_1 \frac{Assaults_{c,s,t}}{Population_{c,s,t}} + \gamma_2 \frac{Robberies_{c,s,t}}{Population_{c,s,t}} + \gamma_3 CountyDemographics_{c,s,t} + \gamma_4 CountyEconomy_{c,s,t}$$
$$+\gamma_5 \frac{NRAmembers_{s,t}}{Population_{s,t}} + \sum_c CountyEffects_c + \sum_t TimeEffects_t + \eta_{s,t} + \varepsilon_{c,s,t}$$

where $c$ denotes a county, $s$ denotes the state that the county is in, and $t$ denotes a year. The main coefficients of interest in this equation are the $\beta$s, and specifically, they interpret $\beta_3$ as representing the effects of executions on the homicide rate.

Following Ehrlich's discussion of the difficulty of making causal inferences in this setting,[79] the authors are sensitive to concerns that their deterrence measures might be driven by other factors, which leads them to run instrumental variables regressions. Essentially, this requires them to look for changes in deterrence caused by factors unrelated to either prevailing homicide rates or the unobserved determinants of crime (like sentence length). They believe that they have identified several such variables: state-level police payroll, judicial expenditures, Republican vote shares in presidential elections, and prison admissions. (Somewhat surprisingly the police, judicial, and prison variables are statewide aggregates, rather than per capita numbers, and the authors choose not to adjust either police payrolls or judicial expenditures to account for inflation.) As such, these variables (plus controls) are included in first-stage regressions for each of the deterrence variables. That is, they only analyze movements in the deterrence variables that are correlated with state police payrolls, judicial expenditures, vote shares, or prison admissions.

Dezhbakhsh, Rubin, and Shepherd generously shared their data and code, and Joanna Shepherd assisted our efforts, enabling us to perfectly replicate all of their results, as shown below in Panel A of Table 7. (Their six main regressions, summarized in their Tables 3 and 4, differ slightly in how they proxy for the expectations of criminals regarding the deterrence variables.[80]) These results report the regression coefficients on the probability of homicide arrest, the probability of a death sentence conditional on arrest, and the probability of execution conditional on a death sentence. For continuity, we report the same standard errors (and as closely as possible the same specification) that the authors do, but will return to this issue below.

78. The authors actually report six main regressions, where each differs slightly in how it measures the deterrence variables and how it deals with observations in which a state had no murders or issued no death sentence. *Id.* This equation shows their preferred specification, Model 4.

79. *See* Ehrlich, *supra* note 3, at 414.

80. *See* Dezhbakhsh, Rubin & Shepherd, *supra* note 11, at 362-63 tbls.3 & 4.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Table 7: Estimating Effect of Executions on Murder Rates and Net Lives Saved: Testing the Sensitivity of the Dezhbakhsh, Rubin, and Shepherd (DRS) Estimates, 1977-1996

| | Dependent Variable: *Annual Homicides per 100,000 Residents*$_{c,t}$ | | | | | |
|---|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) | (6) |
| | **Panel A: Replication of DRS, Estimated Coefficients** | | | | | |
| Probability of Arrest | -4.04*** | -10.10*** | -3.33*** | -2.27*** | -4.42*** | -2.18*** |
| | (0.58) | (0.57) | (0.52) | (0.50) | (0.45) | (0.48) |
| Probability of Death Sentence Given Arrest | -21.80 | -42.41*** | -32.12** | -3.62 | -47.66*** | -10.76 |
| | (18.6) | (13.71) | (16.22) | (14.53) | (10.45) | (13.13) |
| Probability of Execution Given Death Sentence | -5.17*** | -2.89*** | -7.40*** | -2.71*** | -5.20*** | -4.78*** |
| | (0.81) | (0.46) | (0.72) | (0.62) | (0.27) | (0.56) |
| | **Panel B: Replication of DRS, Implied Life-Life Tradeoff[(a)]** | | | | | |
| Net Lives Saved | 36.1*** | 19.7*** | 52.0*** | 18.5*** | 36.3*** | 33.3*** |
| | (5.8) | (3.3) | (5.1) | (4.4) | (1.9) | (4.0) |
| | **Panel C: Allowing Only One Partisanship Variable** | | | | | |
| Net Lives Saved | -24.5*** | -53.8*** | -43.3*** | -17.7*** | -0.9 | -26.1*** |
| | (8.0) | (6.0) | (8.2) | (6.0) | (3.0) | (6.2) |
| | **Panel D: Dropping Texas** | | | | | |
| Net Lives Saved | -21.5*** | 33.7*** | 6.5 | -41.6*** | 32.5*** | -11.3* |
| | (7.6) | (4.4) | (7.9) | (5.6) | (2.1) | (5.9) |
| | **Panel E: Dropping California** | | | | | |
| Net Lives Saved | -26.1*** | 30.1*** | 33.3*** | -28.7*** | 17.8*** | 9.6*** |
| | (7.0) | (3.9) | (6.5) | (4.9) | (2.0) | (4.8) |

Notes: Panel A replicates the estimates of the impact of deterrence variables on murder rates, using the specification and county-level data from Dezhbakhsh, Rubin & Shepherd, *supra* note 11, at 362-63 tbls.3-4. Panel B converts these estimates into net lives saved per execution, showing a net savings of from eighteen to fifty-two lives per execution. Panel C runs the regression as described by Dezhbakhsh, Rubin, and Shepherd, collapsing the partisanship variables into a single instrumental variable indicating the percentage of the Republican vote in the last presidential election (instead of six variables—one for each election); this specification then predicts that each execution will cost between one and fifty-four lives. Panels D and E show highly variable estimates when Texas and California are dropped.

Population-weighted instrumental variables regressions are used. Endogenous independent variables are shown in panel A. Instruments include state-level police payroll, judicial expenditures, Republican vote shares, and prison admissions. Controls include the assault rate; the robbery rate; real per

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership; and county and year fixed effects. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively.

*(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996.

Given the prominence attached to the implied "life-life" tradeoffs, Panel B reports these estimates in terms of the net number of lives saved per execution (evaluated for the average executing state in 1996). Thus, Model 4 shows the basis of the estimate that eighteen lives are saved (on net) by each execution, as trumpeted by Sunstein and Vermeule.[81] Because the estimated coefficients appearing in Panel A are less easily interpreted, we will convert estimates into this "lives saved" metric and report them as such throughout.[82] The evidence collected in Panels A and B superficially appears to show robust and consistent support of the view that execution deters homicide.

Panels C through E show the sensitivity of Dezhbakhsh, Rubin, and Shepherd's results to a number of very simple specification checks, and the fragility of their conclusions becomes immediately evident. Panel C shows our initial attempt to replicate their results; this regression is actually the one described in the text of their paper, but not implemented in their code. One of their instrumental variables—that measuring partisan influence in the state—turned out to be particularly troubling. Specifically, they note that their set of instruments includes "partisan influence as measured by the Republican presidential candidate's percentage of the statewide vote in the most recent election. . . ."[83] The set of results in Panel C implements their model using Dezhbakhsh, Rubin, and Shepherd's instruments but including—as the text cited above suggests—a single variable that denotes the Republican vote share in that state in the most recent presidential election. This single change generates considerably different results from those reported in their paper, suggesting instead a large antideterrent effect. The signs are different, and the magnitudes are larger. Note that for Dezhbakhsh, Rubin, and Shepherd's preferred Model 4, this single change flips the sign of their original estimates:

81. *See* Sunstein & Vermeule, *supra* note 12, at 706.

82. We should note that this is the relevant tradeoff where the thought experiment involves a governor asking about the implications of whether to execute a prisoner on death row. For consideration of the Sunstein and Vermeule argument, the relevant margin is deciding whether to introduce and enforce the death penalty. Computing the life-life tradeoff for this thought experiment requires consideration of a second effect, mediated by changes in the probability of obtaining a death sentence. We follow Dezhbakhsh, Rubin, and Shepherd in reporting the results of the former, but we note that the qualitative conclusions one would draw from our analysis are largely unchanged when considering the latter.

