

Empirical Evaluation of Law: The Dream and the Nightmare

John J. Donohue III *Stanford Law School and NBER*

Send correspondence to: John J. Donohue III, Crown Quadrangle, Stanford Law School, Stanford, CA 94305, USA; E-mail: Donohue@law.stanford.edu

I discuss the empirical revolution in law and economics, and use the analysis of the deterrent impact of the death penalty to chart the tremendous advances in estimating causal effects since the mid-1970s. This story highlights how ostensibly sophisticated studies frequently generate incorrect estimates, and how difficult it is to know what studies should be believed—a difficulty open to being exploited by those (the media, think tanks, and others) who seek to promote clearly weak studies for some private agenda. I offer a hierarchy of methodologies to assist in evaluating empirical studies and some suggestions for promoting the search for truth. (*JEL*: K00, C10, C50, K14)

1. Introduction

The empirical revolution that has swept the academic and business world has, unsurprisingly, also transformed the field of law and economics. Any examination of the *American Law and Economics Review*, the *International Review of Law and Economics*, the *Journal of Law and Economics*, the *Journal of Legal Studies*, the *Journal of Empirical Legal Studies*, the pure economics journals, law reviews, and other scholarly collections reveals that

This article is based on the Presidential address given at the 2011 meetings of the American Law and Economics Association at Columbia Law School. My thanks to Ian Ayres, Ryan Bubba, Daniel Nagin, Christopher Robertson, Peter Siegelman, and Abe Wickelgren for helpful comments on an earlier draft, the students in my Practicum on the Death Penalty at Stanford Law School and Alex Albright for outstanding research assistance, Justin Wolfers for all his brilliant work on the death penalty and deterrence, and Stanford Law School for research support.

American Law and Economics Review
doi:10.1093/aler/ahv007

© The Author 2015. Published by Oxford University Press on behalf of the American Law and Economics Association. All rights reserved. For permissions, please e-mail: journals.permissions@oup.com.

a field that was once dominated by theoretical work now has a large and growing empirical focus. This movement from the theoretical to the empirical has been propelled from the supply side by advances in computing, the development of increasingly powerful yet user-friendly statistical packages, and the growing availability of data, with many datasets available for download with just a few clicks of a mouse. At the same time, the recognition of the new power of empirical analysis has led to continuing academic developments and refinements in econometric and empirical tools.

The allure of empirical data has spread beyond scholarly journals, becoming a growing force in the cultural landscape, with books such as Chicago economist Steven Levitt's *Freakonomics* (not to mention *Superfreakonomics* or even *Think Like a Freak*) and Yale law professor Ian Ayres' *Super Crunchers: Why Thinking-by-Numbers Is the New Way to Be Smart*. These popular books illustrate the force and power of empirical work, which has inspired a new generation of academics to focus on clever ways of using empirical methods to answer important, interesting, and often politically controversial questions.

In law and economics, the empirical revolution has one animating aspiration: to estimate the direction and magnitude of causal effects with sufficient precision to guide policy, using a protocol that is recognized by knowledgeable researchers to generate reliable results. With improved data as well as increasingly more sophisticated empirical tools, the dream of convincingly identifying the direction and magnitude of true causal influences has been achieved in specific instances more than ever before, but far too often the outcome of empirical studies has been broad uncertainty about what, if anything, has been learned. Indeed, at times, the nightmare scenario unfolds where the power and complexity of statistical analysis is arrayed on behalf of incorrect results that therefore mistakenly push legislators and policymakers in the direction of misguided policies.

In this paper, Section 2 begins by noting the commendable and highly desirable growth in empirical studies, but also stressing the dangers posed when methodologically unsound studies are not screened out in the process of peer review and emphasizing the dangers posed by “motivated” researchers and an uninformed public. Section 3 sets forth a methodological hierarchy of empirical studies, and uses the pioneering work of Isaac Ehrlich examining the deterrent effect of the death penalty to illustrate the

strengths and weaknesses of different types of empirical evidence and the advances (and some setbacks) in empirical work that have occurred over the last 40 years (Ehrlich, 1975, 1977).

Section 4 shows how worthless studies are promoted to influence policy on the death penalty, and Section 5 illustrates that some academics can be blinded by their own ideology in ways that undermine their critical capacity to evaluate evidence. The section also discusses details of various studies claiming to find a deterrent effect of the death penalty using invalid instrumental variables approaches, and argues that the best empiricists allow the data to transcend their ideology. Lastly, Section 6 concludes by revisiting the aforementioned themes and seeks to emphasize the importance of allowing empirical evidence to refine and shape our priors rather than having our priors dictate the nature of the empirical evidence we present.

2. Empirical Evidence: Growth, Shortcomings, Dangers

2.1. Growth

An examination of the most cited articles in a discipline can help illuminate the nature of its scholarly discourse. Before 1985, most highly cited articles in economics tended to be major theory pieces. Theory was the dominant force in the discipline, with empirical evaluation playing a clearly secondary role. Over the last three decades, this ranking has been dramatically altered: the single most cited article in all of economics since 1985 is Hal White's "A Heteroskedasticity-Consistent Covariance-Matrix Estimator and a Direct Test for Heteroskedasticity."¹ Few theorists have even heard of, let alone read this work; fewer still have any idea what heteroskedasticity is.² Indeed, eight of the top 11 most cited articles are

1. Virtually all of the empirical articles cited herein either cite White's article or make the adjustment to standard errors that his paper suggests (White, 1980). Without this correction, these articles would tend to exaggerate the statistical significance of their findings.

2. I was once serving as an expert witness in a federal employment discrimination case where a key issue was the need to correct for heteroskedasticity in measuring differences in promotion rates among groups to see if there was evidence of discrimination. After an extended cross-examination on the issue, the judge turned to the jury and said "we will now take our afternoon break and let me admonish you that you are not to discuss the case on the break. In particular, I don't want to hear any discussion of heteroskedasticity." The comment elicited quite a mirthful response from the jury, and

technical econometrics articles (Kim et al., 2006). The growing importance of empirical economics is unmistakable.

The continuing developments in the science of econometrics and the growing role of empirical work in law and economics scholarship has been driven on the demand side by the inherent limitations of theory: theory can, at times, identify the direction of an effect of a law or policy, but it rarely has anything to say about the magnitude of the effect, which is almost always a central issue for optimal policy. As Goethe wrote, “All theory, dear friend, is gray, but the golden tree of life springs ever green.” The comment might suggest that empirical work is a deeper and richer source of knowledge, but that will only be true if it is informed by a strong theoretical foundation. Despite the admonition of Wittgenstein—“Don’t think. Look!”—we must both think and look. The necessity of combining empirical work and theory was reflected in the remarks of the pioneering empirical economist James Heckman, who accepted his Nobel Prize in 2000 “[o]n behalf of all economists who analyze micro-economic data and use micro-econometrics *to unite evidence and theory*, and to evaluate policy interventions of all kinds” (emphasis supplied) (Heckman, 2000).

2.2. Shortcomings

2.2.1. *Imprecise estimates.* But while the growth of empirical evidence in evaluating policy interventions has been both prodigious and commendable, the goal of generating precise and broadly accepted measures providing clear guidance to policymakers is not easily achieved. Consider the academic discussion of the impact of perhaps the most important policy variable undertaken to influence crime over the last four decades—the enormous increase in our rate of incarceration from roughly 500,000 in custody in the early 1970s to a peak of almost 2.5 million behind bars 35 years later. While the evidence is clear that the initial doubling of the prison population had a substantial impact on crime, the issue of how much crime falls from additional increments of increased incarceration is still very much up for debate (Donohue and Siegelman, 1998). (Note each doubling costs twice as much as the prior doubling and generates smaller incremental benefits

speculation from me about whether there was any chance anyone else in the courtroom had any clue about the issue.

in terms of reduced crime.) Specifically, the major studies offer elasticity point estimates ranging from -0.05 to -0.4 , leaving policymakers with an uncomfortably large lack of precision (Donohue, 2009).

Do we have too much incarceration in the United States or not enough? The low-end elasticity estimate implies the U.S. prison population should be *decreased* by 700,000 (since the crime-reducing benefits are far outweighed by the costs of additional incarceration). The high-end estimate implies that the prison population should be *increased* by 1.2 million (since the benefits are so great they outweigh the considerable burdens of mass incarceration). This degree of imprecision in this empirical evidence—while preferable to mere theoretical claims and valuable in clarifying which factors most powerfully influence judgments about the optimal level of incarceration—provides deeply insufficient guidance for policymakers.

2.2.2. *Conflicting estimates.* Even worse, dueling studies generate utterly conflicting estimates with respect to a host of important issues in law and policy—including, but far from limited to, the impact on crime of guns,³ police,⁴ legalized abortion,⁵ and three-strikes laws;⁶ the effect of the death penalty on murder; the impact on stock prices of staggered boards or incorporation in Delaware;⁷ whether securitization of subprime loans led to lower

3. Lott and Mustard (1997) conclude that right-to-carry laws deter violent crimes, however, Aneja et al. (2014) find no evidence of deterrence and provide suggestive evidence that RTC laws are associated with significant increases in aggravated assault.

4. While the majority of studies surveyed by Cameron (1988) found that either there is no relationship between number of police and crime or that increases in the number of police are associated with *increases* in the level of crime, Klick and Tabarrok (2005) provide strong evidence that there is a causal deterrent effect of police size on criminal activity.

5. Donohue and Levitt (2001) provided empirical evidence that the legalization of abortion contributed to the decline in crime in the United States during the 1990s. That finding was then questioned by Foote and Goetz (2008). Recently, Francois et al. (2014) used cross-country evidence from Europe in order to evaluate the validity of the original Donohue–Levitt hypothesis. The authors found that the legalization of abortion had a significant and negative impact on crime rates across Europe.

6. Iyengar (2008) argues that three strikes laws reduce participation in criminal activity, while Marvell and Moody (2001) find little evidence that crime declines due to these laws and even argue that it is likely three strikes laws *increase* homicides.

7. Daines (2001) presents evidence that corporate law improves stock prices in Delaware, but Cremers and Sepe (2014) present evidence that supports the opposite conclusion. Bebchuk and Cohen (2005) “find that staggered boards are associated with an

scrutiny of borrowers,⁸ and whether illegal drugs and alcohol are complements or substitutes.⁹ While the empirical revolution has been exciting and promising, the wide range of divergent conclusions that emerge from different studies has left the public confused about the reasons that so many studies are in conflict and how these studies should be interpreted to influence policy. Indeed, some individuals or organizations use econometrics in order to obscure the truth to support their own preferred policies, which complicates the task of discerning true causal relationships. Of course, the institution of peer review helps to some degree, but in practice peer review has done a poor job of screening out work relying on problematic empirical techniques.

2.3. Dangers

2.3.1. *Failure of peer review.* One example of the mischief that can follow from failures in the peer review process occurred when the journal *Applied Economics Letters* published a paper by Mark Gius that concluded that the federal assault weapons ban (AWB) had led to a 19.3% *increase* in the gun murder rate (Gius, 2014). This paper should never have made it through the peer review process for two reasons.

First, anyone familiar with crime during the period of the federal AWB would be highly skeptical of Professor Gius' finding that the federal AWB sharply increased the gun murder rate. The U.S. murder rate was at its peak in the year before the federal AWB passed and then fell sharply thereafter. When the AWB ended a decade later, homicides rose and remained elevated for 3 years, before a further decline began. Specifically, in 1993 (the year before adoption of the federal AWB) the gun murder rate was 6.2 per

economically meaningful reduction in firm value (as measured by Tobin's Q)," while Cremers et al. (2013) find that "firms adopting a staggered board increase in value, while de-staggering is associated with a decrease in value."

8. Keys et al. (2010) conduct a regression discontinuity analysis and conclude that securitization did dull the incentives of subprime lenders, while Bubb and Kaufman (2014) question the appropriateness of their regression discontinuity design, concluding that the credit score cutoff rules cannot be used to assess the loan behavior of securitizers, and raising doubts that securitization led to lax screening.

9. Crost and Guerrero (2012) as well as Conlin et al. (2005) find evidence that suggests alcohol and marijuana are substitutes. Using the same regression discontinuity approach as Crost and Guerrero (2012), Yörük and Yörük (2011) find that the two substances are complements.

100,000, and by the time the federal AWB ban had elapsed in 2004 it had fallen to 3.4. The 10-year downward trend in the gun murder rate rose was reversed when the federal AWB lapsed, and for the following 3 years this murder rate was never below 3.6 per 100,000. Nothing in any of the extensive writing on the federal AWB specifically or in the broader criminological literature could support the counter-intuitive claim that the federal AWB dramatically *increased* gun murders in the United States.

