From the "Econometrics of Capital Punishment" To the "Capital Punishment" of Econometrics: On the Use and Abuse of Sensitivity Analysis

Hashem Dezhbakhsh.

Paul H. Rubin,

September 2007

Draft 1

* Hashem Dezhbakhsh is a professor of economics at Emory University and Paul Rubin is Samuel Candler Dobbs Professor of Economics and Law at Emory University.

The authors are thankful to Joanna Shepherd for substantial assistance with the paper. While she agrees with the empirical findings in the paper, Joanna has chosen not to continue with the current debate by being a co-author. Moreover, we conclude our contribution to this debate with the current paper to avoid an endless academic squabble.

From the "Econometrics of Capital Punishment" To the "Capital Punishment" of Econometrics: On the Use and Abuse of Sensitivity Analysis

Abstract

The academic debate over the deterrent effect of capital punishment has intensified again with a major policy outcome at stake. About two dozen empirical studies have recently emerged that explore the issue. Donohue and Wolfers (2005) claim to have examined the recent studies and shown the evidence is not robust to specification changes. We argue that the narrow scope of their study does not warrant this claim. Moreover, focusing on our two studies that they have examined, we show the deterrence findings to be robust, while their work has serious flaw in analysis and selectivity in reporting the results. The selectivity is biased toward showing "no deterrence." This highlights the importance of a proper framework for guiding the sensitivity analysis of published work to guard against data-mining and agenda-driven empiricism. We hope that our study generates interests in appropriate ways to do sensitivity analysis of published work as much as it contributes to the debate about capital punishment.

I. Introduction

The deterrent effect of capital punishment has been the subject of passionate debates among scholars. Criminologists pioneered empirical analysis of the issue and found no deterrence (see, e.g., Sellin, 1959; Eysenck, 1970; and Cameron, 1994). Economists entered the debate with Ehrlich's (1975 and 1977) work that introduced econometrics rigor into the analysis and reported a strong deterrent effect. Ehrlich's findings received considerable attention from the policy makers.[1] The attention prompted economists as well as criminologist to reexamine or extend Ehrlich's analysis with mixed results.[2]

During the past ten years, the interest in the issue has been reinvigorated as various states in the U.S. have considered changing their death penalty laws. A number of academic studies have also been added to the death penalty literature, mostly examining the recent evidence. Fourteen of these studies report some deterrence, six show no deterrence, and one is ambiguous.[3] Moreover, Sunstein and Vermeule (2005) draw on the collective deterrence evidence to argue that government has an obligation to act on the evidence and vigorously prosecute the death penalty.

---

[1] The Solicitor General of the United States, for example, introduced Ehrlich's findings to the Supreme Court in support of capital punishment. (Fowler vs. North Carolina, 428 U.S. 904, 1976).

[2] For example, Yunker (1976), Cloninger (1977), Ehrlich and Gibbons (1977), Layson (1985), and Chressanthis (1989) report further deterrence evidence; Bowers and Pierce (1975), Passel and Taylor (1977), Hoenack and Weiler (1980), McManus (1985), Leamer (1983), Cover and Thistle (1988), McAleer and Veall (1989), and Grogger (1990) find no deterrence; and Black and Orsagh (1978) find mixed results.

[3] The studies that show deterrence include Brumm and Cloninger (1996), Ehrlich and Liu (1999), Lott and Landes (2000), Cloninger and Marchesini (2001), Yunker (2002), Dezhbakhsh, Rubin, and Shepherd (2003), Mocan and Gittings (2003), Shepherd (2004), Liu (2004), Zimmerman (2004 and 2006a), Cloninger and Marchesini (2005), Dezhbakhsh and Shepherd (2006), Ekelund, Jackson, Ressler, and Tollison (2006). The no deterrence studies include Bailey (1998), Albert (1999), Sorenson, Wrinkle, Brewer, and Marquart (1999), Stolzenberg and D'Alessio (2004), Berk (2005), and Fagan, Zimring, and Geller (2006). Results reported by Katz, Levitt, and Shustorovich (2003) are mixed. We do not include Fagan (2006) in the above count as it is a compendium of verbal criticisms without any empirical analysis to back the author's assertions.