83. Dezhbakhsh, Rubin & Shepherd, *supra* note 11, at 357.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

instead of saving eighteen lives, each execution leads to eighteen lives *lost*.

The ultimate resolution of this substantial discrepancy lay in the fact that Dezhbakhsh, Rubin, and Shepherd had controlled for "partisan influence" not with a single measure of the Republican vote in the most recent election, but by defining six different political variables reflecting the Republican vote shares in six different presidential elections.[84] To be clear, the diametrically opposed conclusions of Panels B and C reflect the fact that the regression in Panel C implicitly imposes a constant effect of the partisanship variable through time (resulting in a finding that the death penalty leads to a large increase in murders), while Panel B allows it to change (and even change signs) across election cycles (leading to a finding that the death penalty deters murders). Our point is not that one specification is preferable to the other. Indeed, sorting that out would be a difficult task. Rather, the point is to show the incredible sensitivity of Dezhbakhsh, Rubin, and Shepherd's results to how they code their instruments: using the methods described in the paper leads to very different results from those using the minor variation that they actually implemented.

Panels D and E show the sensitivity of these results to sample selection. We return to Dezhbakhsh, Rubin, and Shepherd's preferred specification, but in Panel D we drop Texas from the data; this change also leads to a wide range of estimated effects, with the estimated life-life tradeoff across the six specifications ranging from -42 to +34. In Panel E we drop California and this also dramatically affects the estimates, with estimates ranging from -29 to +30. Of course, both California and Texas are very interesting states, and we do not mean to suggest that they don't contain (substantial) useful information for establishing the deterrent effects of the death penalty. Rather, we mean to simply highlight the sensitivity of the results. Shepherd has also shown that the estimated deterrent or antideterrent effects in this regression vary dramatically across states, a fact that she interprets as reflecting some states not executing enough convicts to reach a threshold where deterrence applies.[85] What is not shown in Shepherd's article is that the same exercise also suggests large effects even in states that do not have capital punishment. Thus, an equally likely interpretation is that the differences across states also reflect different degrees

84. In other words, we had initially thought that for each year and each state, Dezhbakhsh, Rubin, and Shepherd were using a single continuous variable equal to the percentage of the Republican vote in the closest presidential election to that particular year. Instead, they had six different continuous variables so that the effect of voting Republican would be different for each of the six presidential elections between 1976 and 1996. This was accomplished by having a variable set equal to zero for all observations except 1995-1996, when it was set equal to the Republican vote share in that state in the 1996 election, another variable that is all zeroes but for 1991-1994 (when it was set equal to the Republican state vote share in the 1992 presidential election), and similar variables for the 1988 election (1987-1990), the 1984 election (1983-1986), the 1980 election (1979-1982), and the 1976 election (1977 and 1978).

85. Shepherd, *Deterrence Versus Brutalization*, *supra* note 46, at 225-26.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

of misspecification,[86] or simply noise.

In sum, given the sensitivity of these results to rather small and sometimes arbitrary changes, one has little reason to prefer the conclusion that the death penalty will save lives to the conclusion that scores will die as a result of each execution.

### A. *Problems with Invalid Instruments*

We now turn to evaluating in greater detail the instrumental variables procedure employed. Recall that the instrumental variables procedure yields valid results if the raw number of prison admissions, police payrolls, judicial expenditures, and the Republican presidential vote share in each state provide "experiments" which change the deterrence variables, but are not related in any other way to the homicide rate. If these variables are good instruments, then they should be correlated with the endogenous deterrence variables: the probability of arrest for murder, the probability of receiving a death sentence conditional on murder arrest, and the probability of execution given death sentence. It seems fairly clear that each of these instrumental variables will be correlated with crime rates; however, the credibility of this exercise depends vitally on whether the sole mediating links are changes in the murder arrest rate and application of the death penalty. This is a much tougher case to make. While these identifying assumptions are untestable in many applications, in this case there are a number of approaches we can take to examine their plausibility.

The top panel in Table 8 simply replicates Dezhbakhsh, Rubin, and Shepherd's main estimates (again showing the estimates as the number of lives that will be saved per execution). Recall that if the identifying assumptions are true, variation in the instruments should not affect the homicide rate, except through its influence on executions. In Panel B, we restrict the sample to those observations occurring when the state did not have the death penalty.[87] As such, there is no way for changes in the instruments to yield useful experiments changing the execution rate for this subsample. Thus, Panel B can be thought of as depicting the "effect" of "exogenously" executing prisoners in states that have no death penalty (an obvious oxymoron).[88] The number of state-year observations in which there is no death penalty is rather limited—about one-fifth of the sample—and hence the coefficients are not quite as precisely

86. That is, it may be that the relationship between the endogenous deterrence variables and the exogenous instrumental variables varies across states, rather than that the relationship between homicide and deterrence varies.

87. To generate our Panel B estimates, we first run the first-stage regression. Then, we drop all observations for which the state is operating under a legal death penalty regime and run the second-stage regression on this subset of the data.

88. Dezhbakhsh, Rubin, and Shepherd's instruments would pass this test of validity if there was no correlation between the instruments and homicide rates in states without the death penalty. Panel B of Table 8 shows that this is not the case.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

estimated. Nonetheless, the effects are positive in five of the six columns and tend to be larger than the effects estimated for the full sample (Panel A). The most obvious interpretation is that the instruments (or their correlates) affect homicide rates directly—through channels other than death row—and hence that the assumption required for these instrumental-variables estimates to be valid is violated.

Table 8: Estimating Net Lives Saved per Execution: Exploring the Validity of the Dezhbakhsh, Rubin, and Shepherd (DRS) Instrumental Variables, 1977-1996

| **Dependent Variable:** ***Annual Homicides per 100,000 Residents***$_{c,t}$ | (1) | (2) | (3) | (4) | (5) | (6) |
|---|---|---|---|---|---|---|
| | **Panel A: Replication of DRS, Implied Life-Life Tradeoff**[(a)] | | | | | |
| Net Lives Saved | 36.05*** | 19.70*** | 51.99*** | 18.45*** | 36.27*** | 33.26*** |
| | (5.83) | (3.32) | (5.14) | (4.43) | (1.94) | (4.01) |
| | **Panel B: "Effects" in State-Years in Which There Is No Death Penalty** | | | | | |
| Net Lives Saved | 74.00** | 71.48*** | 163.87*** | -70.06*** | 103.01*** | 108.07*** |
| | (29.62) | (8.80) | (21.64) | (15.40) | (5.34) | (14.98) |
| | **Panel C: Restricting the Instrumental Variables to Police Payrolls, Judicial Expenditure, and Prison Admission**[(b)] | | | | | |
| Net Lives Saved | -85.57*** | -36.81 | -71.95*** | -52.30*** | -23.00*** | -85.67*** |
| | (13.72) | (28.30) | (14.91) | (9.15) | (8.14) | (13.62) |
| | **Panel D: Restricting the Instruments to the Republican Vote Share**[(c)] | | | | | |
| Net Lives Saved | 429.43*** | 81.98*** | 286.45*** | 288.76*** | 53.06*** | 242.29*** |
| | (21.16) | (4.56) | (11.06) | (15.66) | (2.24) | (9.33) |

Notes: Panel A replicates Panel B of Table 7, showing the DRS estimates of the number of net lives saved per execution. Specification and data are from Dezhbakhsh, Rubin & Shepherd, *supra* note 11, at tbls.3-4. For further details, see notes to Table 7. Panel B tests the DRS assumption that their instruments only affect homicides through their effect on executions by showing that the predicted number of executions are highly correlated with murder rates even in states with no executions. Panel C shows that if one does not use the Republican vote share as an instrument, the death penalty leads to more murders, while Panel D shows that using only the Republican vote share variables as instruments, the apparent beneficial effect of the death penalty skyrockets.