The second failing in the Gius article, which the referee should have identified, was that Gius' statistical model contained a fatal econometric flaw. The methodology that Professor Gius employed is referred to as a panel data model with state and year fixed effects, which has been one of the workhorses of empirical evaluation in law and economics. Panel data models can be useful in examining a change in law or policy adopted by selected states (preferably at different times) so that one can compare what happens in the states that adopt the legal change to the states that do not adopt the legal change. This is an appealing feature since it can allow the researcher to separate the data into the treated group (the set of states that adopts the law during the relevant data period) and the control group (the set of all other states).

Two requirements are needed to generate estimated effects of a legal change in a panel data model with state and year fixed effects: (1) one can only estimate the effect of a state law for those states that adopt or eliminate the law during the data period (the changing states); and (2) the legal change must not apply to all states at the same time.

Professor Gius notes, in his paper, that he was the first to look at the impact of the federal AWB using a panel data model. In fact, one cannot directly estimate the impact of the federal AWB when controlling for year fixed effects, as Professor Gius claimed to do. Doing so would violate Condition (2) because the federal law turns on and off at the same time for all states, which prevents the treatment and control comparison across states that is the central feature and primary value of this type of model.¹⁰ Thus, the Gius model cannot be estimated in the format he specified.

10. In technical econometric terms, the year fixed effects for the period 1994–2004 that Professor Gius enters into his model would be perfectly collinear with his federal AWB variable. Therefore, one cannot estimate a model that controls for year fixed effects and the federal AWB.

But how did the paper estimate the effect of the federal AWB variable if his explanatory variables were perfectly collinear? In the old days before user-friendly statistical packages were available, the regression would not have run, but now these packages will simply drop a variable to eliminate perfect collinearity, which in this case could be achieved by dropping any one of the year effects over the period 1994–2004. When I checked, I discovered that the year fixed effect for 2004 was the one that was dropped, which meant that the value generated by the statistical package for the federal AWB variable was in fact simply the 2004 year fixed effect. Since dropping a different year's fixed effect would give an entirely different estimate, it should be clear that the coefficient on Professor Gius' federal AWB variable cannot provide useful information about the impact on gun murders of the federal AWB. But, that did not stop the NRA from filing a lawsuit seeking to challenge Maryland's AWB using the Gius article as ammunition against the state law.¹¹ One shudders to think that an NRA-sympathizing judge might rule that the "peer-reviewed" literature establishes that gun murders are sharply elevated by the passage of AWBs.

While one might have expected more in this case from the peer review process, it must be acknowledged that even perfect peer-review cannot ensure that published studies are valid. Any major empirical study will require the researcher to make a large number of potentially influential choices that will not be identifiable without having access to the data and computer program of the authors. Since such access is virtually never available to a referee, only a process of attempted replication could begin to reveal the many serious problems that can infect even major papers in top empirical journals.¹² The result is that even highly talented professors

11. Gius filed an expert report for the NRA, and I submitted a responsive report for the state of Maryland.

12. Economist David Card at Berkeley has advanced the goal of ensuring greater credibility by encouraging his graduate students to replicate major published studies (now that journals like the *American Economic Review* require data and computer programs to be made available to the public). Since researchers have to make hundreds of choices in implementing their regression models, most of which will not be available to referees and editors at the time the publication decision is made, researchers who are motivated only to secure a publication (which may have considerable professional benefits—such as securing tenure), can cherry-pick their results to generate appealing results (even if they are clearly not robust to reasonable adjustments in specification). This process is regularly seen in litigation, with the advantage that the other side can identify the cherry-picking,

with strong empirical backgrounds, let alone judges and policymakers, are unsure of which studies can be relied on as they try to evaluate the published research in any particular area. The problem for judges and policymakers is compounded by the fact that they are often bombarded by “expert” opinions designed to mislead or confuse them in order to obscure rather than reveal the truth.

2.3.2. *Dangers of ideology and motivation.* Very few in the public fully appreciate the range and seriousness of the vexing issues that empirical studies must grapple with: endogeneity, omitted variables, misspecification, bad data, coding errors, inadequate science, and misinterpretation. These issues are compounded by the impact of ideology and self-interest (financial, occupational, and political). Navigating one’s way through this gauntlet of dangers in order to discover a true causal relationship is perhaps the most difficult task in the entire academy. Since those conducting complex empirical research must make many low-visibility but potentially influential choices, one must be particularly wary of researchers who are excessively burdened by ideology or too highly motivated to achieve a particular result. The dangers become acute if such researchers are working with inadequate empirical tools because this can often allow the motivated researcher to simply select an appealing finding rather than discover a causal truth.

Consider the case of Lord William Thomson Kelvin, a genius mathematician and physicist—a “Victorian Superman.” Lord Kelvin was admitted to Glasgow University at age 10. He formulated theories on the topics of electromagnetism, thermodynamics, and the wave theory of light. His place in the scientific pantheon is confirmed by his burial site between Sir Isaac Newton and Charles Darwin in Westminster Abbey.

But all of this talent could not overcome the burdens of ideology. In 1863, Kelvin sought to calculate the age of the earth based on assumptions regarding the time it would take for the planet to cool to its current temperature.

but with the disadvantage that judges and juries may be wholly incapable of sorting out the truth even when one side is clearly pointing it out (Donohue, 2014). As a result, more judges should follow Richard Posner’s example of appointing neutral econometric expert witnesses as special masters to guide them in evaluating econometric testimony. (Sidak, 2013).

He initially estimated the earth to be roughly 100 million years old, but in the ensuing decades, the deeply religious and anti-evolutionist Kelvin kept “refining” his estimates down, ending with a final figure of 20 *million* years. In fact, the true age of the earth is orders of magnitude higher—4.54 *billion* years. The major theoretical difficulty for Kelvin was that he did not know that radioactive decay in the earth’s core keeps resupplying heat. As a result, his cooling equation was missing a key component, which underscores how inadequate theory can impair empirical estimates. But Kelvin almost certainly would have done better had he not been blinded by his religiously based antagonism to the doctrine of evolution, which motivated his efforts to show the earth was too young to enable the full operation of natural selection that Darwin posited.¹³ It was this non-truth-oriented influence that led Kelvin to ignore fossil evidence pointing to a much older age of the earth.

The enormous intelligence and stellar reputation of Kelvin coupled with the evident ingenuity and sophistication of his work persuaded many to accept his conclusions. The unhappy result was that Kelvin’s work “imposed a strait jacket on studies of the Earth’s evolution that cramped geological thinking for the rest of the century” (Burchfield, 1975). The lesson for consumers of empirical research is “beware the ‘motivated’ researcher.” The lesson for the empirical researcher: curb the blinding influence of ideology and other non-truth-oriented influences on your analysis.¹⁴

An empiricist of the highest order will elevate the search for truth over the desire to prove a personally congenial result. One suspects that a scientist such as Charles Darwin would have been far less prone to make such a gross Kelvin-like error since Darwin’s fundamentally empirical cast of mind would not be shackled by priors or overly motivated to seek a particular result. Rather than rejecting a body of conflicting evidence, Darwin was renowned for relentlessly challenging his priors with a meticulous attention

13. Because it was recognized in the 19th Century that the process of natural selection would take tens or hundreds of millions of years to unfold, those who wished to disprove Darwinian selection had a strong incentive to argue that the earth was insufficiently old to allow that process to have occurred.

14. A tragic illustration of where these two lessons were not heeded was offered in the documentary *Inside Job*, which examined the role played by economists in providing reports that both served the interests of the financial industry that hired them and facilitated the financial collapse of 2008, with costly consequences for hundreds of millions of individuals around the world (Chan, 2010).

to potentially disconfirming evidence. This truth-oriented and scientific frame of mind is the ideal for conducting empirical research across the board—whether in law, economics, policy, or medicine.

3. The Developing Empirical Methodology in Law and Economics

3.1. Ehrlich on Deterrence and the Death Penalty

The introduction of serious econometric evaluation of the law emerged roughly 100 years after Kelvin’s brilliantly creative, but ultimately misguided forays into estimating the age of the earth. Isaac Ehrlich’s work on deterrence and the death penalty study was an early and influential contribution that had some of the strengths and weaknesses of Kelvin’s work. In addition to Ehrlich’s time-series study in the *American Economic Review* (1975) claiming that each execution leads to eight fewer murders, he published a cross-sectional study in the *Journal of Political Economy* (1977) that also concluded there was a deterrent effect of the death penalty. The US Supreme Court cited Ehrlich’s time-series study in reinstating the death penalty in 1976. The brief submitted by then Solicitor General Robert Bork (but written by Frank Easterbrook) extolled the Kelvin-like sophistication and path-breaking intelligence of Ehrlich’s work ([Brief for United States as Amicus Curiae, Gregg v. Georgia, 1976](#)). But just as the combination of ideological fervor and inadequate scientific knowledge led Kelvin astray, Ehrlich’s work was also marred by the combination of his idiosyncratic specification choices and the limitations of the empirical tools he employed—even though they were considered “state of the art” by journal editors at that time. Weak tools combined with very strong priors are not a good recipe for sound empirical work.

3.2. A Methodological Hierarchy

To evaluate Ehrlich’s work, note that there is a methodological hierarchy in the ability to convincingly establish the existence of a causal relationship. At the top of the hierarchy is the gold standard for such proof: a randomized study in which there is a treatment and a control group. Such studies are less common in the social sciences than in the medical literature, but even in the former domain there have been a number of tremendously

important randomized experiments.¹⁵ Descriptive statistics such as graphs and percentages are placed at the bottom of the hierarchy to the extent that they are used to capture the full extent of a causal claim. Although they have a pleasing simplicity about them and have the capacity to persuade through appeals to intuition, such tools are quite limited if naively invoked. Unless they are based on a substantial supporting architecture of theoretical or empirical work, graphs and percentages can seldom identify true causal relationships.¹⁶ (Happily, as I hope to demonstrate below, *when the supporting architecture is present*, appropriate graphs can provide useful insights, while stripping away some of the complexity that can obscure what is really driving results in complex econometric models).

Essentially, time series and cross-section studies—Ehrlich’s two approaches—are near the bottom of what I would deem the current methodological hierarchy for determining the causal impact of law and policies:

- (1) Randomized experiment.
- (2) Natural experiment or regression discontinuity.
- (3) Panel data.
- (4) Time series or cross-section analysis.
- (5) Graphs or Percentages.

I must hasten to add that any of these five levels of empirical approaches can be poorly conducted, so I am not suggesting that every higher-order

15. Important early randomized experiments were (1) the income maintenance (or negative income tax) studies conducted from 1968 to 1979 (Munnell, 1986); (2) the RAND Health Insurance Experiment, which randomly assigned individuals to different types of health plans and assessed their behavior and outcomes from 1974 to 1982 (Brook et al., 2006); and (3) the Kansas City policing experiment undertaken in 1972 and 1973 (Larson, 1976). One of the most important randomized experiments in recent years was the Moving to Opportunity program launched during the Clinton Administration and still generating valuable research about the impact of enabling families to move out of the worst inner-city neighborhoods. See Kling et al. (2007), Sanbonmatsu et al. (2006), Ludwig and Kling (2007). Of course, randomized experiments have often gone badly astray—differential attrition is a potentially serious problem—so one always needs to carefully evaluate the quality of any study before its conclusions merit acceptance.

16. A graph could certainly be used as method to display the results of a randomized experiment, but then it is the experiment that is doing the work of establishing the causal relationships, and the graph is merely the method of showing the results in what is often a more compelling manner.

study is sound or more valuable than a well-performed lower-order study.¹⁷ Moreover, at times, data going back many years for states or individuals will not be available, so panel data studies, which require all of the data going back in time of a time-series analysis as well as the data across subjects or jurisdictions of a cross-section analysis, may be impossible to produce.¹⁸ Such restrictions can be particularly constraining in the world of litigation, where it simply may not be feasible to do anything more than a category (4) analysis. Nonetheless, attentiveness to the issues discussed herein can help improve any study seeking to identify a causal relationship.

3.2.1. *The low ranking of national time-series analyses.* To see why Ehrlich's approach merits this low ranking in the methodological hierarchy, consider Figure 1, which shows the homicide and execution rates in the United States over the last century. One can readily see that starting in 1933—importantly, the year Prohibition ended—murder rates

17. This proposed hierarchy conceals much complexity—as one might expect given the thrust of this article. In a natural experiment, factors outside the control of the researcher have led one group of individuals to be treated differently than another in a way that is arguably similar to random assignment. This term can be applied broadly to include (1) regression discontinuity cases, where the researcher looks on either side of some selection cutoff to see if those who are just above or just below the threshold experience different outcomes that can be attributed to this differential treatment (Dunning, 2012); (2) panel data studies when exogenous factors lead states to adopt different laws, thereby facilitating inter-state comparisons; and (3) instrumental variables estimation, which I discuss in Section 5. As the Section 5 discussion illustrates, one can employ an instrumental variable in panel data studies (or even in time-series studies as Ehrlich did in his 1975 work), which then complicates the categorization. Needless to say, it does not elevate the quality of a poorly designed study because it can be referred to as a “natural experiment.” For example, as I emphasize in Section 5, an invalid instrument will not yield useful information on the causal question of interest.