Reacting to Sunstein and Vermeule's policy call, Donohue and Wolfers (2005) conduct what they call a "*thorough assessment of the statistical evidence on this important public policy issue*" and claim to have found "*little evidence for deterrence.*"[4] To determine whether their claim is on solid grounds, we carefully examine their sensitivity analysis of our two studies that constitute the bulk of their sample.[5]

Our findings raise serious doubts about Donohue and Wolfers' analysis, conclusions, and claims.[6] They have made a number of conceptual, statistical, and implementation errors that render their analysis invalid. Moreover, our replication of their work shows that, contrary to the norms of sensitivity analysis, their reporting of the key model estimates is highly selective, favoring results that show "no deterrence." A vast number of their sensitivity estimates that show deterrence are never reported, or even acknowledged. This suggests that the tools to safeguard against data-mining, are used by them to mine the data.

In addition, the scope of Donohue and Wolfers inquiry is far too limited to warrant writing the obituary of the deterrence findings. The sample of studies they choose to examine is small even if the relevant population is just the recent 21 papers listed above. In fact, six of the nine Donohue and Wolfers' tables focus only on two death penalty studies.

The only undisputable finding of Donohue and Wolfers is that the existing deterrence findings can be easily replicated, meeting an important requirement of scientific inquiries. This finding, which is unfortunately lost in the welter of exaggerated claims, is a testimony

---

[4] Donohue and Wolfers (2005), pages 794 and 844.

[5] These two studies are Dezhbakhsh, Rubin, and Shepherd (2003) and Dezhbakhsh and Shepherd (2006). We also need to note that our purpose here is not to offer a point by point response to Donohue and Wolfers' critique of our studies. Our aim is rather to analyze their methods and conclusions.

[6] Zimmerman (2006b) documents several misrepresentations and errors in Donohue and Wolfers' study. Clonninger and Marchesini (2007) state that Donohue and Wolfers' measurement-related criticism of their work is baseless. Mocan and Gittings (2006) also raise doubts about Donohue and Wolfers' claims.

to the credibility of the existing deterrence evidence, since such ease of replication might not be common in social science empirics.

The academic exchange between Donohue and Wolfers and the authors of a few death penalty studies might not be of great interest to every economist, but the issues that this exchange raises about sensitivity analysis are important to all empirical social scientists. Hendry (1980), Lovell (1983), and Leamer (1983, 1985) all criticized data mining and whimsical choices in econometrics. Their classic critiques are as relevant now as they were then. These authors intended to provide guidelines for practitioners to produce more credible evidence by making valid choices. Equally important are the relevance and validity of choices made by sensitivity investigators who reexamine the work of other authors. They too need to follow protocols and systematic procedures to reexamine results and then impartially reporting the findings. After all, counter evidence based on invalid assumptions cannot be credible.

Our study shows that more work is needed in this area to guard against biased and agenda-driven approaches to sensitivity analysis. Specific guidelines and statistical protocol are needed to ensure objective and scientific sensitivity analysis of published studies. We hope that our study generates interests in proper ways to do sensitivity analysis as much as it contributes to the debate about capital punishment, that seems to be heating up again with significant policy outcome at stake (see Tanner, 2007).

The paper is organized as follows: Section II briefly discusses data mining and sensitivity analysis in economics, and how such analysis can be abused. Section III describes the shortfalls of the Donohue and Wolfer's study, focusing on biased reporting and misleading presentation of results. This section also provides further evidence on robustness of the results reported in Dezhbakhsh and Shepherd (2006). Section IV reveals statistical errors made by Donohue and Wolfers and the implications of such errors for their claim about the deterrent effect of capital punishment. Section V offers a summary and some concluding remarks.