*(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996.

*(b)* Panel C regression includes the Republican vote share variables as controls, but not as instruments.

*(c)* Panel D regression includes police payrolls, judicial expenditure, and prison admissions as controls, but not as instruments.

There exists an alternative way to test the validity of instrumental

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

variables, based on Jerry Hausman's overidentification test.[89] The logic of an overidentification test is that if the "experiments" in deterrence generated by the instrumental variables are valid, then the results from one set of experiments should be similar to those from another set of experiments. The specific system of equations offered by Dezhbakhsh, Rubin, and Shepherd cannot be estimated unless they have three instruments (because they need at least one exogenous instrument for each of their three endogenous variables); they actually employ four separate instruments (or nine, if the six Republican vote-share variables are counted separately). Thus, an overidentification test essentially suggests that if these instruments are all valid, then the coefficients should remain stable as we drop some subset of the instruments. Shepherd discusses these regressions, stating that "tests for overidentification indicate that the model is correctly specified and employs valid instruments."[90] We subjected these models to a battery of overidentification tests and could not find any evidence consistent with this claim. For instance, Panel C shows what happens when the partisanship variables are no longer regarded as instruments.[91] We see that the "experiments" generated by the combined forces of police payrolls, judicial expenditures, and prison admissions suggest that more executions lead to substantially *more* homicides. Panel D shows the complementary set of regressions: the six partisanship variables are retained as instruments, but police payrolls, judicial expenditures, and prison admissions are included as control variables. The variation induced by these variables yields dramatically different and implausibly large estimates of the deterrent effect of the death penalty.

The massive change in these coefficients suggests that at least some of these instrumental variables are not valid instruments. The large deterrent effect noted in their baseline regressions appears to be driven entirely by the partisan variables. As an aside, recall that Dezhbakhsh, Rubin, and Shepherd received their county data from John Lott, who had created the dataset to examine the impact of laws affording the right to carry concealed handguns. Like Lott, Dezhbakhsh, Rubin, and Shepherd use the exact same Republican vote-share variables as instruments in their analysis. In so doing, Lott was implicitly assuming that this political variable was influencing homicide only through its impact on arrest rates and the likelihood of adoption of a right-to-carry concealed handgun law. But in using the same Lott instruments, Dezhbakhsh, Rubin, and Shepherd assume that the political variables only influence crime rates through their effect on murder arrests, death sentences, and execution. Thus, it seems difficult to reconcile the competing assumptions made by these

89. *See* Jerry A. Hausman, *Specification Tests in Econometrics*, 46 ECONOMETRICA 1251 (1978).

90. Shepherd, *Deterrence Versus Brutalization*, *supra* note 46, at 227.

91. That is, we include the partisanship variables as control variables—in both first- and second-stage regressions.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

two sets of authors about how this political variable influences crime in a state.[92] In fact, Shepherd has used three of the four Dezhbakhsh, Rubin, and Shepherd instruments—police expenditure, judicial expenditure, and percentage voting Republican in the last presidential election—as instruments in analyzing the deterrent impacts of three other legislative measures: California's strike-based sentencing scheme on crime,[93] truth-in-sentencing legislation,[94] and sentencing guidelines.[95] The use of the same instruments in multiple studies underscores that the requirements for valid instrumentation of the death penalty must be violated if these instruments are influencing crime through these other avenues unrelated to execution.

An additional way to test whether variation in these instruments causes (or reflects) changes in crime markets not mediated by the death penalty (thus invalidating the crucial identifying assumption) is to test whether the variation in executions generated by them is correlated with other crimes for which the death penalty does not apply. We have run these separate regressions using each of the FBI index crimes as individual dependent variables, but otherwise applying the Dezhbakhsh, Rubin, and Shepherd specification.[96] The results are not encouraging for Dezhbakhsh, Rubin, and Shepherd, as they suggest that executions cause more rape, assault, burglary, and larceny, and less auto theft and homicide; the effects on robbery are inconclusive. In terms of statistical significance, the relationship between the homicide and execution rates is typically less reliable (statistically significant) than that between the execution rate and rape, aggravated assault, burglary, and larceny.[97]

---

92. As a further aside, note that Rubin and Dezhbakhsh rerun Lott's analysis, applying these same variables as instruments for concealed handgun laws, referring to this method as "more appropriate." Paul H. Rubin & Hashem Dezhbakhsh, *The Effect of Concealed Handgun Laws on Crime: Beyond the Dummy Variables*, 23 INT'L REV. L. & ECON. 199, 206 n.11 (2003).

93. Joanna M. Shepherd, *Fear of the First Strike: The Full Deterrent Effect of California's Two- and Three-Strikes Legislation*, 31 J. LEGAL STUD. 159 (2002).

94. Joanna M. Shepherd, *Police, Prosecutors, Criminals, and Determinate Sentencing: The Truth About Truth-in-Sentencing Laws*, 45 J.L. & ECON. 509 (2002).

95. Joanna M. Shepherd, *Are Criminals Like Us? Risk Attitudes, Sentencing Guidelines, and Increased Crime* (Emory Law & Econ. Research Paper No. 04-03, 2004), *available at* http://ssrn.com/abstract=370421 (last visited Dec. 4, 2005).

96. For obvious reasons, we need to drop aggravated assault and robbery as controls when either is the dependent variable; for other index crimes and in all other respects, we leave their specification unchanged.

97. Note that Dezhbakhsh, Rubin, and Shepherd discuss this approach directly in their paper:

> We also repeat the analysis, using as our dependent variable six other crimes: aggravated assault, robbery, rape, burglary, larceny and auto theft. If executions were found to deter other crimes besides murder, it may be the case that some other omitted variable that is correlated with the number of executions is causing crime to drop across the board. However we find no evidence of this. Of the thirty-six models that we estimate (six crimes and six models per crime), only six exhibit a negative correlation between crime and the number of executions. These cases are spread across crimes with no consistency as to which crime decreases with executions.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Given the apparent problems with these instrumental-variables estimates, it seems reasonable to try to figure out what is going on and to see whether the estimates are consistent with their theory. Specifically, Dezhbakhsh, Rubin, and Shepherd provide a theoretical rationale for their instruments:

> Police and judicial-legal expenditure . . . represent marginal costs of enforcement. More expenditure should increase the productivity of law enforcement or increase the probabilities of arrest, and of conviction, given arrest. Partisan influence is used to capture any political pressure to "get tough" with criminals, a message popular with Republican candidates. . . . Prison admission is a proxy for the existing burden on the justice system; the burden may affect judicial outcomes.[98]

Table 9 reports the Dezhbakhsh, Rubin, and Shepherd first-stage regressions—always a useful diagnostic, but something not shown in their paper. For brevity, we simply show the coefficients from their preferred specification (see Model 4 in Table 8).

Having estimated the first-stage regression, we can compute the (reduced-form) effects of a change in each of the instrumental variables on the homicide rate. This value is shown in the final column, which comes from multiplying the coefficient in each column by the coefficient of the relevant instrument in the second-stage regression. Note that contrary to their theorizing, increases in police spending and judicial spending are associated with a *higher* murder rate. Moreover, the coefficients on the Republican share of the vote in the six individual elections—which we saw in Panel C of Table 7 to have such a powerful effect on the deterrence estimates—change substantially from election to election. That is, the effect on deterrence policy of having more Republican voters bounces back and forth across various elections, again counter to the theoretical rationale that Republican majorities would be tougher on crime. Moreover, these estimates bounce around in a particularly counterintuitive manner: increased voting for Reagan in 1980 was associated with a deterrent effect, while the effects of Reagan in 1984 were equal and opposite; increased voting for Bush in 1988 was associated with an antideterrent effect, while states voting strongly for Bush in 1992 had the opposite result.