18. Note that different types of data are needed for particular questions. For example, while panel data with state and year fixed effects would be superior for identifying the effect of, say, a liberalized state divorce law, such panel data (or time series data) could not answer the issue of why some cities have high murder rates than others. The panel data fails because it assumes that there are some enduring traits of the cities that cannot be captured in quantitative variables (the city fixed effect in this case). The research question is trying to ascertain what those fixed effects are. With rich enough data of a cross-section of cities, you could try to answer why the 2012 murder rate in Dallas (12.4 per 100,000) is twice that of slightly larger Las Vegas (5.1), but if there really are enduring traits that cannot be captured in explanatory variables (the assumption behind the city fixed effect in the panel data analysis), the cross-section analysis will be unable to explain these factors.

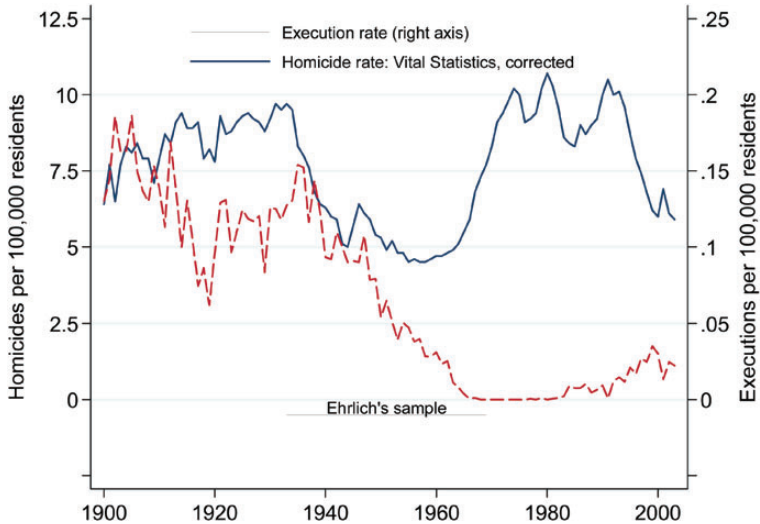


Figure 1. Homicide and Execution Rates in the United States.

and executions rates both fell substantially for the next 30 years. This, of course, does not provide much support for the deterrence thesis that Ehrlich was advancing since, it would seem, the rates of execution and homicide move in the same direction, while the deterrence hypothesis requires there be a negative correlation between executions and homicides.

So how then does Ehrlich generate the eight lives saved per execution prediction? This estimate becomes possible because there was a sharp jump in homicides from 1963 to 1969 at the same time the nation moved from a very low rate of execution to virtual abolition. Thus, while the 80% drop in execution rate risk after 1933 was accompanied by three decades of falling homicides rates, Ehrlich would have us believe the final drop to no executions in the late 1960s suddenly caused an explosion in homicides. Strike one against Ehrlich—the thesis seems implausible in light of a comprehensive view of the evidence (the same first strike against the Gius study on AWBs and gun murders). Remember we must think and look!

Indeed, in light of this pattern over the full 1933–1969 period, one may wonder how Ehrlich was able to overcome the influence of 30 years of contradictory evidence with a few years of superficially supportive evidence at the end of his sample. By specifying his regression model using the log of

the execution rate, Ehrlich gave greater weight to the ostensibly larger percentage change in the execution rate (moving from low to near zero) that occurred in the late 1960s. If the model had used the actual execution rate rather than the log form, no statistically significant deterrent effect would be seen (Passell and Taylor, 1977, pp. 448–449). Strike two against Ehrlich—his results are highly sensitive to the functional form of the execution rate variable. This chosen form implausibly suggests that massive reductions in the rate of execution over three decades when homicides were falling should be weighed less heavily than the brief transition from very low to near-zero rates of execution, which is deemed to have unleashed a dramatic rise in the rate of murder.

But the most basic problem with Ehrlich’s time-series analysis is that it fails in the crucial quest to rely on a treatment and a control to establish causality. Time-series analyses do not handle the counterfactual well, which is the key to causal analysis. What we need to consider is “what would have happened to the homicide rate if the execution rate had not declined?” If we had evidence for that, we could compare what did happen (seen in Figure 1) with this counterfactual scenario. This difference would more plausibly illuminate the causal influence of the change in the execution rate.¹⁹

3.2.2. Superior panel data models with state and year fixed effects. An analogy to a medical study examining whether a drug reduces headache pain may be helpful. We would never conduct an experiment that only observed the treated group that received the drug; we also need to consider an appropriate control group in order to make valid comparisons. Figure 2 moves beyond Ehrlich’s national time-series analysis to consider a possible treatment and control comparison by looking at the homicide rates in the United States and Canada. Figure 2 captures the theory behind panel data models with state and year fixed effects, which we discussed above concerning

19. Indeed, the synthetic control technique first implemented by Abadie and Gardeazabal (2003) was designed to devise a plausible counterfactual against which one could compare a traditional time-series outcome. This technique has been evolving over the last decade and is gaining popularity as another tool for identifying causal relationships. I am currently working with coauthors on a paper that uses this approach to assess the impact of laws allowing citizens to carry concealed handguns. It will be worth noting whether this relatively new technique will surpass traditional panel data analyses in estimating causal effects.

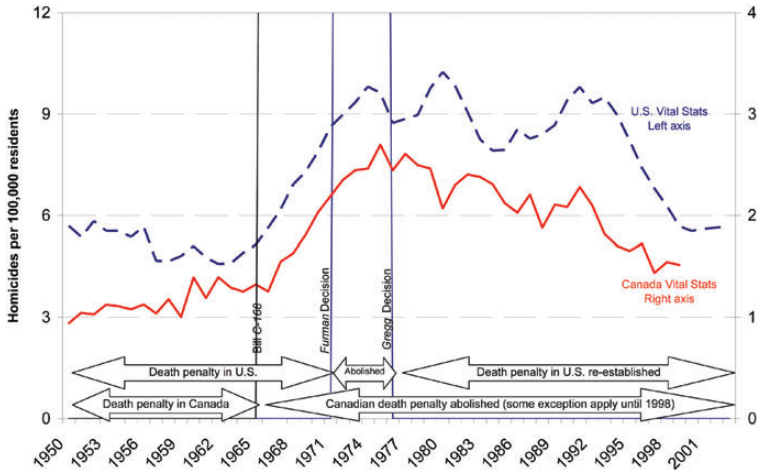


Figure 2. Homicide Rates and the Death Penalty in the United States and Canada. *Note:* This graph previously appeared in [Donohue and Wolfers \(2005\)](#). It is reprinted here with the permission of the *Stanford Law Review* and the Board of Trustees of the Leland Stanford Junior University.

the impact of AWBs on gun murders. Note that panel data analyses stand above the Ehrlich approaches in the methodological hierarchy. One immediately sees that while murder rates in Canada and the United States are at very different levels (Canada’s murder rate is only one-third as large as that in the United States—which is the “state fixed effect,” or more precisely in this example, the “country fixed effect”), they tend to move in similar patterns over time (reflecting the “year fixed effects”). While both Canada and the United States got rid of the death penalty at different times, both experienced the same increase in homicides in the late 1960s. Then, both countries experienced substantial crime drops in the 1990s, but Canada did *not* reintroduce the death penalty.

The story that Ehrlich tried to tell breaks down when one adheres to higher methodological standards of establishing causal proof by looking at panel data. If the upturn in executions in the United States in the late 1990s had caused the drop in homicides, then we would not expect to see a similar drop in homicides in Canada, which did not restore the death penalty. Panel data studies rank higher than time-series analyses in the above methodological hierarchy since they allow a comparison of the murder rates in the

treatment (United States) and control (Canada) jurisdictions. The intuition behind Figure 2 is clear: one could not attribute the United States drop in crime to the return of the death penalty if Canada enjoyed the same drop without restoring the death penalty.²⁰

Figure 3 offers another illuminating panel data-like comparison. It depicts the time paths of homicide rates for the states that varied the rate of executions over this period (the treated group) and the states that never had the death penalty over the same period (the control group). Note that the non-death penalty states had lower murder rates on average than the death penalty states (again a “state fixed effect”). Importantly, the two sets of states follow extremely similar paths with parallel upticks and downturns despite the fact that one of these sets of states did not have the death penalty at any point from 1960 to 2000 (that is, the two groups have similar “year fixed effects”). Thus, the late 1960s jump in the murder rate in the states without the death penalty cannot have been caused by eliminating the death penalty since these states never had the death penalty (meaning nothing changed for these states). Moreover, the drops in crime in the 1990s were similar in both groups even though one set of states had a death penalty and the other set did not. Figure 3’s treatment-control comparison, which is lacking in the time-series analysis, shows once again that the role for a causal influence for the death penalty on homicide is far smaller than the overall drop in the national time series—strike 3 for Ehrlich.

The panel data-like comparisons for Canada versus the United States and for abolitionist states versus the remaining states help illuminate the drawbacks of time-series studies. Simply comparing national rates of execution

20. Of course, it could be the case that during the 1990s Canada did other things that the United States did not do to try to lower their crime rate without reintroducing the death penalty. But the data do not support this hypothesis for the two obvious anti-crime measures: more prisons and more police. While the U.S. incarceration and police employment rates are estimated to have *increased* by 57 and 14% during the 1990s, these rates actually *decreased* in Canada—by 6 and 10%, respectively (Zimring, 2006; Donohue, 2007). Thus, the statistical controls that would be implemented in a full panel data analysis would actually make the Canadian drop in crime—achieved without any executions—look even more pronounced than the graph depicts. This underscores once again that while the raw graphs may provide insight and have strong intuitive appeal, the final resolution of complicated causal questions must rest on a more probing statistical evaluation than graphs alone can provide. Here, the empirical reality undermining deterrence of the death penalty is even stronger than the graphs suggest.

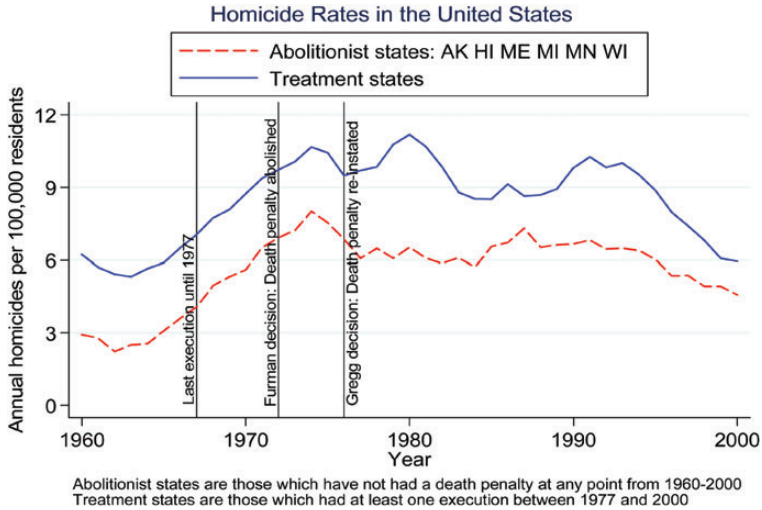


Figure 3. Homicide Rates in Abolitionist and Executing States.

Note: This graph previously appeared in [Donohue and Wolfers \(2005\)](#). It is reprinted here with the permission of the *Stanford Law Review* and the Board of Trustees of the Leland Stanford Junior University.

and homicides will be much less informative on the issue of deterrence than linking executions with changes in homicide to see whether areas that do not have the same pattern of executions are experiencing the same changes in homicides. If there is no *difference* in the patterns in the treatment and control groups (the *different* groups), there is no basis for claiming that the death penalty influences homicide rates (hence the terminology of a “difference in differences” study).

After his time-series work was criticized, Ehrlich published a second paper in 1977 based on a cross-sectional analysis of murder rates across states for 2 years—1940 and 1950. Figure 3 hints why the cross-sectional analysis—looking across states at a single point in time—is not a promising approach to establishing the causal influence of the death penalty. On its face, the figure indicates that the murder rate is *higher* in death penalty states. While this might seem to refute the deterrence hypothesis, a causal attribution is difficult when there are many factors that influence crime. All we know from this figure is that on balance the various influences on crime that lead to lower murder rates in the abolitionist states outweigh any

conceivable deterrent effect since non-death penalty states have lower rates of murder. Perhaps—one might conjecture—even though there are some enduring features that make death penalty states more murderous, the death penalty actually helps to suppress the murder rate below the level it otherwise would be. The panel data model recognizes that there are such stable differences between states that are hard to capture with specific variables but that can lead to a higher or lower murder rate. These enduring traits are captured in the “state fixed effects” variables. The cross-section study cannot control for such traits, and therefore it is a less reliable guide to establishing causal relationships—hence its lower ranking in the methodological hierarchy.