II. Sensitivity Analysis in Economics

In a fascinating historical recount, Stephen Stigler (1986) discusses how various sciences adjusted their approaches to take advantage of the emerging tools of modern probability and statistics. Astronomers, for example, abandoned averaging methods to adopt linear models. Psychologists redesigned their experimental method. Social scientists, however, at first did not embrace linear models or experimental methods, as they were trying, with much frustration, to identify causal links by brute force and massive data bases.[7]

Despite some successes, however, the complexity of social phenomena continues to challenge the limits of empirical methods. In economics, the explosion of data, wide dissemination of econometric tools, canned programs, and computing capacity all offer researchers many choices, generating biased as well as unbiased research. Consequently, zealous researchers who seek new discoveries at any cost continue to create controversies and frictions in empirical economics.

Data Mining

[7] Economists had their own share of these frustrations, reflected, for example, in the melodramatic exchange between economist William Stanley Jevons who had combined various prices into one index and his critics who were opposed to the notion of an index created from prices of vastly different commodities (See, Stigler, 1986 and Jevons, 1869).

A hazardous practice that economists have been preached against is data mining, also known as fishing or data grubbing.[8] The practice involves estimating many, sometimes hundreds of, models in hope of obtaining a good fit. The exhaustive search can be guided by statistical metrics, such as coefficient of determination and Akaike Information Criterion, or by researcher's predispositions and even ideology.

Lovell (1983) shows that when searching for best regressors, data mining can lead to unusually large significance levels (large type-I error), therefore, invalidating the traditional methods of inference. The scientific integrity of the work is also compromised when the search is agenda driven and its extent is veiled from the reader.[9] Other prominent econometricians who voiced concerns about econometric practice include Leamer (1983) who called for taking "the con out of econometrics," Hendry (1980) who characterized econometric practice as Alchemy, and Sims (1980) who argued that empirical macroeconomics is divorced from reality. Leamer's critiques (1983 and 1985) are more relevant to the current discussion as they focus on sensitivity analysis as a tool to examine the reliability of statistical inference.

Sensitivity Analysis

Econometric inference is usually based on a presumed model. A change in model assumptions—such as functional form, explanatory variables, or parametric restrictions—can alter the results. Sensitivity analysis refers to estimating models with alternative specifications. Accordingly, if various estimates of the parameter(s) of interest are similar,

[8] See, e.g., Lovell (1983), Chatfield (1995), and Hand (1998).
[9] Irrespective of the researcher's incentives and modality of search, data mining can lead to reporting a random pattern as systematic or structured evidence.

then the finding is robust. This tool is intended to guard against data mining and to enhance the credibility of empirical research (Chatfield, 1995).

To provide a framework for sensitivity analysis, the uncertainty in nonexperimental inference can be categorized into two parts: the sampling uncertainty and misspecification uncertainty.[10] The former is rooted in sample changes and the latter in specification changes. For illustration, consider two spaces, one consisting of all possible model assumptions (assumption space) and the other consisting of all possible inference (inference space). There is a correspondence between the two spaces that applied econometricians try to map. If the mapping includes only a small segment of the assumption space, then the inference will be of limited use. Moreover, if this segment is selected through data mining, particularly with predisposition, then the inference will be suspect, and often invalid.

Sensitivity analysis involves expanding the mapping to a broader segment of the assumption space and reporting the corresponding inference. But a critical issue is what segments of the assumption space to cover in the expanded mapping. Constraints have to be imposed since a full mapping is impossible. The researcher's prior belief may also play a role here. To avoid arbitrariness and whimsy, however, informed conventions need to be used to impose discipline on such priors (Leamer, 1983 and 1985). Accordingly, the analysis is conducted only over relevant segments of the assumption space. The range of the corresponding inference suggests whether the original inference is fragile or robust.