---

Dezhbakhsh, Rubin & Shepherd, *supra* note 11, at 365 n.21. That is, while they claim that six of thirty-six estimates showed a significant pseudo-deterrent effect and were spread across crimes with no consistency, we found six of six estimates for auto theft and two of six robbery estimates yielded significant pseudo-deterrent effects. Moreover, they neglected to mention that all six rape estimates, all six assault estimates, four of six robbery estimates, all six burglary estimates, and all six larceny estimates yielded a statistically significant pseudo-antideterrent effect. Both the pseudo-deterrent and pseudo-antideterrent estimates suggest that the instrumental variables are correlated with other developments in crime markets, which would render them invalid instruments for Dezhbakhsh, Rubin, and Shepherd's analysis.

98. *Id.* at 357.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Table 9. Do the Dezhbakhsh, Rubin, and Shepherd Instruments Have the Predicted Effects on Endogenous Deterrence Variables in Their First-Stage Regressions? (1977-1996)

| | Dependent variable | | | Net Effect on Homicide Rate[(a)] |
|---|---|---|---|---|
| | Probability of Arrest | Probability of Death Sentence Given Arrest | Probability of Execution Given Death Sentence | |
| | (1) | (2) | (3) | (4) |
| Police Spending | 0.03<br>(0.023) | -0.002***<br>(0.000) | -0.05***<br>(0.004) | 0.08 |
| Judicial Spending | -0.22***<br>(0.034) | 0.01***<br>(0.001) | -0.04***<br>(0.006) | 0.58 |
| Prison Admission | 0.01***<br>(0.002) | -0.0001***<br>(0.000) | 0.004***<br>(0.000) | -0.04 |
| 1976 * Republican Vote Share (Ford) | -0.66**<br>(0.311) | 0.03<br>(0.083) | 0.49***<br>(0.053) | 0.08 |
| 1980 * Republican Vote Share (Reagan I) | 0.16<br>(0.202) | 0.004<br>(0.004) | 0.02<br>(0.036) | -0.45 |
| 1984 * Republican Vote Share (Reagan II) | -0.64***<br>(0.196) | 0.04***<br>(0.004) | 0.29***<br>(0.035) | 0.54 |
| 1988 * Republican Vote Share (Bush I) | -0.25<br>(0.216) | 0.06***<br>(0.004) | -0.03<br>(0.038) | 0.41 |
| 1992 * Republican Vote Share (Bush II) | -0.04<br>(0.215) | 0.05***<br>(0.004) | 0.14***<br>(0.039) | -0.45 |
| 1996 * Republican Vote Share (Dole) | -0.82***<br>(0.212) | 0.01**<br>(0.004) | 0.96***<br>(0.040) | -0.77 |
| *N* | 48,070 | 51,143 | 57,637 | |
| | **Second Stage** | | | |
| Coefficients | -2.27***<br>(0.50) | -3.62<br>(14.53) | -2.71***<br>(0.62) | |

Notes: Using the data, source, and specification from *supra* note 11, at 363 tbl.4, Model 4, this table illustrates the impact of the Dezhbakhsh, Rubin, and Shepherd instrumental variables on the three endogenous deterrent variables (Columns 1 through 3) and on homicide rates (Column 4). Contrary to their articulated rationale for these instruments, police spending, judicial spending, and Republican vote share in 1976, 1984, and 1988 correlate with higher murder rates. The police and judicial spending variables are expressed in billions of dollars. Coefficients on prison admissions and vote share variables have been multiplied by 1000 and 100, respectively. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5% or 10%, respectively.

*(a)* Column 4 is a simple calculation reflecting the direct effect of a change in each independent variable on the homicide rate, as mediated through each of the endogenous variables. That is, Column 4 is the sum of the first stage coefficients multiplied by the corresponding second-stage coefficients (listed in the bottom row).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

B. *Problems with Statistical Significance*

At this point we have shown that the Dezhbakhsh, Rubin, and Shepherd results are highly sensitive in a range of dimensions and that both the sign and magnitude of the estimates vary wildly. From a statistical standpoint, what is most surprising is that each estimate—while often dramatically different from other estimates—also appears to be estimated quite precisely. That is, the standard errors on all of these results are quite small, and the statistical significance of the results quite substantial. This invites the inference that the statistical significance of these results is considerably overstated.

To better illustrate that the Dezhbakhsh, Rubin, and Shepherd model is not yielding reliable estimates of the effect of an additional execution on murder, we ran the following experiment using their preferred specification as our base model. We took the time series of the independent variables for each county and matched it to the time series of the homicide rate for a random county. Thus, the independent variables are, *by construction*, unrelated to the dependent variables (conditional on year fixed effects).[99] We then ran the Dezhbakhsh, Rubin, and Shepherd regression (using their preferred Model 4) and collected the relevant coefficients. We repeated this process 1000 times and, hence, generated the distribution of the estimated effects across 1000 instances *in which there is no true underlying relationship*.

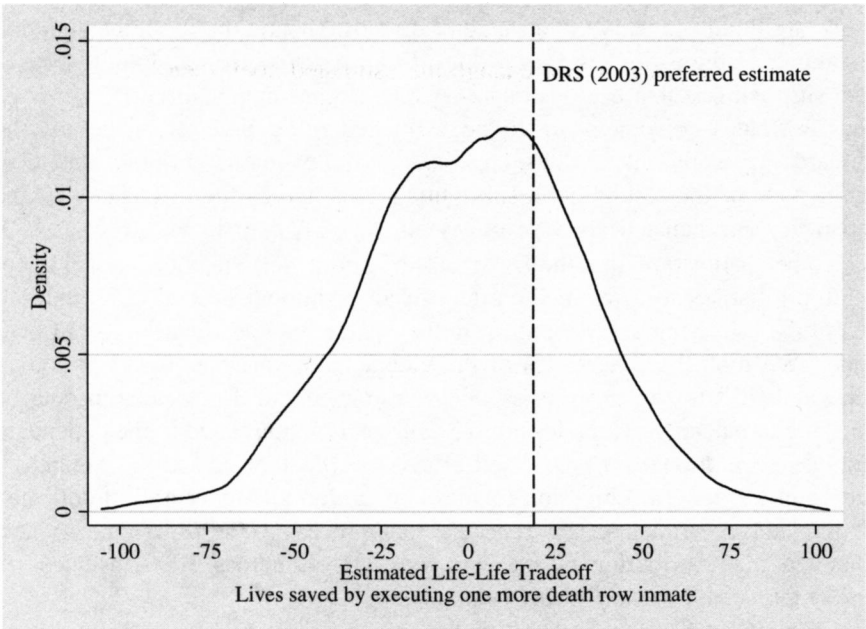
Figure 7 depicts the probability density function of these estimates, and highlights where the Dezhbakhsh, Rubin, and Shepherd central estimate falls in this distribution. In these experiments, the *uncorrelated* data yielded coefficients at least as large as their estimate 30% of the time, and it yielded coefficients with an absolute value at least this big 56% of the time. That is, this exercise suggests that even if there is absolutely no relationship between the death penalty and murder, there is a substantial probability that the Dezhbakhsh, Rubin, and Shepherd model will, by chance, generate results suggesting there is a large and statistically significant effect. By contrast, the $t$-statistic that they reported ($t = 4.4$) suggests that under the same null, estimates as large as theirs occur less than 0.001% of the time.