In light of these methodological deficiencies, the claim of [Posner \(2006\)](#) that “[e]arly empirical analysis by Isaac Ehrlich found a substantial incremental deterrent effect of capital punishment” should be understood to mean only that this was Ehrlich’s conclusion, not that this conclusion was warranted. In fact, Ehrlich used a weak tool and needed a particular logarithmic specification estimated over an idiosyncratic period to generate that conclusion. As the comparisons in [Figures 2 and 3](#) suggest, better panel data tools would have produced no evidence that executions deter homicide. Of course, one needs to do a more comprehensive statistical assessment of the panel data than the simple graphs of [Figures 2 and 3](#) can provide, but when this more comprehensive statistical treatment is done correctly, the deterrence claim for the penalty evaporates, as I discuss in further detail below.

While Ehrlich’s work misled a generation of readers about the impact of the death penalty and might conceivably have improperly contributed to the Supreme Court’s resurrection of the death penalty in 1976, it also had some positive effects. First, it stimulated others to march down the empirical path, thereby encouraging one of the great academic shifts in law and economics over the last four decades. Secondly, it heralded—perhaps prematurely—that econometric techniques could be used to address important questions of law and policy. Ehrlich’s work also serves an important pedagogical function in establishing the baseline from which impressive improvements in empirical estimation have occurred since the mid-1970s. Ehrlich’s papers appeared in the *American Economic Review* and the *Journal of Political Economy*—two of the most eminent journals in economics. Very few

empiricists in economics or law were using panel data at the time Ehrlich was writing, and few appreciated its superiority over the approaches that Ehrlich employed. Today, no AER or JPE editor would even send a time-series or cross-section study of law or policy to a referee for review. The paper would be rejected summarily, because superior panel data are available, and there would never be a reason to rely on a lower rung of the methodological hierarchy when higher quality approaches are available at reasonable cost.

What were jewels in the 1970s (at least in the eyes of some top journal editors at that time) would essentially be deemed worthless by the empirical standards of today. Consequently, we should relegate Ehrlich's work to the same place in the history of science that Kelvin's work on the age of the earth resides. If a paper could not be published today because its inherent weaknesses have been revealed, then it should no longer be offered in support of the conclusion that it once (unjustifiably) reached. There is progress in science, and work that falls short of the improved, contemporary scientific standards is no longer worthy of any credence in public policy debates. Unfortunately, as we will now see, that lesson has yet to be fully learned.

4. The Tendentious Use of Empirical Work

4.1. Exaggerated Findings That the Death Penalty Deters

One consequence of the empirical revolution is that advocates want to cite supportive empirical studies to sway decision-makers and the public. A now discredited study, such as Ehrlich's, can appear in the advocate's arsenal as though it were a state of the art finding instead of an exemplar of the more primitive econometric methodology of the past. Take for example David Muhlhausen from the Heritage Foundation, who spoke in 2007 in a Senate Judiciary Committee Hearing about the death penalty and related empirical studies. In this hearing, Muhlhausen said, "the recent studies using panel data techniques have confirmed *what we learned decades ago*: Capital punishment does, in fact, save lives. Each additional execution appears to deter between three and 18 murders" (emphasis supplied) ([Muhlhausen, 2007](#)). In fact, we now know that nothing was learned about the deterrent effect of the death penalty from Isaac Ehrlich, just as nothing was learned

about the actual age of the earth from Kelvin's work—other than to show that their methodologies could not shed light on their respective topics.

But Muhlhausen also refers to modern panel data studies, which we have seen reside at a higher place in the methodological hierarchy. What do they show? Despite the debilitating weaknesses in the studies claiming deterrence, various scholars have repeatedly testified to Congress that they point with near unanimity toward a finding that the death penalty deters homicide.²¹ Muhlhausen fails to reference every study disputing a finding of deterrence for capital punishment, including the following articles finding no evidence of deterrence that appeared prior to his testimony: [Katz et al. \(2003\)](#), [Donohue and Wolfers \(2005\)](#), [Berk \(2005\)](#), [Fagan \(2006\)](#), and [Fagan et al. \(2005\)](#). Furthermore, there is no reference to the critiques of the studies that Muhlhausen does rely on by [Donohue and Wolfers \(2005\)](#) and [Berk \(2005\)](#). Indeed, the list of articles finding no deterrence has only continued to grow since his testimony: [Zimring \(2008\)](#), [Cohen-Cole et al. \(2009\)](#), [Donohue and Wolfers \(2009\)](#), [Hjalmarsson \(2009\)](#), [Kovandzic et al. \(2009\)](#), and [Zimring et al. \(2010\)](#). Indeed, in 2012, a panel of the National Research Council concluded that none of the studies that Muhlhausen cited as evidence of a deterrent effect of the death penalty were reliable ([National Research Council, 2012](#)).²²

Muhlhausen's recitation of the empirical evidence illustrates one of the negative effects of the empirical revolution—studies can be churned out, cited, and relied on in the corridors of power regardless of their

21. Consider both Joanna Shepherd's 2004 congressional testimony before the Subcommittee on Crime, Terrorism, and Homeland Security ("In the economics literature in the past decade . . . there is a very strong consensus . . . all of the modern economic studies in the past decades have found a deterrent effect") as well as remarks by [Rubin \(2006\)](#) before the Senate Judiciary Committee ("The literature is easy to summarize: almost all modern studies and all the refereed studies find a significant deterrent effect of capital punishment. Only one study questions these results") ([Hearing before the Subcommittee on Crime, Terrorism, and Homeland Security, 2004](#)).

22. In addition, Muhlhausen's one-sided review cited the panel data study by [Ekelund et al. \(2006\)](#) as supportive of deterrence, ostensibly because the paper did assert that each additional execution was associated with a reduction in homicides. But Muhlhausen failed to mention that the Ekelund model also included a control for the presence of a capital punishment regime, which had the opposite sign and overwhelmed the execution effect. Correctly interpreted, the Ekelund study actually undermines rather than supports the deterrence hypothesis [Donohue and Wolfers \(2009, p. 256\)](#).

methodological soundness. What the scholarly review by the august National Research Council properly deemed to be an absence of evidence can appear in the hands of advocates and the press to be overwhelming econometric support. It has never been easier for ideologues and those with an ax to grind to employ weak tools in order to support dubious propositions with methodologically unsound approaches, and for advocates to cite to a mountain of arcane and seemingly sophisticated econometric studies that have no scholarly value in ascertaining true causal relationships.

4.2. Using the Media to Create Weak Studies and Vouch for their Soundness

This lamentable process unfolded once again following the September 25, 2007 U.S. Supreme Court grant of certiorari in *Baze v. Rees*, which was a challenge to the constitutionality of lethal injection. The pro-death penalty forces swung into action. On November 2, 2007, the *Wall Street Journal* published a national time-series study claiming that every execution saves 74 lives—far more than the paltry eight that Ehrlich had offered 30 years earlier ([Adler and Summers, 2007](#)). The two Pepperdine professors who authored this article had never written anything on the death penalty or on criminal justice and did not publish their research in any academic journal. The entire study was contained in the eleven-paragraph, editorial page article and used a national time-series analysis comparing the number of murders and executions in the nation over the last 25 years to identify the effect of the death penalty. Despite the fact that a 1978 National Research Council report had “conclude[d] that the results of the [Ehrlich time-series] analyses on capital punishment provide[d] no useful evidence on the deterrent effect of capital punishment” ([National Research Council, 1978](#), p. 62), a marketing professor (Dr. Adler) and his management professor colleague (Dr. Summers) were able to quickly generate a new national time-series study that was even less credible than Ehrlich’s, yet able to gain national prominence and influence as the public and decision-makers proved wholly unable to assess scholarly merit. Note that a phone call to an assistant professor—indeed to almost any empirically trained graduate student—at any top economics department would likely have been enough to alert the WSJ of the unreliability of the Adler–Summers work.

The web page of the National Center for Policy Analysis (NCPA) suggests the strategy behind the Adler and Summers piece, which it simultaneously posted at the time of the WSJ publication. NCPA describes itself as “a nonprofit, nonpartisan public policy research organization” whose “goal is to develop and promote private alternatives to government regulation and control, solving problems by relying on the strength of the competitive, entrepreneurial private sector.” NCPA’s webpage also explains its approach to influencing policy:

Ideas tend to filter through a hierarchy. They start in the realm of intellectuals. Through conferences, speeches, briefings and reports written for lay readers, the audience expands. The ideas begin to appear in newspaper editorials. Special interests may find an idea to their liking and help it along. Gradually, more and more people become aware of it. Politicians are often the last to climb on board. Still, it’s a process that has been repeated again and again.

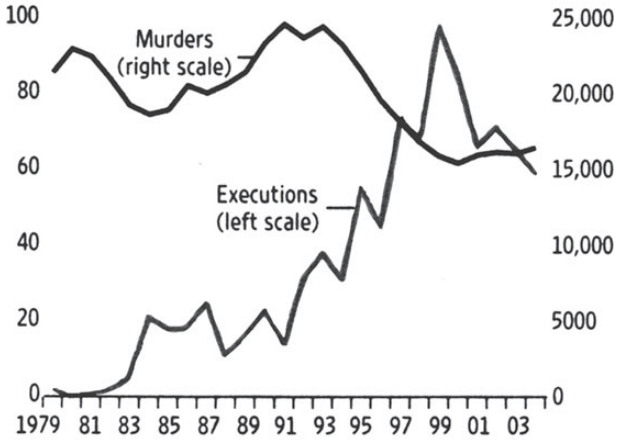
NCPA realized that the coin of the realm today is an empirical study that, whatever the reality, looks “scholarly” so that it can be used to rally the relatively uninformed public and influence governmental decision-makers.

Figure 4 captures the heart of the Adler and Summers article, which would have you believe that the rise in the death penalty in the 1990s explains the massive drop in murders that occurred during the Clinton years. One should now be able to discern the misleading nature of Figure 4 from our earlier discussion of the panel data-like graphs in Figures 2 and 3, since Adler and Summers use the time-series dimension to generate the striking claim that 74 lives are saved per execution. Despite the impressive visual appeal of their graph (underscoring that some graphs illuminate while others deceive), our methodological hierarchy would dictate that their time-series study should be dismissed outright since panel data studies are available. However, this is information that many readers are not aware of, and the WSJ certainly made no effort to inform them. This op-ed took advantage of the lack of expertise of decision-makers and the public in order to convince them of the legitimacy of particular political positions for which no sound empirical support existed.²³

23. When I mentioned this campaign to Gary Becker, he responded that “the left does the same thing.” The job of the academic empiricist is to maintain high standards for empirical studies regardless of whether the study is endorsed by the left or right (and regardless of whether the researcher falls on the left or right).

The Deterrence Effect

Relationship between executions and murders,
1979-2004



Sources: Roy Adler and Michael Summers

Figure 4. The Adler and Summers Wall Street Journal Graph.

With the faulty study launched, the next step in ensuring its widespread acceptance would be for some seemingly knowledgeable experts to vouch for its findings. At this point, ABC7 stepped in with an article “Can executions decrease murder rates?” which stated “that a new study [by Adler and Summers] is adding to the long-running debate on whether the death penalty is a deterrent to crime. Researchers in Southern California claim it is—capital punishment works” (Lee, 2008). ABC7 larded praise on the Adler–Summers study from the head of the Institute of Criminal Justice: “As far as the study itself goes, the research, design, the analysis, interpretation, they all appear to be valid. The death penalty is statistically a deterrent to the crime of murder.” Of course, this was highly inaccurate. Every methodological aspect of the study was wanting, but the apparent strategy paid off.

When the U.S. Supreme Court decided the lethal injection case of *Baze v. Rees*, 553 U.S. 35 (2008), it stated: “there has been a recent surge in scholarship asserting the deterrent effect of the death penalty, see, e.g., Mocan & Gittings, Getting Off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment, 46 J. Law & Econ. 453 (2003); Adler &

Summers, Capital Punishment Works, Wall Street Journal, Nov. 2, 2007, p. A13” ([Supreme Court of the United States, 2008](#)). While the [Mocan and Gittings \(2003\)](#) study, despite its flaws, was an academic panel data analysis published in a peer-reviewed journal, the Adler and Summers study did not merit citation by the WSJ, let alone by the U.S. Supreme Court.²⁴ This is the nightmare scenario of the empirical revolution, where studies that would be clearly recognized as invalid by serious academics can be quickly injected into the public discourse, and the help of think tanks and the media can generate enough prominence for the flawed work to lead to citation in an opinion of the U.S. Supreme Court.²⁵

5. Evaluating the Theory and Empirics of the Death Penalty

5.1. Ideology Blinds Certain Academics

It is perhaps unsurprising that media outlets and special interest think tanks would ignore the importance of econometric methodology and economic principles in championing studies, and that the judiciary—including the Supreme Court—might struggle to sort out the wheat from the chaff. But even elite academics can be blinded by ideology in a way that undermines their critical judgment, as I learned during a talk I gave a number of years back at Gary Becker’s workshop at the University of Chicago. As I discussed the fragility of studies claiming a deterrent effect of capital punishment, a University of Chicago economics professor exclaimed, “Come on, we know the death penalty deters. If you had the death penalty for illegal parking, no one would do it.” The remark is badly misconceived.