For example, in a linear regression model, a researcher who wants to examine the sensitivity of her results with respect to explanatory variables must first identify the variables that must be included in the regression. These are called focus variables and the

[10] See, e.g., Leamer (1983 and 1985). Also, note that in statistics parlance, model uncertainty is another alias for specification uncertainty; see, e.g., Chatfield (1995).

remaining variables are called doubtful variables. The coefficients of focus variables are called free parameters. For each free parameter a range of estimates can then be obtained by adding various combinations of doubtful explanatory variables to the equation and re-estimating it.[11] A narrow range suggests a robust finding. The above method is known as extreme bounds analysis, and can be viewed as a part of global sensitivity analysis that expands the inquiry in various directions in the parameter and inference spaces.

Fishing for Sensitivity Evidence

Sensitivity analysis is a useful tool for examining robustness, or fragility, of a statistical result. It can be used to dispel doubts or to create doubts about an inference. It is also a powerful tool that can be used to manipulate findings. Leamer as well as others have expressed concern about the potential for such abuses. Leamer (1985) suggests an organized and systematic approach to avoid manipulation or whimsy. In reference to the studies that attempted to examine Ehrlich's (1975) death penalty finding, he asserts "*These disorganized studies of fragility are inefficient, haphazard, and confusing.*" McAleer, Pagan and Volker (1985) criticize the heavy reliance of unstructured and nonsystematic sensitivity analysis on whimsical choices made by researchers. Referring to such analysis, they state "*…Hardly a good method for getting rid of the con artists. Instead it gives them enormous scope in generating any result they like*."

Moreover, Granger and Uhlig (1990) go one step further and warn against including econometrically irrelevant estimates in sensitivity analysis. They argue, for example, that estimates with poor fit, low Durbin-Watson, or other specification ailments should not be presented as evidence of robustness or fragility. Ehrlich and Liu (1999) also argue that

[11] The original estimates of the free parameters are from one of these combinations.

theory-based restrictions on individual estimates should be further used in sensitivity analysis. Therefore, not every specification check is worthy of consideration and not all evidence of fragility can be taken seriously, especially when the accepted norms for avoiding bias are ignored. For a more detailed discussion of this issue see the Appendix.

When researcher does not adhere to a set of guidelines to perform the sensitivity checks in a systematic and methodical fashion, then the exercise can potentially turn into a fishing expedition, a con job, or groping in the dark. The culprits of such abuses can be a data miner who tries to convince the readers that his results are resilient, or a zealous skeptic who tries to cast doubt on a published study.

The data miners' abuses have been the subject of much criticism, leading to a higher scrutiny before publication.[12] The doubt casters' mischief, however, needs more scrutiny, because the practice is just as damaging. Indeed, the relative ease of obtaining irrelevant or invalid evidence of fragility (as discussed above), high likelihood of publishing such findings, and potential controversy that will result all provide incentives for empirical scavengers to fish for counter evidence at any cost. After all, obtaining someone else's data and tweaking their model over and over till a different inference emerges is much easier than collecting your own data, building an econometric model, developing a statistical strategy, and making an inference.

A number of safeguards for preempting the manipulation of sensitivity analysis come to mind.[13]

[12] For example, as the result of these criticisms, many journal referees now require that authors perform specific sensitivity checks rather than making a general statement that results are robust.
[13] This list is not intended to be exhaustive and is drafted drawing on the literature on sensitivity analysis and the authors experiences.

1) The process (of fragility search) should be transparent, well documented, and reproducible.

2) The direction of the specification searches needs to be clearly explained and justified on theory and econometric grounds following statistical conventions.

3) Blatantly invalid estimates should not be included in establishing the range of an inference.

4) Results should not be reported selectively. Given the large number of sensitivity checks, economy of space requires presenting some results and not others. But the presented results should be representative, and the selection should not be agenda driven and misleading. More importantly, all estimates of the free parameters (related to the focus variables) should be presented.