It is now well known that there are at least two problems with the standard errors that Dezhbakhsh, Rubin, and Shepherd report. First, the data are highly autocorrelated, which leads to substantial underestimates of standard errors (and thus overestimation of precision). To explain briefly, this year's homicide and execution rates often closely resemble last year's, and so to treat the two observations as independent experiments would understate uncertainty about the relationship between the two. Second, despite the fact that the dependent

99. Formally, this is a randomization test, using block randomization. *See* BRYAN F.J. MANLY, RANDOMIZATION, BOOTSTRAP AND MONTE CARLO METHODS IN BIOLOGY (2d ed. 1997). We also obtained qualitatively similar results when randomizing the residuals instead of the independent variable, as suggested in Peter E. Kennedy, *Randomization Tests in Econometrics*, 13 J. BUS. ECON. STAT. 85 (1995).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 7. Dezhbakhsh, Rubin, and Shepherd (DRS) Distribution of Estimates Under the Null of No Deterrent Effect




variable is measured at the county level, the independent variables of interest in these regressions are measured at the state level. If there are state-specific shocks through time—reflecting factors like unmodelled changes in state policies, changes in state criminal markets, and the like—then this again will lead standard OLS methods to overstate their precision. The intuition is that by disaggregating to the county level, one might gain a false sense of security that each county provides an independent experiment, when counties within a state are likely to be subject to correlated shocks.

Both of these facts are already well understood in the empirical literature,[100] and indeed, Eric Helland and Alex Tabarrok have made these points quite explicitly regarding Lott and Mustard's investigation of the right-to-carry concealed handgun laws.[101] The exercise depicted in Figure 7 provides one way of assessing statistical significance in light of autocorrelation, but it does not further take account of the correlation across counties within the same

100. *See* Brent R. Moulton, *An Illustration of a Pitfall in Estimating the Effects of Aggregate Variables in Micro Units*, 72 REV. ECON. & STAT. 334 (1990) (on clustering); Bertrand, Duflo & Mullainathan, *supra* note 45 (on autocorrelation).

101. *See* Eric Helland & Alex Tabarrok, *Using Placebo Laws To Test "More Guns, Less Crime,"* 4 ADVANCES ECON. ANALYSIS & POL'Y 1 (2004). Given that the Dezhbakhsh, Rubin, and Shepherd data are a near-identical version of the Lott and Mustard data and that the structure of their estimating equations is similar, it seems natural to suspect that the same issues arise.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

state. As such, we followed Marianne Bertrand, Esther Duflo, and Sendhil Mullainathan[102] and reestimated the Dezhbakhsh, Rubin, and Shepherd models, correcting the standard error estimates to take account of correlation both across counties within states and within states and counties through time. These adjustments obviously do not change the estimated coefficients, and thus the estimated life-life tradeoff for Dezhbakhsh, Rubin, and Shepherd's preferred Model 4 remains at 18.5. However clustering by county leads the standard error to rise from 7.1 to 37.6, and clustering by state leads the estimated standard error to rise further to 51.3; block-bootstrap standard errors yielded similar estimates. That is, the 95% confidence interval around their central estimate ranges from the suggestion that each execution causes 82 more murders to each execution saving 119 lives.

Some of these same problems with statistical inference recur in Paul Zimmerman's 2004 study.[103] While several aspects of his approach are similar to those of Dezhbakhsh, Rubin, and Shepherd, there are two important differences: he exploits state-level data (over the sample from 1978 to 1997), and he uses a different set of instrumental variables. Specifically, Zimmerman argues that characteristics of homicides affect the resolve of the authorities to apply the death penalty, and so he employs variables describing homicides in the current and previous year as his instrumental variables.[104] Analyzing the subset of variation in executions that is correlated with his instruments, Zimmerman's preferred estimate suggests that each execution saves 19 lives, and his reported 95% confidence interval ranges from 7 to 31 lives. While we cannot test his identifying assumption (although we may be skeptical about it), we can test whether his results reflect chance, or a more fundamental correlation. Using Zimmerman's data, we reran his regressions so as to correct the standard error for clustering within states through time; we also estimated block-bootstrap standard errors. These exercises suggested that the true 95% confidence interval runs from each execution causing 23 homicides to each preventing 54 homicides.

---

102. *See* Bertrand, Duflo & Mullainathan, *supra* note 45, at 249.

103. *See* Zimmerman, *supra* note 11, at 163.

104. Thus, Zimmerman's instruments include: an indicator for whether an offender was released from death row in the previous year; an indicator of whether there was a botched execution in the previous year; and both contemporaneous and once-lagged values of the proportion of murders committed by strangers, by nonwhites, and under nonfelony-related circumstances. Of course if certain classes of homicides simply vary more than others, their share in the total will be directly correlated with the homicide rate, invalidating the use of these variables as instruments.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

## VI. A PARTIAL RECONCILIATION: LACK OF STATISTICAL POWER AND REPORTING BIAS

Our analysis of the effects of judicial and legislative experiments yielded quite inconclusive results. Neither adoption nor abolition of the death penalty could reliably be causally linked to homicide rates. Our reanalysis of Katz, Levitt, and Shustorovich's data shows that even with the largest samples analyzed in the literature, it is difficult to isolate any robust correlation between homicide rates and changes in the intensity with which the death penalty applies. That this is true even when analyzing data from fifty states over the period from 1934 through 2000 is perhaps surprising, although this could be taken to buttress the view that the true effect is reasonably close to zero.

A set of studies has analyzed execution data over much shorter, more recent (post-moratorium) time periods and purports to find reliable relationships between executions and homicides.[105] While the published estimates in this set of studies point to a deterrent effect, our reanalysis shows that small changes in specifications, samples, or functional form can dramatically change the results. Indeed, several of the more expansive specifications point to an antideterrent effect of the death penalty. What then is to be made of this highly volatile set of estimates? Unless one has a particularly strong prior belief about the "correct specification" (and we do not believe that economic or econometric theory are sufficiently well developed here that one would be warranted), one cannot confidently conclude that the evidence points to either deterrent or antideterrent effects. The difficulty in drawing strong conclusions is not simply one of the statistical (in)significance of the estimates: even when coefficient estimates are plagued by wide confidence intervals, they are still informative as to the "most likely" effects of the death penalty; yet, the "most likely" effect varies too widely across specifications to provide much guidance.

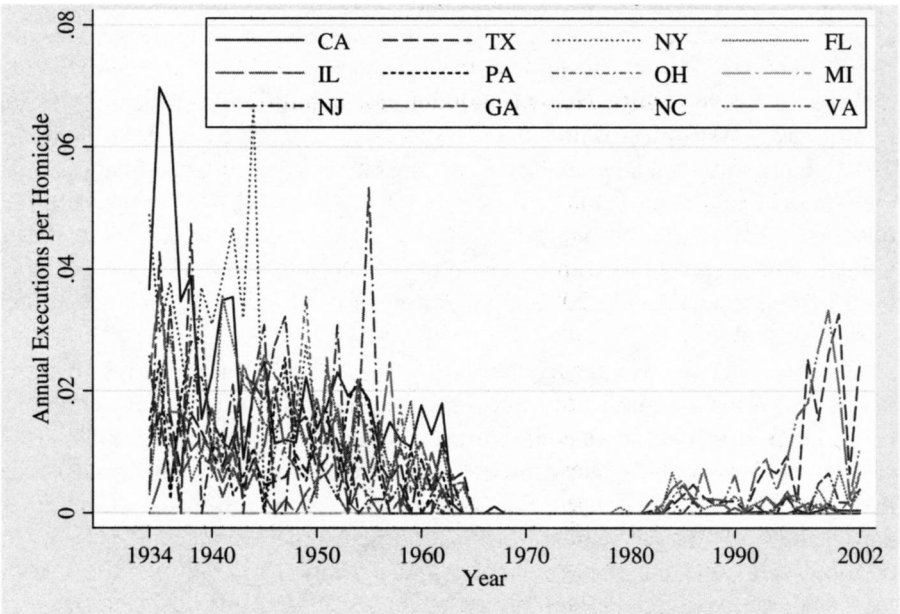
Moreover, it seems unlikely that *any* study based only on recent U.S. data can find a reliable link between homicide and execution rates. Figure 8 illustrates the difficulty facing researchers fixated on recent data, showing execution rates from 1934 to 2002 for the twelve largest states (accounting for around 60% of the U.S. population). The clear message is that there has been very little variation in execution rates since 1960 with which to reliably estimate any effects. Among these twelve states, there were very few executions between the early 1960s and the mid-1990s, and since then, only Texas and Illinois provide much variation. Moreover, the difficulty of finding reliable estimates is exacerbated by the fact that homicide rates typically show tremendous volatility both year to year and decade to decade.