24. Table 2 in [Donohue and Wolfers \(2009\)](#) shows that the Mocan and Gittings OLS finding of deterrence was the product of a coding error, and their 2SLS finding of a statistically significant deterrent effect (which Donohue and Wolfers argued used invalid instruments) was eliminated when the standard errors were clustered (as they should be).

25. This discussion underscores that some graphs are revealing while others are misleading. Figures 2 and 3, which capture the insights of valid panel data models, illuminate that the scope of any influence of the death penalty on murder rates in the United States is small. Figure 4 suggesting an enormous impact of the death penalty (as murders fell from 25,000 to 15,000) is misleading (based on its reliance on a limited national time-series analysis). Much complex statistical analysis undergirds the conclusions about why the first two graphs are meaningful (see discussion in note 20, above) and Figure 4 is not.

Though, of course, it is true that illegal parking would decline if the penalty for illegal parking were raised from, say, a \$15 fine to execution, this is irrelevant to the issue of whether the death penalty in America deters homicide. The real question the professor should have been considering is: if the penalty for illegal parking were life in prison without the possibility of parole, would illegal parking decline further if we raised the penalty to include capital punishment? One expects most University of Chicago economists to understand the importance of marginal analysis and the law of diminishing returns but there is something about the affection for capital punishment that seems to eliminate the capacity for critical thought. The Chicago economist's comment might have made sense if a convicted murderer who did not receive a sentence of death would otherwise face a relatively favorable prospect—perhaps only a few years in jail. But escaping the death penalty is no picnic—in 2005, roughly 132,000 US inmates had life sentences and 37,000 of these had no chance for parole. Did killing another 1,250 in the last 30 years have an additional deterrent effect beyond the alternative of life without possibility of parole? That is a very different question from that posed by the initial parking violation hypothetical. Again, an empirical analysis must be structured to reflect the correct theoretical framework if it is to have any chance of reaching a meaningful conclusion.

While one should not ordinarily attribute much significance to stray comments in a seminar, the remark about illegal parking is illuminating in that it shows the power of ideology to dislodge even a member of the University of Chicago economics department from the customary centrality of marginal analysis in modern economics. The hypothetical is also economically uninformed from a second perspective to the extent that it fails to consider the costs of the death penalty versus other possible crime-fighting expenditures. One of the defining characteristics of the modern death penalty in the United States is its high cost. For example, over the last 35 years, the state of California has spent roughly \$4 billion to execute thirteen individuals (Alarcon and Mitchell, 2010). Even if it could be proved that these thirteen executions *each* deterred, say, eighteen murders (the *high-end* estimate offered by Muhlhausen, above, from the flawed and discredited panel data models), this would not be a sign that the death penalty reduced the rate of murder. Given the average cost of a California

police officer of \$86,040, the \$4 billion could have been used to hire roughly 46,490 police officers which, if appropriately assigned, would be expected to prevent 552 murders (and much other crime) in California—considerably more than twice the number saved under even the most wildly optimistic (albeit discredited) view of capital punishment.²⁶ In other words, since the death penalty is a costly and inefficient system, its use will lead to more murders by wasting resources that could be expended on crime-fighting measures that are known to be effective. It might seem that our hapless Chicago economist not only missed the class on marginal analysis but also on the need to consider both costs and benefits. But the problem is not lack of knowledge, but rather presence of ideology, which blinds even the most astute to evidence that would threaten cherished priors.²⁷

5.2. Panel Data Evidence

At this point, it is worth turning to the panel data evidence concerning the issue of whether the death penalty deters murder. Probably the best ordinary least squares (OLS) panel data study to address this question was [Kovandzic et al. \(2009\)](#), which found no support for the view that executions or death penalty laws reduce murder. [Donohue and Wolfers \(2005, 2009\)](#) also found that the other OLS panel data studies that purported to find a deterrent effect failed to do so when a variety of errors were corrected. But

26. This increase in the police force would be about a 68% increase in the number of police and sheriff's patrol officers in California. ([Bureau of Labor Statistics, 2013](#)). With an elasticity of crime with respect to police of 0.30 ([Klick and Tabarrok, 2005](#)), one would expect to save about 552 lives per year from a base of 2,707 murders in California (the average number of murders over the period 1978–2013, [Harris, 2013](#)). This is considerably more than twice the number of lives saved by the California death penalty under the unrealistically optimistic estimate of 18 lives saved per execution ($13 \times 18 = 234$), emanating from the discredited panel data studies that the NRC panel found to be uninformative. (Other states have less expensive death penalty regimes than California, but their police are also substantially less costly: the Bureau of Labor Statistics, *supra*, indicates that average police costs outside California are only \$58,720 per officer.) Furthermore, unlike the death penalty, which would not deter lesser crimes, adding police would also lead to significant drops in all crime categories. Note that the Klick and Tabbarok paper was specifically cited by the NRC panel on the death penalty as a study generating reliable estimates of the impact of added police on crime.

27. This provides an argument for seeking comments from those inclined to disagree with one's conclusions since they may be more attentive to weaknesses in a study than those who are pleased with its findings. Sadly, though, this often becomes an opportunity to kick sand in the air rather than to really identify valid deficiencies.

the NRC panel on deterrence and the death penalty concluded that all of the death penalty studies, including Kovandzic et al., were “not informative.” That word choice is unfortunate since in my view it is clearly worth knowing that the best OLS panel data models do not support the deterrence hypothesis. This evidence is not conclusive; indeed, a single statistical study cannot be dispositive without the development of a supporting literature.

A panel data study still may fail because of omitted variable bias or endogenous adoption of the death penalty, but knowing that the best available panel data study shows robust results that have withstood serious testing from opponents is certainly informative. It will obviously be a judgment call as to when a careful panel data study will be sound enough to warrant consideration in important public policy debates, and it is wise to set a high bar. The question for the NRC panel is whether it could have been more discriminating in judging some studies—perhaps those that only generated statistically significant results because of a failure to cluster standard errors or those using clearly invalid instrumental variables—as clearly below the bar, and others as being closer to the “best available” standard regardless of their defects.²⁸

The NRC panel hoped that a methodology superior to OLS panel data analysis could be brought to bear on this question. The concern is worth addressing because of the possibility of bias in the OLS estimates that could be remedied by the use of yet another methodological advance—an instrumental variable approach. This approach is highly desirable if one has a good instrument—which we will define momentarily—because one could then hope to secure better (unbiased) estimates of the impact of the death penalty. But it is no surprise to anyone in this field that it is

28. Of course, the NRC panel was correct in noting that the failure to reject the null hypothesis (of no deterrence) does not establish the truth of the null hypothesis. Indeed, it is much harder to establish a null effect because this would require such a small standard error that one could confidently establish a zero effect. But the Supreme Court has weighed in on this issue in stating that unless the imposition of the death penalty “*measurably* contributes to one or both of these goals [deterrence and retribution], it ‘is nothing more than the purposeless and needless imposition of pain and suffering,’ and hence an unconstitutional punishment.” *Atkins v. Virginia*, 536 U.S. 304, 318-19 (2002) (emphasis supplied). If years of intense effort by highly motivated researchers have not produced a single credible study that supports the deterrence effect of the death penalty, it seems clear that there is no basis for saying that the death penalty “measurably contributes” to deterrence.

extremely difficult to find a valid instrument for a crime policy intervention like the death penalty. Moreover, using an instrument that is inadequate or invalid *cannot* yield valid estimates—such estimates clearly would be “uninformative.”²⁹

5.3. Are the OLS Panel Data Results Biased?

But before turning to a discussion of instrumental variables, we can also make some headway by considering the type of bias that might infect OLS panel data estimates, which was clearly an appropriate concern for the NRC panel. As we have seen, the resurrection of the death penalty after the moratorium of the 1970s blossomed into a large number of executions in the 1990s, largely driven by political influences. Indeed, it was the crime spikes of the late 1980s that set the stage for this punitive response. This pattern suggests the first type of potential bias in OLS estimates of the impact of the death penalty over the last 30 years: the spiking murder rate led to a jump in executions followed by the likely mean reversion as crime returned to more customary levels. This, of course, would make the death penalty look *more* effective in reducing crime than in fact it was.

Secondly, a “get tough on crime” attitude might lead to longer jail sentences, increased use of life without parole, harsher prison conditions, and more aggressive policing, as well as increased use of the death penalty. This was indeed the response to the crime rises of the 1980s. Without adequate controls for these other factors, OLS estimates will make the death penalty look better than it is (a classic form of omitted variable bias). The NRC panel recognized this potential bias in the existing death penalty studies but it failed to appreciate the importance of the fact that the major potential problems with OLS estimates of the impact of the death penalty may well

29. It is important to realize that an invalid instrumental variables analysis tells you nothing and indeed may be particularly dangerous in creating a false sense of precision. In contrast, a well-done OLS panel data analysis is certainly worth doing—at the very least, one can say that “controlling for all of the factors that we can quantify and using our best available panel data models, we find X,” while noting whatever weaknesses might exist in the analysis. This analysis will be even more informative when one has a legitimate basis for evaluating the nature and sign of any possible biases if they are likely only to strengthen the initial finding. As Keynes noted, “it is better to be roughly correct than precisely wrong.”

be that they are biased in *favor* of finding deterrence.³⁰ But if the best OLS panel data studies show no deterrence, even when the biases are operating in *favor* of a finding of deterrence, one may be able to draw two conclusions: (1) the OLS findings of no deterrence are more compelling than the skeptics might first believe, and (2) if a good instrument could be found, it would only serve to *strengthen* the OLS findings of no deterrence rather

30. The NRC panel agreed with the [Donohue and Wolfers \(2009\)](#) discussion that some studies were not precise in thinking about the execution risk that actually confronted potential murderers. For example, the studies measuring this risk by using an identifier for whether the state had a death penalty law meant that New York was categorized as a death penalty state for a number of years after 1995, even though no one was ever executed before the death penalty was eliminated. Moreover, the District Attorney of Manhattan specifically announced that he would not seek the death penalty yet the enormous drop in murders occurring in Manhattan after 1995 was then in part attributed to the death penalty. Did the criminals take cognizance of the death penalty law or the conflicting announcement of the Manhattan DA? Certainly, the studies did not consider this issue. Still, one would imagine that the risk of execution would at least be related to the frequency of executions, and again we see no evidence that this rate of execution influences the rate of murder in the best panel data studies.

One could also ask whether the omitted variables such as the other harsh anti-crime measures that were being adopted in some states might undermine the claim that the bias runs in favor of deterrence. In other words, might the states that were the most aggressive in using the death penalty be less aggressive in other measures (while the states without the death penalty made up for this lack by being particularly aggressive in other anti-crime dimensions)? I have never seen any evidence to that effect in the United States (or anywhere else for that matter). Harshness in one domain tends to carry over across all domains since the same dynamics that produce it in one arena tend to operate across the board. Indeed, a panel data model from 1979 to 2000 shows that as executions rise in a state, the rates of police and incarceration rise compared with other states (consistent with the thesis of uniform harshness that is suggestive of bias in favor of a finding of deterrence). A potentially more plausible modification of the challenge to the bias point is that some of the non-death penalty states, such as New York, might have implemented more effective anti-crime strategies, and it is hard to capture police strategies and effectiveness in lengthy panel data studies (as the measurement issues are vexing).

These issues underscore the considerable challenge of knowing when a study is credible enough to be informative. The NRC report did point to the very good paper on police and crime by [Klick and Tabarrok \(2005\)](#) as being in this category, and I agreed with that conclusion. But if one applies a very exacting standard of not being able to identify a possible weakness in a study before a paper is above the credibility threshold, then one might even conclude that study was not informative. Klick and Tabarrok showed that an elevated terror alert led to more police patrolling the streets and substantially lower crime. But if the heightened terror risk led people to stay home (an issue the authors address in citing mid-day subway ridership numbers), the drop in crime could be the product of citizen apprehension rather than police effectiveness. In the end, one needs to make very nuanced judgments about what type of evidence is persuasive.

than undermine them. The corollary of the second conclusion is that if an instrumental variable approach generates a finding of deterrence that does not appear in the OLS studies, it is likely to be an invalid instrument that is picking up the influence of some other factor that influences crime rather than the impact of the death penalty.