5) If the sensitivity analysis involves several studies, then similar approaches should be applied to all studies and results should be presented without any bias toward a particular finding.

6) It is prudent to make the sensitivity results available to authors whose work has been examined, not as a courtesy, but as a safeguard against misunderstanding, misreporting, or other errors and omissions. This should precede any publication attempt to maintain the integrity of the scientific inquiry.

7) To gain credibility, the sensitivity analysis should go through the standard peer review to detect its errors and flaws before the study is published.

III. Sensitivity Diagnostics and Biased Reporting

Donohue and Wolfers (2005) claim that their sensitivity checks should raise doubts about the recent evidence on the deterrent effect of capital punishment. They examine a few studies which include ours (Dezhbakhsh, Rubin, and Shepherd, 2003; and Dezhbakhsh and Shepherd, 2006). Our intention here is to show how following the "*don't*"s rather than the "*do*"s of sensitivity analysis can distort evidence and mislead readers. We do not offer a point by point response to their claims. Most of the shortcomings that we expose relate to Donohue and Wolfers' analysis of our studies, but some of these criticisms apply equally to their analysis of other studies that find deterrence.[14]

In this section we focus on Dezhbakhsh and Shepherd's (2006) study.[15] This study uses panel data for fifty states during the 1960-2000 period to examine the deterrent effect of capital punishment, using the moratorium as a "judicial experiment." Its results, based on a set of robustness checks involving 96 models, suggest that capital punishment has a deterrent effect, and that executions have a distinct effect which compounds the deterrent effect of merely (re)instating the death penalty.

Donohue and Wolfers reestimate some of the baseline models in this study under a number of arbitrarily assumptions. We first challenge the relevance of these assumptions. But to put Donohue and Wolfers' claims to test, we then adopt these assumptions anyway and redo their analysis . Our results reveal that Donohue and Wolfers present the key

[14] For example, in their replay to Donohue and Wolfers (2005) criticism of their work, Mocan and Gittings (2006) expose errors that Donohue and Wolfers have made in estimation of risk measure. They further show that appropriate extension of the specification in their original paper (Mocan and Gittings, 2003) does not alter their deterrence finding. Zimmerman (2006b) also documents several misrepresentations and errors in Donohue and Wolfers' study.

[15] Donohue and Wolfers examine a prepublication version of this study which is available at http://law.bepress.com/cgi/viewcontent.cgi?article=1017&context=alea. While there are some differences between this version and the published version, the main findings are quite similar.

estimates quite selectively. This selectivity is neither random nor benign; it, rather, tends to disguise the robustness of our results and misrepresent the findings in favor of "no deterrence."

Design of Sensitivity Checks

We had performed a large number of sensitivity checks in Dezhbakhsh and Shepherd (2006). In addition to the main models, another 96 distinct regression equations were estimated that show robustness of the results with respect to changes in regressors, functional form, data, estimation method, and the possibility of capital punishment being a spurious finding stemming from common crime patterns. The prepublication version that Donohue and Wolfers examine includes 84 robustness checks, which they never acknowledge. They perform several additional checks that we discuss next.

Omitting Texas: Donohue and Wolfers argue that Texas has far more executions than other states and, therefore, should not have been included in the analysis. This is despite their own admission that Texas is not an outlier, as it shows close alignment with the estimated regression line.[16] We believe that Texas provides useful evidence for the deterrence hypothesis, because it applies the death penalty. Any deterrence from capital punishment is rooted more in the *exercise* of the law, than in status of the law.[17] Moreover, given that Donohue and Wolfers argue that executions have too little variations to be useful in explaining changes in the homicide rate,[18] dropping a state where executions are prevalent and vary from year to year does not make sense. In fact, if one uses the frequency

[16] See Donohue and Wolfers (2005) page 815.
[17] See Cloninger and Marchesini (2007) for a discussion of this point.
[18] See Donohue and Wolfers (2005) pages 796- 797.

of executions as a sample selection criterion, then states with too few executions must also be dropped.