The difficulty of discerning reliable *correlations* between execution policy

105. *See* Dezhbakhsh, Rubin & Shepherd, *supra* note 11; Mocan & Gittings, *supra* note 11; Zimmerman, *supra* note 11.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 8. Execution Risk by State: Twelve Largest States




and homicides becomes even sharper when attempting to use instrumental variables methods to isolate *causal* effects because these methods focus on only the subset of the variation in executions that is deemed "exogenous." For most plausible sets of instrumental variables, only a small number of executions can be thought of as yielding the sorts of "experiments" that this method requires, so it is commensurately more difficult for these estimates to yield robust and significant estimates. Indeed, in the previous Part we saw that realistic approaches to measuring the standard errors in existing instrumental-variables estimates pointed to an extremely large degree of uncertainty about their true effects.

All told, estimates in the existing literature appear to be quite fragile in light of small changes to specification, sample, or functional form. Estimates from a variety of approaches yielded different signs and vastly different magnitudes, a pattern of results that is at least partly reconciled by more appropriate treatments of standard errors suggesting that much of this is natural sampling error. All of this said, Sunstein and Vermeule's reading of the literature led them to see a persistent pattern of robust deterrent effects reported in these same papers. What explains this disjunction? One possibility is simply that the published estimates are a nonrepresentative sample of the wider universe of estimates that we have sought to present. If this were true, then even a careful reading of published results would suffer from a simple sample selection bias.

"Reporting bias" refers to the possibility that published results are an

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

unrepresentative sample. There are several reasons why this might occur. The "file drawer problem" refers to the tendency of researchers not to report on approaches that "didn't work out," in the sense of not yielding statistically significant estimates. Alternatively, "publication bias" arises when journals only publish estimates that meet standard tests of statistical significance. "Data mining" or "specification search" may also occur if career-driven or ideologically motivated researchers face incentives to report specifications that yield statistically significant evidence or estimates in favor of their preferred position. That said, it is worth emphasizing that reporting bias may occur without any of the authors being aware of it: they might simply want to report useful findings, and evidence falsifying a null hypothesis is typically regarded as more valuable.

Fortunately, we can test for reporting bias.[106] The intuition for this test begins by noting that different approaches to estimating the effect of executions on the homicide rate should yield estimates that are somewhat similar. That said, some approaches yield estimates with small standard errors, and hence these should be tightly clustered around the same estimate, while other approaches yield larger standard errors, and hence the estimated effects might be more variable. Thus, there is likely to be a relationship between the size of the standard error and the variability of the estimates, but *on average there should be no relationship between the standard error and the estimated effect.* By implication, if there is a correlation between the size of the estimate and its standard error, this finding suggests that reported estimates comprise an unrepresentative sample. One simple possibility might be that researchers are particularly likely to report statistically significant results, and thus they only report on estimates that have large standard errors if the estimated effect is also large. If this were true, we would be particularly likely to observe estimates that are at least twice as large as the standard error, and therefore coefficient estimates would be positively correlated with the standard error.

In Figures 9 and 10, we compile each of the reported estimates of the average number of homicides prevented per execution in recent state or county panel-data studies, as well as the reported standard errors. To ensure that this sample is representative of the literature, we included all of the reported panel data estimates from the various papers cited by Sunstein and Vermeule, a list that coincides with Shepherd's congressional testimony.[107]

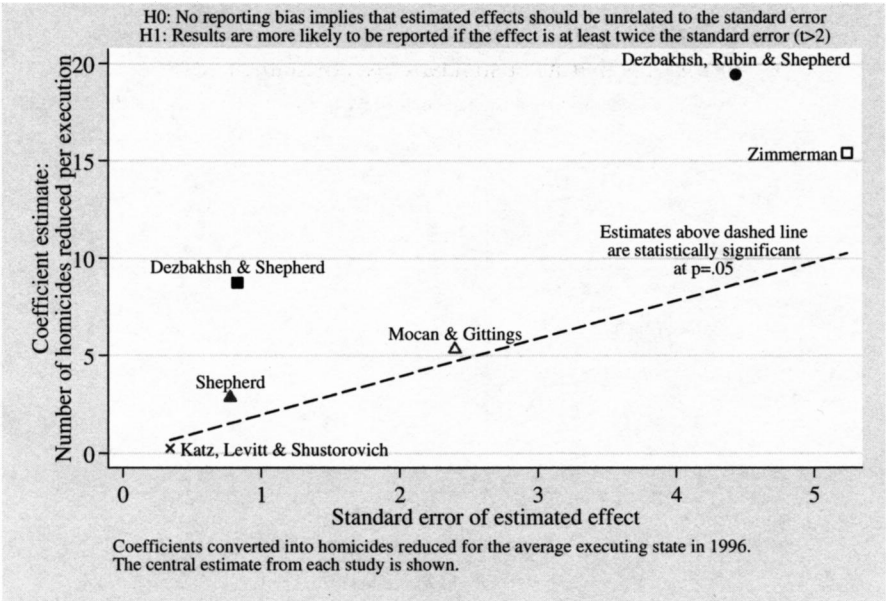
---

106. *See* Orley Ashenfelter, Colm Harmon & Hessel Oosterbeek, *A Review of Estimates of the Schooling/Earnings Relationship, with Tests for Publication Bias*, 6 LAB. ECON. 453 (1999).

107. Compiling the sample still involved some judgment calls. Our goal was to include all comparable aggregate estimates for the average impact of an execution on homicide rates across death penalty jurisdictions. Thus, we included the Mocan and Gittings, *supra* note 11, estimates of the effects of commutations or death row removals as estimates of the effects of an execution foregone, but we omitted the Paul R. Zimmerman, *Estimates of the Deterrent Effect of Alternative Execution Methods in the United States*, 65 AM. J. ECON. & SOC.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 9. Reporting Bias in Estimated Effects of Executions on Homicide: Preferred Estimates Across Six Studies




Recall that if there is no reporting bias, then estimates of the effects of executions should be clustered around the same mean, albeit in a "cone" shape, as the variability of estimates rises (linearly) with the standard error. Moreover, there should be as many estimates in the top half of the cone as in the bottom half, and the estimated effect should be uncorrelated with the standard error. Instead, these data are strongly consistent with evidence of reporting bias. Figure 9 shows the "central" or "preferred" estimate from each study, and its corresponding standard error.[108]

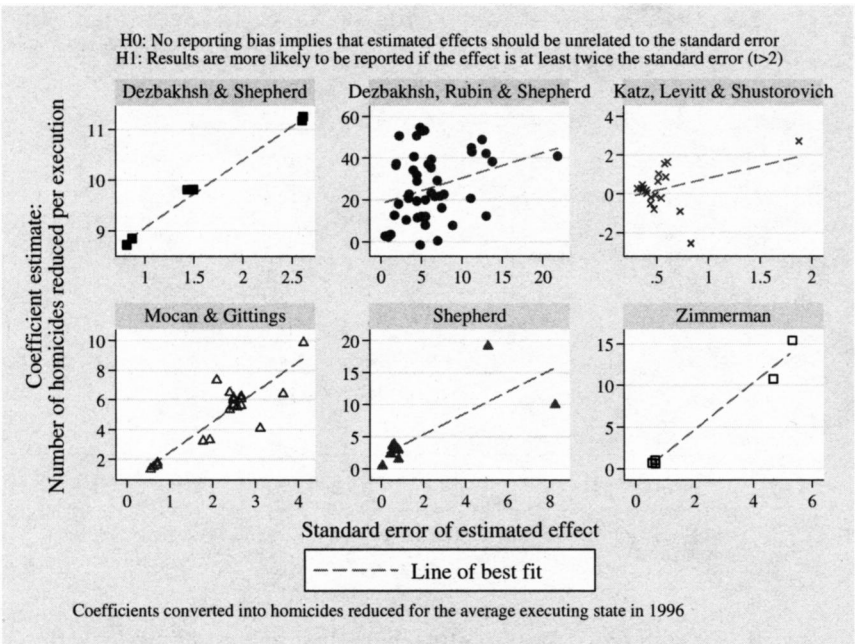
First, note that the reported estimates appear to be strongly correlated with their standard errors: we find a correlation coefficient of 0.88, which is both large and statistically significant. Second, among studies with designs that yielded large standard errors, only large positive effects are reported, despite the fact that such designs should be more likely to also yield small effects or