5.4. Instrumental Variables

5.4.1. *Definition and assessing validity.* With this by way of background, we can now turn to an assessment of the instrumental variables estimates of the impact of the death penalty. But here our expectation that this approach will likely strengthen the finding of no deterrence—at least if the instruments are valid—conflicts with the conclusions of these instrumental variables studies. Reviewing a few of them, [Sunstein and Vermeule \(2005\)](#) wrote that they are “powerful,” “impressive,” “sophisticated multiple regression studies” showing that “capital punishment powerfully deters killings.” Our corollary above suggests a different interpretation: these studies are truly uninformative because they are not using valid instruments (as the NRC panel also found).

Note a valid instrument has a very precise definition: it must be a variable that is highly correlated with the variable of interest (say “the number of executions in a given state and year”) but does not influence crime except through its relationship with the number of executions. A nice illustration of a conceptually attractive instrument for discerning the impact of police on crime is provided by [Levitt \(1997\)](#), who uses mayoral elections as an instrument. This factor influences the number of police (as mayors increase the police presence more in mayoral election years to curb crime), but does not influence crime except through its influence on the number of police (assuming the mayor is not telling the police not to count the crimes that do occur).³¹ These two conditions establish the validity of the instrument for estimating the effect of the police on crime. As we explore the instruments

31. Even this wonderfully illustrative (and innovative!) instrument is not beyond question as the parenthetical suggests—there certainly is evidence of police efforts to cook the books to make crime look better. Moreover, to establish instrument validity one would have to ensure that sheriffs, prosecutors, and judges are not up for election at the same time as the mayors and using their power to shrink crime in ways other than through the increase in the police presence.

used in the death penalty studies, we must ask if they plausibly conform to the definition of a valid instrument. If the answer is no, then the study simply *cannot* yield informative estimates.

5.4.2. *The validity of the DRS instruments.* Let us begin with the main instrumental variables study relied upon by [Sunstein and Vermeule \(2005\)](#), which is [Dezhbakhsh et al. \(2003\)](#). In this particular paper, the authors follow Ehrlich’s lead in positing that homicide rates are a function of three deterrence variables: the probability of execution conditional on a death sentence, the probability of a death sentence conditional on arrest, and homicide arrest rates. Their main regression (for their preferred model 4) is as follows:

$$\begin{aligned} & \frac{Murders_{c,s,t}}{(Population)_{c,s,t}/100000} \\ &= \beta_1 \frac{Homicide\ Arrests_{c,s,t}}{Murders_{c,s,t}} + \beta_2 \frac{Death\ Sentences_{s,t}}{Arrests_{s,t-2}} \\ &+ \beta_3 \frac{Executions_{s,t}}{Death\ Sentences_{s,t-6}} + \gamma_1 \frac{Assaults_{c,s,t}}{Population_{c,s,t}} \\ &+ \gamma_2 \frac{Robberies_{c,s,t}}{Population_{c,s,t}} + \gamma_3 County\ Demographics_{c,s,t} \\ &+ \gamma_4 County\ Economy_{c,s,t} + \gamma_5 \frac{NRAMembers_{s,t}}{Population_{s,t}} \\ &+ \sum_c County\ Effects_c + \sum_t Time\ Effects_t + \eta_{s,t} + \varepsilon_{c,s,t} \end{aligned}$$

While this equation may look daunting, it merely says that the murder rate is a function of the three mentioned deterrence variables (rates of homicide arrest, death sentencing, and execution) and then a number of other measures designed to capture the current level of crime (assaults and robberies), as well as the demographic and economic aspects of each county and the prevalence of NRA members. The last two variables in the final line of the equation simply refer to the geographic (here county) and year fixed effects that we have previously spoken of in discussing the state panel data analyses.

But what instruments do the authors bring to this estimation? Is there any plausibly exogenous factor such as Levitt’s mayoral election year variable that can serve as an instrument for the number of police? Sadly, the answer

is no. The instrumental variables Dezhbakhsh, Rubin, and Shepherd (DRS) offer are state-level nominal police payroll, nominal judicial expenditures, Republican vote shares in presidential elections, and prison admissions. While I could believe that these variables are correlated with executions in a given state and year, it seems highly implausible that they would influence crime *only* through death sentences and executions. Indeed, one expects the police and the level of incarceration to have major direct effects on rates of murder and all of these direct effects will show up in the instrumental variables estimates that are supposed to be capturing *only* the impact of the death penalty. Moreover, to the extent that Republicans pushed for tougher anti-crime sanctions in the 1990s, the share of the Republican vote is capturing more than just the increased use of the death penalty but all of the tougher measures of policing and sentence enhancement that would influence crime in ways other than through the impact of the death penalty. In short, the instruments appear invalid on their face.³²

A further illustration of the problematic nature of these instruments is provided by showing that, DRS' results are extremely sensitive to small changes in the way the instruments are specified or in the data included in the regression. To see this, consider the below Table 1 from [Donohue and Wolfers \(2005\)](#): Panel A of Table 1 replicates DRS' results perfectly, showing the impact of deterrence variables on murder rates. Then, Panel B converts the results from Panel A to net lives saved per execution, which are easier to interpret. Model 4 in Panel B shows that DRS' preferred model purports to find that 18.5 lives are saved by each execution. We will now discuss how Panels C through E reveal the high degree of sensitivity of the original DRS results.

Panel C collapses the partisanship variables—the percent Republican vote share in the last Presidential election—into a single instrumental variable (rather than allowing for six separate variables based on each Presidential election). Amazingly, while Panel B had suggested that many lives would be saved per execution, Panel C suggests just the opposite: the death

32. Even if these instruments were not conceptually invalid, they seem to be improperly constructed. The police, judicial, and prison variables are not per capita numbers, but rather they are statewide aggregates, which is the incorrect form when trying to explain murder *rates*. In addition, the police and judicial expenditure numbers are not adjusted to account for inflation, while standard practice would do so. [Donohue and Wolfers \(2005, p. 821\)](#).

Table 1. Estimating Effect of Executions on Murder Rates and Net Lives Saved: Testing the Sensitivity of the DRS Estimates, 1977–1996

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

Note: This table previously appeared in [Donohue and Wolfers \(2005\)](#). It is reprinted here with the permission of the *Stanford Law Review* and the Board of Trustees of the Leland Stanford Junior University. 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.

penalty actually induces more homicides! The preferred Model 4 in Panel C possesses a similar magnitude to the estimate of Model 4 in Panel B but since the sign is negative, it means that instead of saving eighteen lives, this one tweak in the instrumental variable now implies that eighteen lives will be lost for each execution.

Panels D and E, which show estimates when Texas and California are dropped, reveal large and negative statistically significant coefficients

for DRS's preferred Model 4—meaning that every execution will lead to between twenty-nine and forty-two additional homicides. Looking across each of those rows, one sees highly variable results in both directions, ranging from forty-two lives lost to thirty-four lives saved. Given the illustrated sensitivity of these results, “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” (Donohue and Wolfers, 2005).

Table 2 from Donohue and Wolfers, reprinted below, provides an additional check on the robustness of the DRS instrumental variables estimates by first dropping the Republican vote share instrument (Panel C) and then dropping all of the instruments *except* the Republican vote share variables (Panel D). Once again, the estimated effect of the death penalty is seen to swing dramatically. In Panel C, the death penalty is estimated to cost a massive number of lives, while in Panel D the death penalty is deemed to save an utterly implausibly large number of lives. Indeed, the array of estimates in Table 1 and Table 2's Panels B–D reveal that one can simply select a particular model in a particular row and generate virtually any estimated effect of the death penalty that one wanted. Now one begins to see the dangers of the use of invalid instruments since they can make estimates bounce wildly. Selecting among this extremely variable set of estimates is a near-impossible task.

At this point, it should be clear that the DRS instruments are not generating robust and plausible estimates of the impact of the death penalty. Again, recall that for the DRS instruments to be valid they must not directly influence murder except through the channel of executions. Donohue and Wolfers (2005) test whether this requirement is truly met by examining whether the instruments impact the homicide rate when restricting the sample to those observations within states that did not have the death penalty. This test, in Panel B of Table 2, can be illuminating because it is not possible for the instruments to influence homicide rates through executions since there are no executions in the states used in this analysis. If the instruments are valid, they should have no influence on the murder rate where there are no executions (that is, the estimates in Row B of Table 2 should be small and statistically insignificant). As the table shows, however, Panel B generates large and highly statistically significant estimated effects in each of the six models. The simplest conclusion is that the instruments (or their

Table 2. Estimating Effect of Executions on Murder Rates and Net Lives Saved: Testing the Sensitivity of the DRS Estimates, 1977–1996

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

Note: This table previously appeared in [Donohue and Wolfers \(2005\)](#). It is reprinted here with the permission of the *Stanford Law Review* and the Board of Trustees of the Leland Stanford Junior University. 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.

^(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.

correlates) impermissibly impact murder rates *directly*—not only through the channel of executions. Once again, the validity of the instrumental variables is undermined.

Rubin responded to the Donohue and Wolfers’ challenge to the DRS instruments by observing that the instruments that DRS employed have been used to explain concealed gun laws ([Lott and Mustard, 1997](#); [Rubin and Dezhbakhsh, 2003](#)), truth-in-sentencing legislation ([Shepherd, 2002a](#)), California’s three strikes law ([Shepherd, 2002b](#)), and sentencing guidelines ([Shepherd, 2004](#)). Rubin explains that “[m]ost of our instrumental variables have been used in numerous empirical papers because previous researchers believed (often based on empirical testing) that the instruments were as uncorrelated with crime rates as one was likely to find” ([Rubin, 2006](#)). In Table 2, Panel B undermines the last claim of Rubin since the DRS instruments are in fact highly correlated with crime even where there

is no death penalty. Moreover, the claim that these instruments are valid because they are used to test other policy interventions is patently incorrect. If a variable like the percent Republican vote is used to assess the impact on crime of other policy interventions such as RTC laws, then it is being assumed that this variable influences crime only through its effect in encouraging adoption of RTC laws. But the same instrumental variable that influences crime directly *only* through its effect on RTC laws cannot simultaneously influence crime *only* through the operation of the death penalty. This is simply a logical impossibility. Rather than justifying the DRS instruments, Rubin's quote reinforces the point that one simply cannot use these instruments to estimate the impact of the death penalty on murder (Donohue and Wolfers, 2006).

5.5. Infirmitities in the Instrumental Variables Estimation

In *Baze v. Rees*, 553 U.S. 35 (2008), Justice Stevens cited research by Donohue and Wolfers in order to justify the claim that "there remains no reliable statistical evidence that capital punishment in fact deters potential offenders." Justice Scalia responded to Stevens by saying that Stevens' conclusions "are not supported by the available data." In support of this, Justice Scalia cited only the single article by Cass Sunstein and Adrian Vermeule that was mentioned above. Sunstein quickly responded to Scalia's citation, indicating in a piece with Justin Wolfers that his views had evolved: "In short, the best reading of the accumulated data is that they do not establish a deterrent effect of the death penalty" (Sunstein and Wolfers, 2008). In other words, the sole authority that Scalia relied upon in support of the deterrence argument now rejects that position.

To be a bit more precise, there is no valid statistical support for the proposition that the death penalty has ever had a net deterrent effect on murder in the United States over the period from 1934 to the present. The valid OLS panel data studies consistently reveal no relationship between the death penalty and the overall rate of murder. The *only* regression outcomes that superficially support deterrence are instrumental variables estimates using invalid instruments and highly problematic specifications. Moreover, these instrumental variables estimates—including the Mocan and Gittings study cited by the Supreme Court in *Baze*—are all statistically insignificant if one clusters the standard errors—as one should in panel crime data studies since,

otherwise, serial correlation in panel data can lead to major underestimation of standard errors (Wooldridge, 2010; Angrist and Pischke, 2009).³³

5.5.1. *Should priors dominate empirical evaluation?* Just prior to publication of Donohue and Wolfers (2005), Gary Becker wrote that “I favor the death penalty because and only because I believe it has a ‘sizeable’ deterrent effect . . . the preponderance of the evidence does indicate that capital punishment deters” (Becker, 2005a, 2005b). Many economists reflexively accept the deterrence claim adhering to price theoretic claims based on downward sloping demand curves. But even if that theoretical story were accurate, it would not tell whether it is ten lives saved per execution or one life saved per ten executions or per 10,000 executions. Unquestionably, empirical support is needed to claim an effect is “sizeable.” When the 2005 Donohue and Wolfers paper was published, Becker wrote to me and said that he liked the piece but still believed the death penalty “does deter, although far less than 10 to 1, or even 3 to 1. I argue that even if it is 1-1, and perhaps even less, it is worth doing.” Becker admits, “if the direct evidence on deterrence is murky, either one weights a lot one’s priors . . . or one says there is no effect. I prefer the former by a long shot.”