However, throwing out data is a bad idea that most statistics textbooks warn against. Statisticians have, indeed, devoted a lot of efforts to develop methods that deal with outliers and influential data points to encourage practitioners to use such data rather than discard them.[19] Throwing out an observation randomly only reduces the estimation accuracy, without biasing the resulting estimates. But, if the omission is based on a criterion—such as Donohue and Wolfers' "too many execution criterion"—then the outcome will be inaccuracy as well as estimation bias due to sample selectivity.[20]

***De Facto* Versus *De Jure* Laws:** Another sensitivity check that Donohue and Wolfers perform on our results involves changing the binary moratorium variable to make a distinction between states that apply their death penalty law and states that do not. Accordingly, a state with a death penalty law is classified as a *de jure* death penalty state in any given year if it has not executed anyone during the previous ten years. A state that has executed at least one convict during the previous ten years is considered a *de facto* death penalty state. Although the idea may appear reasonable at first—especially if one ignores the arbitrariness of Donohue and Wolfers' ten year window and their failure to use alternative windows—it has flaws that surface upon careful examination.

For example, the *de facto* vs. *de jure* distinction is irrelevant considering that we did control for the application of execution in our various models by including as regressor(s) the number of executions and/or number of executions lagged by one period. More importantly, the way Donohue and Wolfers implement their *de fact* vs. *de jure* distinction is

[19] See, e.g., Cook (1977, and 1982) and Welsch (1982).
[20] See, e.g., Wooldridge (2003), chapter 9 for the perils of non-random data omission.

awkward and inappropriate. Our original moratorium binary variable equals one for every year during a state's moratorium and zero otherwise. Adjusting this variable to accommodate Donohue and Wolfers' *de facto* vs. *de jure* distinction requires changing some of the zeros to one. Correct construction of the new binary variable requires not confusing the *ex ante* with the *ex post*. So, when the moratorium for a state is lifted, the value switches to zero and stays zero for ten years. If the state has not executed anyone by then, the binary variable will take value of one from that point on until the next execution. For years prior to moratorium, past executions will determine whether the binary value is zero or one.

What we find peculiar is that instead of the above straight forward extension, Donohue and Wolfers implement their *de facto* law using two binary variables—one for active death penalty states and the other for inactive death penalty states. Given that each state is classified in one of the two categories in any given year, using two binary variables is quite unusual and inconsistent with standard practice as well as their own work; see, e.g., their Table 2. Using two binary variables requires dropping the intercept and absorbing its effect into these binary variable in order to avoid a dummy trap.[21] The results we report later suggest that Donohue and Wolfers *de facto* moratorium finding is perhaps driven by the awkward implementation of *de facto* vs. *de jure* distinction rather than the distinction itself.

Cluster Correction: The easy fix that software packages offer for correcting the clustering effect, or cluster samples, has persuaded practitioners to use it, and often as a knee-jerk reaction. This is done without much attention to the statistical procedure that is applied when this option is invoked and the statistical properties of the resulting inference in

[21] Dummy-trap refers to a situation where a linear combination of two or more regressors are equal to a column of ones and ,thus, perfectly multi co-linear with the regression intercept, making estimation impossible.

terms of type I and type II errors. The correction adjusts the variance covariance matrix of the least squares coefficient estimates for intra group dependencies. The method is only asymptotically valid[22]—in other words, when the number of groups (clusters) relative to the size of each group approaches infinity—a condition that clearly does not hold in the current case. Moreover, the correction is subject to the same criticism as heteroskedastic-consistent covariance estimation. For example, Leamer (1994) refers to the correction as white-washing the problem and expresses doubt about its effectiveness in correcting the size of an estimation interval.