(forthcoming 2006), *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=355783, estimates of the effects of execution broken down by execution method, the Shepherd, *Deterrence Versus Brutalization*, *supra* note 46, estimates broken down by state, and the Shepherd, *Murders of Passion*, *supra* note 46, estimates of the effect of executions on particular homicide types (although we include the aggregate estimates).

108. The central estimates are from Dezhbakhsh, Rubin & Shepherd, *supra* note 11, at 363 tbl.4, col.1; Dezhbakhsh & Shepherd, *supra* note 33, at tbl.7, col.1; Katz, Levitt & Shustorovich, *supra* note 10, at 327 tbl.2, col.6; Mocan & Gittings, *supra* note 11, at 464 tbl.2, col.1; Shepherd, *Murders of Passion*, *supra* note 46, at 310 tbl.3, col.1; and Zimmerman, *supra* note 11, at 183 tbl.4, col. 2.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Figure 10. Reporting Bias in Estimated Effects of Executions on Homicide: Reported Estimates Within Each Study




even large negative effects. And third, we observe very few estimates with *t*-statistics smaller than two, despite the fact that the estimated deterrent effect required to meet this burden rises with the standard error.

Moreover, while Figure 9 focuses only on the central estimate from each study, Figure 10 shows the pattern of estimated coefficients and standard errors reported *within* each study. Typically these various estimates reflect an author's attempt to assess the robustness of the preferred result to an array of alternative specifications. Yet within *each* of these studies (except Katz, Levitt, and Shustorovich) we find a statistically significant correlation between the standard error of the estimate and its coefficient, which runs counter to one's expectations from a true sensitivity analysis.

In light of this analysis, it is probably not surprising that our sensitivity tests—sampling from the universe of unreported results—yielded more frequent and larger negative (that is, antideterrent) estimates and far more fragile estimates of the deterrent effect of the death penalty. Moreover, to the extent that we report only small deviations from a set of specifications that are likely afflicted by reporting bias, future researchers sampling from a wider array of econometric specifications and samples may find even more conflicting signals.

In sum, if the death penalty had a sufficiently powerful effect on murder rates (in either direction), we are confident that it would emerge from panel

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

data across all fifty states over a nearly seventy-year period. Relatively small effects—either stimulating or deterring homicide—will be hard to tease out, though, given the wide swings in homicide rates. Indeed, these wide swings might lead researchers to find spuriously large effects in small subsets of the data.

We are led to conclude that there exists profound uncertainty about the deterrent (or antideterrent) effect of the death penalty; the data tell us that capital punishment is not a major influence on homicide rates, but beyond this, they do not speak clearly. Further, we suspect that our conclusion that econometric studies are highly uncertain about the effects of the death penalty will persist for the foreseeable future. Quite simply, it is difficult to foresee any states providing a sharp enough policy shock for social scientists to reliably estimate an effect on homicide rates.[109] Consequently, we strongly suggest that substantial caution is required in interpreting *any* studies purporting to show that recent data can speak more clearly than earlier studies allowed.

## CONCLUSION

We have surveyed data on the time series of executions and homicides in the United States, compared the United States with Canada, compared non-death penalty states with executing states, analyzed the effects of the judicial experiments provided by the *Furman* and *Gregg* decisions comparing affected states with unaffected states, surveyed the state panel data since 1934, assessed a range of instrumental variables approaches, and analyzed two recent state-specific execution moratoria. None of these approaches suggested that the death penalty has large effects on the murder rate. Year-to-year movements in homicide rates are large, and the effects of even major changes in execution policy are barely detectable. Inferences of substantial deterrent effects made by authors examining specific samples appear not to be robust in larger samples; inferences based on specific functional forms appear not to be robust to alternative functional forms; inferences made without reference to a comparison group appear only to reflect broader societal trends and do not hold up when compared with appropriate control groups; inferences based on specific sets of controls turn out not to be robust to alternative sets of controls; and inferences of robust effects based on either faulty instruments or underestimated standard errors are also found wanting.

Whether or not the death penalty has a deterrent effect is—as Sunstein and Vermeule rightly argue—a very important question. If policymakers are willing to debate the issue based on the consequences of capital punishment (as Sunstein and Vermeule urge them to do), then it is crucial to try to establish

109. For instance, note that the recent Illinois execution moratorium yielded a change in execution risk much smaller than the sorts of shocks seen during the first half of the century. For more information, see Figure 8, *supra*.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

reliable evidence on whether executions deter or stimulate crime. As such, it seems reasonable to appeal to econometric pyrotechnics. Unfortunately, our survey of the literature suggests that too often these pyrotechnics have yielded heat rather than light.

In general, those interested in policy debates should insist upon clarity and intuitive plausibility in all aspects of research design and analysis. This is especially true in domains where research may be driven by ideology and advocacy motives; these incentives may lead researchers to use econometric sophistication to silence debate rather than enlighten policymakers. While sophistication holds an obvious allure (especially for academics), intuitive plausibility should always be preferred in the realm of real-world policy. Unfortunately, the history of the death penalty debate is replete with examples of plausibility being sacrificed on the altar of sophistication.

In many ways, our tour of the recent death penalty literature brings the debate full circle to the explosion of interest in the topic almost a half-century ago. Thorsten Sellin's research showed a clear realization of the value of conducting before and after comparisons, contrasting "treatment" states with "controls" unaffected by policy changes.[110] As Sellin recognized, it is important to compare effects in jurisdictions that are otherwise subject to similar shocks.[111] Even so, in 1975 Ehrlich argued instead for sophistication, claiming "that the statistical methods used by Sellin and others to infer the non-existence of the deterrent effect of capital punishment do not provide an acceptable test of such an effect."[112] Yet despite the technical sophistications of Ehrlich's approach, he clearly sacrificed plausibility, arguing that he could isolate which movements in the aggregate U.S. homicide rates were caused by changing execution policy and thereby estimate the deterrent effect of capital punishment. The subsequent literature, aptly summarized in a National Academy of Sciences report,[113] confirmed that Ehrlich's strong conclusions about the deterrent effects of capital punishment were unwarranted.

A quarter of a century later, a small surge of studies has appeared claiming that recent data and new econometric methods overturn the earlier consensus. Sunstein and Vermeule appear to believe this claim. Despite the sophistication of the studies on which that claim is based, our analysis shows that they either fail to account for developments in unaffected states, apply sophisticated methods in an entirely inappropriate manner, or yield results which are clearly not robust to small changes. Moreover, not only are panel data not "a newly available form of data,"[114] but they also formed the basis of Sellin's research method. While he did not bury his comparisons in jargon, Sellin's method

110. *See* Sellin, *Homicides*, *supra* note 1, at 135.

111. *Id.*

112. Ehrlich, *supra* note 3, at 398.

113. *See* DETERRENCE AND INCAPACITATION, *supra* note 5.