But encouraging people to fall back on relatively uninformed priors, often gives greater power to the ideologues and the uninformed than a searching examination of the empirical evidence would warrant. We certainly want to avoid the scenario that James Heckman has warned against where ideologues mask their preferences in theory and trot out weak empirical evidence as confirmation. As Heckman colorfully noted, “[r]igorous theory and bullshit empirical work can co-exist. It leaves the rigorous theorists to make up the numbers they want” (Scheiber, 2007). Academic research should entail a search for truth, not a rhetorical effort to fool others

33. Bertrand et al. (2004) recommend clustering the standard errors by state. Aneja et al. (2011, 2014) use placebo tests on panel crime data to show that, without clustering, randomly assigned explanatory variables will appear significant between 20 and 75% of the time (instead of the correct 5%). Not one of the panel data studies claiming to find a deterrent effect of capital punishment shows statistically significant results after clustering. Note that the clustering issue is similar to the focus of the Hal White paper we alluded to earlier as the most-cited article in economics in that they both are concerned with the proper estimation of standard errors, which is critical to correctly identifying statistical significance.

into believing what the researcher already believes based on inadequate evidence.

5.5.2. *A more attractive model for empiricists.* A more attractive model for empiricists can be found in Heckman's scholarship (Donohue, 2002). After winning the Nobel Prize, Heckman responded to a question by saying he was most proud of his work evaluating the impact of Title VII of the Civil Rights Act, which prohibited racial discrimination in employment, on black textile workers in the South. As a Chicago school economist, Heckman had a theoretical bias against government intervention that led him to set out to prove that Title VII had not helped African-American workers. After 10 years of study,

in which he applied his techniques that helped gauge how workers would have fared without the act, he surprised himself with his conclusion. "I truly did want to show the government didn't have an impact," Heckman said. "But I convinced myself it had a huge effect" (Liesman, 2000).

Heckman, like Kelvin, labored for many years on an important research question, but what made the difference between success and failure was that Heckman labored in pursuit of truth, while Kelvin struggled to promote a religious ideology. While ideology can profitably lead one to explore a certain question, it should not influence how one resolves the question.

Figure 5 reproduced from Heckman's *American Economic Review* article on the impact of law on the economic welfare of blacks illustrates employment shares by race and sex in the South Carolina textile industry in the period 1910–1977. Until the mid-1960s, the share of blacks in the textile industry remained constant at low levels "despite growth in the quality and quantity of black schooling and despite economic scarcity resulting from tight labor markets." It is in 1965, the year Title VII became effective, that the black share in the industry suddenly starts to improve. While many of his Chicago School colleagues would have liked to see Heckman conclude that black progress was due to a tight labor market, this was not what he found. After an exhaustive examination of county data, he documented that the state time series was not just capturing a tight labor market story but instead showed that the law had its intended effect on black economic welfare (Donohue and Heckman, 1991).

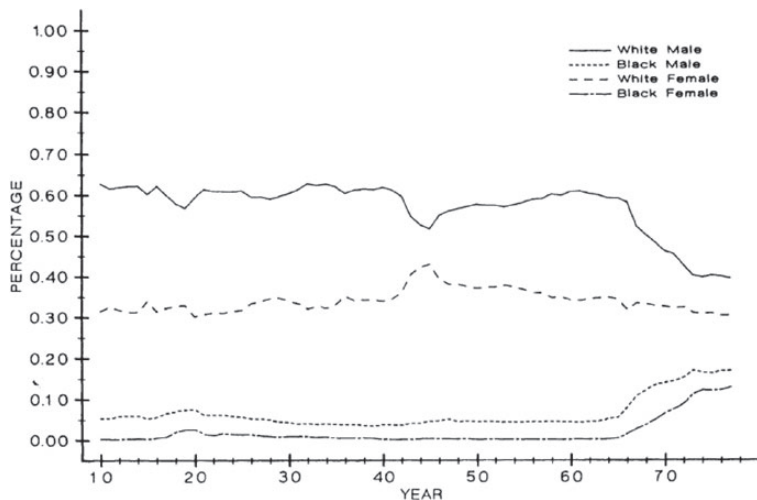


Figure 5. Aggregate Employment Shares in the South Carolina Textile Industry.
Note: This graph previously appeared in Heckman and Payner (1989). It is reprinted here with the permission of the American Economic Association.

In order to maintain an honest search for truth in the realm of empirical law and economics, researchers should conform their beliefs to the valid empirical evidence instead of forcing empirical evidence to conform to prior beliefs. When the evidence is murky, researchers should follow in Heckman's footsteps and push for better data and research approaches. There is an enormous difference between having a theoretical reason to believe something and having empirical support for that belief. Just as we would not want to cling to an estimate of the age of the earth that is off by 4.52 billion years, we do not want our decision-makers to be misinformed by those who would subordinate truth to maintain conclusions that support revered theories or that are procured to promote economic interests.

At the same time that we are striving to cultivate the goal of truth-seeking from individual researchers, we also need to shore up the social science institutions that will foster that goal. The growing trend among top journals to require data and computer code from all researchers should be encouraged. In addition, we must develop the norm that such information will be shared even if not mandated as a requirement of publication, and tenure and appointments committees and deans should recognize and reward those who comply with this practice. The empirical researcher's job is to create

public goods, and all researchers should make their data available online for anyone to use. This is typically avoided because of fears that errors will be discovered, but that is exactly why the practice is essential. Having errors found in data and code that a researcher has made available to all makes the sharing researcher part of the search for truth. Hoarding one's data to protect personal benefits increases the risks both that shoddy, unprincipled, or simply erroneous work will be published and rewarded and that the public will be misguided on important public policy issues.

6. Conclusion

Over the past 30 years, the field of law and economics has transformed dramatically from one dominated by theoretical pieces to one bursting with empirical studies. This change has generated a host of pathbreaking studies, and I assess how a hierarchy of methodologies has developed that can assist in evaluating the confidence one can repose in a particular study. The empirical revolution—sometimes referred to as the credibility revolution—is also shifting the nature of academic discourse. While in the past, sophisticated and rhetorically skilled arguments in support of particular beliefs have resided at the top of the scholarly hierarchy, increasingly it will be studies that credibly reveal important truths about the world in methodologically sound ways that will assume a position of dominance.

But there is a heated debate within the academy among elite empiricists about the standards of methodological soundness. All top empirical academics should recognize that the Gius study on assault weapons as well as the studies by Ehrlich or Adler and Summers on the death penalty are not worthy of credence, but I have already alluded to the disputes over far more solid empirical work that is still rejected by purists who are unwilling to see any value in empirical studies that are not in category 2 or above (that is, randomized studies, natural experiments, or regression discontinuity studies). Thus, the NRC report of 2012 on deterrence and the death penalty stated that even the very well-done OLS panel data study by [Kovandzic et al. \(2009\)](#) or the [Zimring et al. \(2010\)](#) study comparing the murder rate in Singapore (which ratcheted up its use of the death penalty) with that in Hong Kong (which abolished it at the same time)—both of which found no evidence of deterrence—were “not informative.” This may be the purist

fallacy of letting the perfect be the enemy of the good: if it were really true that such studies were not informative, one would have to conclude that perhaps 95% of published empirical work in law and economics and virtually every empirical study introduced in litigation should not be undertaken.

To see how revolutionary the “credibility revolution” can be, consider the most prominent work of Andrei Shleifer of the Harvard economics department, which concludes that common law systems are more successful at promoting prosperity than civil law systems (La Porta et al., 2008). Although this has been among the most influential and most cited work in all of economics, Jonathan Klick rejects it as inconsistent with the demands of modern econometrics: “This kind of cross-sectional comparison has no chance of sorting out these issues, and conclusions based on this analysis are close to worthless in terms of having confidence in causality” (Klick, 2013, p. 908). Heads sometimes roll during major revolutions. But who are we to believe—the revolutionaries or the thousands of scholars who have cited Shleifer’s work? This is a difficult challenge for those seeking answers to these important substantive questions.

Klick observes that “Modern empirical microeconomic work focuses on what are referred to as natural or quasi-experiments, where the researcher attempts to exploit seemingly random variation that affects the policy of interest” (Id. at 907). Such natural experiments provide wonderful vehicles for estimating true causal effects, but they provide cold comfort to an expert witness in litigation or consultant or academic who is asked to opine on an important policy issue but only has observational data at his disposal and cannot wait around for a natural experiment to occur. There are levels of quality in studies of observational data that must also be acknowledged if good work is to be recognized and relied upon and poor work is to be rejected.

The tremendous advances that have occurred in our understanding of estimating causal effects have illuminated how frequently ostensibly sophisticated studies can generate incorrect estimates. While any serious econometrician would grasp the weakness of the Adler and Summers study, it is a substantial challenge to determine whether a study conducted using modern methods is fully reliable. Until the demands of methodological rigor are better understood, the difficulty that decision-makers, the public, and even talented academics have in knowing what studies can be believed will

continue to be exploited by those who seek to push agendas without regard to truth.

References

- Abadie, Alberto, and Javier Gardeazabal. 2003. "The Economic Costs of Conflict: A Case Study of the Basque Country," 93 *American Economic Review* 113–32.
- Adler, Roy D., and Michael Summers. 2007. "Capital Punishment Works," *Wall Street Journal* November 2, A13.
- Alarcon, Arthur L., and Paula M. Mitchell. 2010. "Executing the Will of the Voters: A Roadmap to Mend or End the California Legislature's Multi-Billion-Dollar Death Penalty Debacle," 44 *Loyola of Los Angeles Law Review* S41–109.
- Aneja, Abhay, John Donohue, and Alexandria Zhang. 2011. "The Impact of Right-to-Carry Laws and the NRC Report: Lessons for the Empirical Evaluation of Law and Policy," 13 *American Law and Economics Review* 1–68.
- . 2014. The impact of right-to-carry laws and the NRC Report: the latest lessons for the empirical evaluation of law and policy. NBER Working Paper No. 18294.
- Angrist, Joshua, and Jorn-Steffen Pischke. 2009. *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton, NJ: Princeton University Press.
- Bebchuk, L., and A. Cohen. 2005. "The Costs of Entrenched Boards," 78 *Journal of Financial Economics* 409–33.
- Becker, Gary. 2005a. "More on the Economics of Capital Punishment," Available at: <http://www.becker-posner-blog.com/2005/12/more-on-the-economics-of-capital-punishment-becker.html> (accessed November 17, 2014).
- . 2005b. "Further Comments on Capital Punishment-BECKER," Available at: <http://www.becker-posner-blog.com/2005/12/further-comments-on-capital-punishment-becker.html> (accessed November 17, 2014).
- Berk, Richard. 2005. "New Claims about Executions and General Deterrence: Déjà Vu All Over Again?" 2 *Journal of Empirical Legal Studies* 303–30.
- Bertrand, Marianne, Esther Duflo, and Sendhil Mullainathan. 2004. "How Much Should We Trust Differences-in-Differences Estimates?" 119 *Quarterly Journal of Economics* 249–75.
- Brief for United States as Amicus Curiae, *Gregg v. Georgia*, 428 U.S. 153 (1976) (No. 74-6257, 75-5394, 75-5491, 75-5706 and 75-5844).
- Brook, Robert H., Emmett B. Keeler, Kathleen N. Lohr, Joseph P. Newhouse, John E. Ware, William H. Rogers, Allyson Ross Davies, Cathy D. Sherbourne, George A. Goldberg, Patricia Camp, Caren Kamberg, Arleen Leibowitz, Joan Keesey, and David Reboussin. 2006. "The Health Insurance Experiment," *RAND Health*.