Besides the above criticism, there is the issue of implementation. We believe that Donohue and Wolfers use the cluster option incorrectly. Cluster correction is a method used to deal with spatial correlation in cluster samples. Clusters are cross section groups that are nested within other groupings—for example, students in a school when the sample includes different schools or employees in a firm when the sample includes different firms; see, e.g., Wooldridge (2002, chapter 6) for a discussion. In the current case, one can envision states to form a cluster, because there are many unobservable national factors that can affect crime in various states at any given time—e.g., political mood, federal support for crime fighting programs, media effect, and global factors affecting domestic crime. These contemporaneous influences cause the regression errors to be contemporaneously correlated across states. The magnitude of the correlation can increase with inter-state linkages and proximities.

[22] See, e.g., Wooldridge, 2002, chapter 1.

While Donohue and Wolfers might have intended to correct for this problem, they seem to have used a wrong code forcing, years to form clusters rather than states.[23] For example, the option *cluster(state)* indicates that each state is a cluster which is made up of over time observations for that state. The option that we used in our study is *cluster(year)*.[24] This option forms clusters that include various states at a given time. If one accepts Donohue and Wolfers argument, elsewhere in their paper,[25] that various counties within a state should form a cluster, then various states at a given time should also form a cluster, as we use in our study and not the opposite that they use in their sensitivity checks. In fact, we detected Donohue and Wolfers' error when we used both options to examine the sensitivity of the results to their choice of models (see the results in the following subsection).

Alternative Execution Measures: Donohue and Wolfers criticize using the number of executions as one of the deterrence variables and use, instead, three other measures in their sensitivity checks. One of their choices is executions divided by (lagged) homicide rate. We believe this choice is neither relevant for our study nor valid. In fact, we measured deterrence with conditional probabilities of arrest/homicides, conviction/arrests, and executions/convictions in Dezhbakhsh, Rubin, and Shepherd (2003) where the goal was to use a choice-theoretic model with various execution probabilities as deterrence factors. In Dezhbakhsh and Shepherd's (2006), however, we use the number of executions as a communication factor that signals to potential criminals the state's will and determination to fight capital crimes. Donohue and Wolfers' measure is invalid because if they intend to use

[23] This error is due to ignoring specific coding of the STATA command; see, *STATA User's Guide*, Release 8, page 273.
[24] We find Donohue and Wolfers' assertion that we used OLS method disingenuous. We had communicated to them that we used weighted least squares with robust standard errors and cluster correction (e-mail available from the authors upon request). In fact, when attempting to replicate our work, they themselves use weighted least square which is GLS and not OLS; see Donohue and Wolfers (2005) Table 2 and top of page 805.
[25] See Donohue and Wolfers (2005), page 834.

a probability based choice-theoretic model, they need to include all three conditional probabilities and correctly calculate ratios that capture probabilities. By the same argument, their second deterrence measure—the executions/prisoners ratio—is also flawed.

Donohue and Wolfers' third deterrence measure is executions per 100,000 residents. They argue that our non-per capita measure causes the effect of one execution to be different in states with different population. True! But that is exactly what we want.[26] We believe that an execution in a densely populated state with more crimes, more criminals, and more potential criminals has a stronger deterrent effect, in terms of the number of lives saved, than an execution in a sparsely populated state with few crimes and few potential criminals. So dividing the number of executions by population makes no sense. We reiterate that we use executions as a signaling factor in Dezhbakhsh and Shepherd (2006).

Time Trends: To control for state and time specific factors we use a set of state and time specific binary variables. Four time specific binary variables were used each corresponding to one of the decades in our sample (60s, 70s, 80s, and 90s). This choice was based on the observation of crime trend similarities during various decades. These similarities are also pointed out by Donohue and Wolfers.[27] As an alternative they use year specific binary variables in their sensitivity checks. We have no problem with their choice and use it in our assessment of their sensitivity analysis.