114. Sunstein & Vermeule, *supra* note 12, at 711.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

essentially comprised a difference-in-differences approach; in his insistence on comparing otherwise similar states, Sellin predicted the subsequent emergence of matching estimators. His methods are not only intuitively plausible, but they are not too far from the current state of the art in empirical microeconomics.[115]

As we have applied somewhat updated econometric techniques to Sellin's methods, we have found that his conclusions remain essentially unchanged. The U.S. data simply do not speak clearly about whether the death penalty has a deterrent or antideterrent effect.[116] The only clear conclusion is that execution policy drives little of the year-to-year variation in homicide rates. As to whether executions raise or lower the homicide rate, we remain profoundly uncertain.

Sunstein and Vermeule argue that capital punishment is morally required if it saves lives. Their assessment of the currently published empirical literature leads them to the view that lives would indeed be saved, which in turn prompts them to call for an increase in the number of executions. Moreover, they argue that it is not sufficient to raise reasonable doubt about the claim that executions will reduce the number of murders, as they argue for a version of the precautionary principle, and hence "the existence of legitimate questions is hardly an adequate reason to ignore evidence of severe harm."[117]

In light of our reanalysis of the data, we would strongly urge them to reassess their conclusion about what is known or knowable about the impact of the death penalty. And we do not mean simply to raise "legitimate questions," but rather to urge them to reconsider fundamentally whether existing data can be sufficiently informative as to form the basis of capital punishment policy at all.[118] The estimated effects of capital punishment on homicide rates change dramatically even with small changes in econometric specifications. Aggregating over all of our estimates, it is entirely unclear even whether the preponderance of evidence suggests that the death penalty causes more or less murder.[119]

---

115. David Card and Alan Krueger's landmark minimum-wage study has been an important catalyst for this style of research, and it shares much of the flavor of Sellin's methods. Card and Krueger were interested in the employment consequences of the minimum wage, so they examined the evolution of employment in New Jersey, comparing it with the evolution of employment among a control group of unaffected firms in eastern Pennsylvania. *See* David Card & Alan B. Krueger, *Minimum Wages and Employment: A Case Study of the Fast-Food Industry in New Jersey and Pennsylvania*, 84 AM. ECON. REV. 772, 773 (1994).

116. Conceivably, a careful study of international statistics might provide richer data with which to illuminate the deterrent question, although (depending on which countries are examined) this might raise an additional question whether responses to the use of the death penalty in countries with very different cultural backgrounds and legal institutions would be relevant to the United States.

117. Sunstein & Vermeule, *supra* note 12, at 715.

118. *Id.*

119. As such, our conclusions most closely match those of Steven Levitt. For a

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

Alternatively, to frame the issue as a Bayesian would, one's posterior belief about the deterrent effect of the death penalty surely looks a lot like one's prior belief. We can be sure that the death penalty does not cause or eliminate large numbers of homicides, but we learn little else from the data. As such, there is little evidence to convince believers in the deterrent hypothesis otherwise, as there is little to persuade believers in the competing brutalization hypothesis. Thus, it remains for Sunstein and Vermeule either to accept that their argument provides no useful guidance to policymakers or to argue that the death penalty is morally required if one has a strong enough prior belief. In light of their suspicions that "cognitive processes contribute to large mistakes, at least on questions of fact,"[120] one suspects that they would also be led to agree that—in light of the highly uncertain evidence—their argument has little prescriptive content.

To the extent that there is a prescription in Sunstein and Vermeule's argument, it is to emphasize the importance of a direct interplay between crime research and (highly politicized) policymaking. Unfortunately, recent history on this score is not particularly encouraging. Isaac Ehrlich's econometric evaluation of the deterrent effect of the death penalty breathed new life into the pro-death penalty movement. Even though Ehrlich's 1975 study was to be later discredited, the real problem was not that a flawed empirical paper had been written, but rather that there were those who leapt to use it as a tool to advance the goal of reinstating capital punishment in the United States before the validity and reliability of the work had been fully explored. In the words of the National Academy of Sciences report on Ehrlich's work: "[I]t seems unthinkable to us to base decisions on the use of the death penalty on Ehrlich's findings, as the Solicitor General of the United States has urged. They simply are not sufficiently powerful, robust, or tested at this stage to warrant use in such an important case."[121] More recently, numerous legislators, and even former Attorney General John Ashcroft, have been willing to rely on the findings of John Lott[122] as constituting powerful evidence that right-to-carry

---

particularly sharp articulation, see Douglas Clement, *Does the Death Penalty Deter Homicide? New Economic Studies Seek the Answer to an Age-Old Question*, REGION, June 2002, *available at* http://minneapolisfed.org/pubs/region/02-06/debate.cfm. Clement reports:

> "What's interesting about this is that it mirrors so closely the Ehrlich debate of the '70s," said Chicago's Levitt, "which basically all came down to if you tweak his specification at all, you get numbers that are totally different." And reaching a definitive answer about deterrence could well be impossible since current execution rates may be too low to provide sufficient empirical data. "I really think not that the answer is 'yes' or 'no,'" said Levitt, "but that there's not enough information to figure it out. There may never be enough. It may just be a question that can't be answered."

*Id.*

120. Sunstein & Vermeule, *supra* note 12, at 740.

121. DETERRENCE AND INCAPACITATION, *supra* note 5, at 358.

122. JOHN R. LOTT, JR., MORE GUNS, LESS CRIME: UNDERSTANDING CRIME AND GUN CONTROL LAWS (2d ed. 2000).

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms

laws would save lives.[123] It was only subsequent to the legislative and judicial debates that a consensus emerged in the scientific community, and in each case the early research did not withstand the scrutiny of a National Research Council report.[124]

These episodes suggest the potential dangers awaiting those who might wish to short-circuit the full process of scientific inquiry and validation and rush directly to the legislative forum when initial empirical findings seem to support a favored policy position. Lamentably, studies that were later utterly discredited continue to influence policy since the evidentiary burden required to reverse course appears to be high. In our view, Sunstein and Vermeule are a bit cavalier in thinking that we can start down a path of greater reliance on the death penalty today and then turn away from it if the evidence later proves that there is no effect (or even a pernicious effect). It is far better to insist on a stronger foundation of statistical proof before advancing a policy position, particularly one that the polity may be predisposed to embrace without regard to the degree of social scientific support.

As Hashem Dezhbakhsh—the lead author of the Dezhbakhsh, Rubin, and Shepherd study—argued with respect to John Lott's work: "The academic survival of a flawed study may not be of much consequence. But, unfortunately, the ill-effects of a bad policy, influenced by flawed research, may hurt generations."[125] While Dezhbakhsh was referring to John Lott's research on guns, his insight is equally applicable to the debate over capital punishment.

---

123. In 2002, eighteen state attorneys general referenced Lott's work in a letter to Attorney General John Ashcroft supporting his interpretation of the Second Amendment as protecting the right of individuals to bear arms. *See* Letter from Bill Pryor, Alabama Attorney General, to John Ashcroft, U.S. Attorney General (July 8, 2002), *available at* http://www.ago.state.al.us/ag_items.cfm?Item=81; *see also* 146 CONG. REC. S349 (daily ed. Feb. 7, 2000) (statement of Sen. Craig); 145 CONG. REC. H8645 (daily ed. Sept. 24, 1999) (statement of Rep. Doolittle).

124. NAT'L RESEARCH COUNCIL, FIREARMS AND VIOLENCE: A CRITICAL REVIEW (Charles F. Wellford et al. eds., 2004).

125. Hashem Dezhbakhsh, *First Person: More Guns, Less Crime? Hashem Dezhbakhsh Disagrees*, EMORY REP., Sept. 27, 1999, *available at* http://www.emory.edu/EMORY_REPORT/erarchive/1999/September/erseptember.27/9_27_99dezhbakhsh.html.

This content downloaded from 130.91.144.125 on Thu, 11 Aug 2016 18:21:09 UTC
All use subject to http://about.jstor.org/terms