- Retrieved from http://www.rand.org/pubs/research_briefs/RB9174/index1.html (accessed December 19, 2014).
- Bubb, Ryan, and Alex Kaufman. 2014. “Securitization and Moral Hazard: Evidence from Credit Score Cutoff Rules,” 63 *Journal of Monetary Economics* 1–18.
- Bureau of Labor Statistics. 2013. “Occupational Employment Statistics,” Available at: <http://www.bls.gov/oes/current/oes333051.htm> (accessed December 19, 2014).
- Burchfield, Joe. 1975. *Lord Kelvin and the Age of the Earth*. New York: Sci. Hist. Pub.
- Cameron, Samuel. 1988. “The Economics of Crime Deterrence: A Survey of Theory and Evidence,” 41 *Kyklos* 301–23.
- Chan, Sewell. 2010. “Academic Economists to Consider Ethics Code,” *The New York Times*. December 31, B1.
- Cohen-Cole, Ethan, Steven Durlauf, Jeffrey Fagan, and Daniel Nagin. 2009. “Model Uncertainty and the Deterrent Effects of Capital Punishment,” 11 *American Law and Economics Review* 335–69.
- Conlin, Michael, Stacy Dickert-Conlin, and John Pepper. 2005. “The Effect of Alcohol Prohibition on Illicit-Drug-Related Crimes,” 48 *Journal of Law and Economics* 215–34.
- Cremers, Martijn, and Simone M. Sepe. 2014. “Wither Delaware? Limited Commitment and the Financial Value of Corporate Law,” IAST – TSE Conference in Law and Economics 2014. Available at SSRN: <http://ssrn.com/abstract=2519238> (accessed December 19, 2014).
- Cremers, Martijn, L. P. Litov, and S. M. Sepe. 2013. Staggered boards and firm value, revisited. Working Paper. Available at SSRN: <http://ssrn.com/abstract=2364165> (accessed December 19, 2014).
- Crost, Benjamin, and Santiago Guerrero. 2012. “The Effect of Alcohol Availability on Marijuana Use: Evidence from the Minimum Legal Drinking Age,” 31 *Journal of Health Economics* 112–21.
- Daines, Robert. 2001. “Does Delaware Law Improve Firm Value?” 62 *Journal of Financial Economics* 525–58.
- Dezhbakhsh, Hashem, Paul H. Rubin, and Joanna M. Shepherd. 2003. “Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data,” 5 *American Law and Economic Review* 344–76.
- Donohue, John J. 2002. “The Search for Truth: In Appreciation of James J. Heckman,” 27 *Law & Social Inquiry* 23–34.
- . 2007. “Understanding the 1990s Crime Drops in the U.S. and Canada,” 49 *Canadian Journal of Criminology and Criminal Justice* 552.
- . 2009. “Assessing the Relative Benefits of Incarceration: The Overall Change Over the Previous Decades and the Benefits on the Margin” in

- Steven Raphael and Michael Stoll, eds., *Do Prisons Make Us Safer?* New York: Russell Sage Foundation.
- . 2014. “An Empirical Evaluation of the Connecticut Death Penalty System Since 1973: Are There Unlawful Racial, Gender, and Geographic Disparities?” 11 *Journal of Empirical Legal Studies* 637–96.
- Donohue, John J., and James Heckman. 1991. “Continuous versus Episodic Change: The Impact of Civil Rights Policy on the Economic Status of Blacks,” 29 *Journal of Economic Literature* 1603–43.
- Donohue, John J., and Steven Levitt. 2001. “The Impact of Legalized Abortion on Crime,” 116 *Quarterly Journal of Economics* 379–420.
- Donohue, John J., and Peter Siegelman. 1998. “Allocating Resources among Prisons and Social Programs in the Battle Against Crime,” 27 *The Journal of Legal Studies* 1–43.
- Donohue, John J., and Justin Wolfers. 2005. “Uses and Abuses of Empirical Evidence in the Death Penalty Debate,” 58 *Stanford Law Review* 791–846.
- . 2006. “Letter: a Reply to Rubin on the Death Penalty,” 3 *Economists’ Voice* 1–3.
- . 2009. “Estimating the Impact of the Death Penalty on Murder,” 11 *American Law and Economics Review* 249–309.
- Dunning, Thad. 2012. *Natural Experiments in the Social Sciences: A Design-Based Approach*. Cambridge, UK: Cambridge University Press.
- Ehrlich, Isaac. 1975. “The Deterrent Effect of Capital Punishment: A Question of Life and Death,” 65 *American Economic Review* 397–417.
- . 1977. “Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence,” 85 *Journal of Political Economy* 741–88.
- Ekelund, Robert, John Jackson, Rand Ressler, and Robert Tollison. 2006. “Marginal Deterrence and Multiple Murders,” 72 *Southern Economic Journal* 521–41.
- Fagan, Jeffrey. 2006. “Death and Deterrence Redux: Science, Law and Causal Reasoning on Capital Punishment,” 4 *Ohio State Journal of Criminal Law* 255–320.
- Fagan, Jeffrey, Franklin E. Zimring, and Amanda Geller. 2005. “Capital Punishment and Capital Murder: Market Share and the Deterrent Effects of the Death Penalty,” 84 *Texas Law Review* 1803–67.
- Foote, Christopher, and Christopher Goetz. 2008. “The Impact of Legalized Abortion on Crime: Comment,” 123 *Quarterly Journal of Economics* 407–23.
- Francois, Abel, Raul Magni-Berton, and Laurent Weill. 2014. “Abortion and Crime: Cross-Country Evidence from Europe,” 40 *International Review of Law and Economics* 24–35.
- Gius, Mark. 2014. “An Examination of the Effects of Concealed Weapons Laws and Assault Weapons Bans on State Level Murder Rates,” 21 *Applied Economics Letters* 265–67.
- Harris, Kamala D. 2013. *Homicide in California, 2013*. Sacramento, CA: Criminal Justice Statistics Center.

- Hearing before the Subcommittee on Crime, Terrorism, and Homeland Security to consider H.R. 2934, the Terrorist Penalties Enhancement Act of 2003, to strengthen penalties for terrorist offenses. 108th Congress, Second Session, April 21, 2004, p. 24.
- Heckman, James J. 2000. "Microdata, Heterogeneity and the Evaluation of Public Policy," Lecture Presented at the Bank of Sweden Nobel Memorial Lecture in Economic Sciences, Stockholm, Sweden, December 8.
- Heckman, James, and Brook S. Payner. 1989. "Determining the Impact of Federal Antidiscrimination Policy on the Economic Status of Blacks: A Study of South Carolina," 79 *American Economic Review* 138–77.
- Hjalmarsson, Randi. 2009. "Does Capital Punishment Have a 'Local' Deterrent Effect on Homicides?" 11 *American Law and Economics Review* 310–34.
- Iyengar, Radha. 2008. I'd rather be hanged for a sheep than a lamb: the unintended consequences of three-strikes' laws. NBER Working Paper No. w13784.
- Katz, Lawrence, Steven D. Levitt, and Ellen Shustorovich. 2003. "Prison Conditions, Capital Punishment and Deterrence," 5 *American Law and Economics Review* 318–43.
- Keys, Benjamin, Tanmoy Mukherjee, Amit Seru, and Vikrant Vig. 2010. "Did Securitization Lead to Lax Screening? Evidence from Subprime Loans," 125 *The Quarterly Journal of Economics* 307–62.
- Kim, E. Han, Adair Morse, and Luigi Zingales. 2006. "What Has Mattered to Economics Since 1970," 20 *Journal of Economic Perspectives* 189–202.
- Klick, Jonathan. 2013. "Shleifer's Failure," 9 *Texas Law Review* 899–909.
- Klick, Jonathan, and Alexander Tabarrok (2005). "Using Terror Alert Levels to Estimate the Effect of Police on Crime," 48 *Journal of Law and Economics* 267–79.
- Kling, Jeffrey R., Jeffrey B. Liebman, and Lawrence F. Katz. 2007. "Experimental Analysis of Neighborhood Effects," 75 *Econometrica* 83–119.
- Kovandzic, Tomislav, L. Vieraitis, and D. Boots. 2009. "Does the Death Penalty Save Lives? New Evidence from State Panel Data, 1977 to 2006," 8 *Criminology and Public Policy* 803–43.
- La Porta, Rafael, Florencio Lopez de-Silanes, and Andrei Shleifer. 2008. "The Economic Consequences of Legal Origins," 46 *Journal of Economic Literature* 285–332.
- Larson, Richard C. 1976. "What Happened to Patrol Operations in Kansas City? A Review of the Kansas City Preventive Patrol Experiment," 3 *Journal of Criminal Justice* 267–97.
- Lee, Vic. "Study: Can Executions Decrease Murder Rates?" ABC7 San Francisco. N.p., February 13, 2008. Web. 20 Nov. 2014.
- Levitt, Steven D. 1997. "Using Electoral Cycles in Police Hiring to Estimate the Effect of Police on Crime," 87 *American Economic Review* 270–90.

- Liesman, Steve. 2000. "Two U.S. Economists Win Nobel Prize—Heckman, McFadden Cited For Analyzing Decisions Made About Lifestyles," *Wall Street Journal*. October 12, A2.
- Lott, John R., Jr., and David B. Mustard. 1997. "Crime, Deterrence, and Right-to-Carry Concealed Handguns," 26 *The Journal of Legal Studies* 1–68.
- Ludwig, Jens, and Jeffrey R. Kling. 2007. "Is Crime Contagious?" 50 *Journal of Law and Economics* 491–518.
- Marvell, Thomas B., and Carlisle E. Moody. 2001. "The Lethal Effects of Three-Strikes Laws," 30 *The Journal of Legal Studies* 89–106.
- Mocan, Naci, and R. Kai Gittings. 2003. "Getting Off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment," 46 *Journal of Law and Economics* 453–78.
- Muhlhausen, David B. 2007. "The Death Penalty Deters Crime and Saves Lives," Testifying on June 27, 2007 before the Subcommittee on the Constitution, Civil Rights, and Property Rights of the Committee on the Judiciary of the United States Senate. Available at: <http://www.heritage.org/research/testimony/the-death-penalty-deters-crime-and-saves-lives> (accessed November 19, 2014).
- Munnell, Alicia H. 1986. "Lessons from the Income Maintenance Experiments: An Overview," in *Proceedings of a Conference Sponsored by the Federal Reserve Bank of Boston and the Brookings Institution*, Melvin Village, NH.
- National Research Council. 1978. *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washington, DC: National Academies Press.
- . 2012. *Deterrence and the Death Penalty*. Washington, DC: National Academies Press.
- Passell, Peter, and John B. Taylor. 1977. "The Deterrent Effect of Capital Punishment: Another View," 67 *The American Economic Review* 445–51.
- Posner, Richard A. 2006. "The Economics of Capital Punishment," 3 *Economists' Voice* 1–3.
- Rubin, Paul. 2006. "Statistical Evidence on Capital Punishment and the Deterrence of Homicide," Testifying on February 1, 2006 in Hearings on "An Examination of the Death Penalty in the United States," Senate Judiciary Committee. Available at: <http://mcadams.posc.mu.edu/blog/TestimonyPaulRubin.pdf> (accessed November 17, 2014).
- Rubin, Paul H., and Hashem Dezhbakhsh. 2003. "The Effect of Concealed Handgun Laws on Crime: Beyond the Dummy Variables," 23 *International Review of Law and Economics* 199–216.
- Sanbonmatsu, Lisa, Jeanne Brooks-Gunn, Greg J. Duncan, and Jeffrey R. Kling. 2006. "Neighborhoods and Academic Achievement: Results from the MTO Experiment," 41 *Journal of Human Resources* 649–91.
- Scheiber, Noam. 2007. "Freaks and Geeks; How Freakonomics Is Ruining the Dis-mal Science," *The New Republic*, April 2, 27–31.

- Shepherd, Joanna M. 2002a. “Police, Prosecutors, Criminals, and Determinate Sentencing: The Truth about Truth-in-Sentencing Laws,” 45 *Journal of Law and Economics* 509–33.
- . 2002b. “Fear of the First Strike: The Full Deterrent Effect of California’s Two- and Three-Strikes Legislation,” 31 *Journal of Legal Studies* 159–201.
- . 2004. “Are Criminals Like Us? Risk Attitudes, Sentencing Guidelines, and Increased Crime,” Emory Law & Econ Research Paper No. 04-03.
- Sidak, Gregory. 2013. “Court-Appointed Neutral Economic Experts,” 9 *Journal of Competition Law & Economics* 359.
- Sunstein, Cass, and Adrian Vermeule. 2005. “Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs,” 58 *Stanford Law Review* 703–50.
- Sunstein, Cass, and Justin Wolfers. 2008. “A Death Penalty Puzzle,” *The Washington Post*, <http://www.washingtonpost.com/wp-dyn/content/article/2008/06/29/AR2008062901476.html> (accessed November 19, 2014).
- Supreme Court of the United States. *Baze et al. v Rees, Commissioner, Kentucky Department of Corrections, et al.* No. 07-5439. Argued January 7, 2008; Decided April 16, 2008. Available at: <http://www.supremecourtus.gov/opinions/07pdf/07-5439.pdf>.
- White, Halbert. 1980. “A Heteroskedasticity-Consistent Covariance-Matrix Estimator and a Direct Test for Heteroskedasticity,” 48 *Econometrica* 817–38.
- Wooldridge, Jeffrey M. 2010. *Econometric Analysis of Cross-Section and Panel Data*. Cambridge, MA: MIT Press.
- Yörük, Barış, and Ceren Yörük. 2011. “The Impact of Minimum Legal Drinking Age Laws on Alcohol Consumption and Marijuana Use: Evidence from a Regression Discontinuity Design Using Exact Date of Birth,” 30 *Journal of Health Economics* 740–53.
- Zimring, Franklin E. 2006. *The Great American Crime Decline*. New York City: Oxford University Press.
- . 2008. “Criminology and Its Discontents: The American Society of Criminology 2007 Sutherland Address,” 46 *Criminology* 255–66.
- Zimring, Franklin E., Jeffrey Fagan, and David T. Johnson. 2010. “Executions, Deterrence, and Homicide: A Tale of Two Cities,” 7 *Journal of Empirical Legal Studies* 1–29.