Donohue and Wolfers' Biased Reporting of Sensitivity Results

Despite our serious reservations about the specification choices that Donohue and Wolfers have made for their sensitivity checks on our results, we conducted a sensitivity

[26] We are surprised that Donohue and Wolfers question this intuitive assumption. As an analogy, one can point to the diverse effect that one minute of TV advertising has in markets with different population densities, and, therefore, different viewer densities. Advertisers pay more for dense markets because of the bigger impact.

[27] Donohue and Wolfers (2005), page 796.

reanalysis using their suggested specifications. Our analysis can serve two purposes. First, it adds to the extensive robustness checks reported in Dezhbakhsh and Shepherd (2006). Second, it can ascertain how accurately Donahue and Wolfers carry out their sensitivity checks and how objectively they report their findings. The outcome will not only put to test the credibility of Donahue and Wolfers' claims but also inform future sensitivity analyses.

For this analysis, we use our base models 1, 2, and 3 that each include one of the three deterrence variables—number of executions, number of executions (lagged), and binary moratorium variable—and model 7 that includes all three variables.[28] These models are reestimated under a variety of Donohue and Wolfers' specifications as listed below. We use two of the deterrence measure suggested by Donohue and Wolfers. These include executions/population in 100,000 and the de facto moratorium as defined earlier. We use the words *de jure* to refer to the results based on the statutory moratorium and *de facto* to refer to the results based on Donohue and Wolfers suggested moratorium variable.

We also correct for the argued clustering effect once by identifying states as members of a cluster, using *cluster(year)* option in STATA, and once by identifying years as members of a cluster, using *cluster(state)* option in STATA to account for possible *within* cluster dependencies. We also estimate models with and without Texas. Finally, we include year fixed effects (dummies) or decade fixed effects (dummies) in addition to the state fixed effects that are included in all of the estimated models. Overall, these specifications result in 80 different regressions, each with one or more deterrence estimates.

Results: The deterrence coefficient estimates along with the corresponding t-statistics for all eighty cases are reported in Table 1. The models are numbered 1 through 80, each reported in one column. Since the dependent variable is murder, a negative coefficient

[28] See Dezhbakhsh and Shepherd (2006), Table 8.

estimate for execution variables suggests deterrence. There are 144 coefficient estimates in Table 1, out of which 106, or more than 73.5%, suggest deterrence at statistically significant levels.[29] Note that none of these models produces a significant estimate with an opposite sign (more executions, more crimes).

Models 1, 2, 3 and 4 in Table 1 are our original baseline models. When the number of executions is replaced with Donohue and Wolfers' suggested number of executions/population, we get models 5, 6 and 7. Note that all estimates in these three models are highly significant, suggesting that the deterrence finding does not change when we use Donohue and Wolfers' execution measure. Models 8-11 show that dropping Texas by itself does not alter the highly significant deterrence results. The same is true for dropping Texas and using the execution/population, as can be seen from models 12-14.

Moreover, models 15-18 suggest that capturing the time fixed effect with year dummies instead of decade dummies does not alter the deterrence results for the execution variables. Models 19-25 suggest that using alternative execution measures (or dropping Texas) along with states as cluster members and year dummies does not alter our findings either.

Using *de facto* instead of *de jure* moratorium does not have any effect on our baseline results, as can be seen from models 29 and 30. Also, models 31-40 show that using various combinations of *de facto* moratorium, excluding Texas, using executions/population as deterrence measure, or changing from decade dummies to year dummies have little effect on the deterrence finding. Note in none of these models the *de facto* moratorium is insignificant.

---

[29] For these parameter tests, the null is the absence of deterrence and the alternative is the presence of deterrence. Given our interest, the one sided deterrence alternative is appropriate.

Models 41-54 suggest that changing the cluster members from states to years by itself or in combination with dropping Texas and/or using execution/population as deterrence measure has very little effect on our results; all 22 deterrence coefficient estimates in these models are significant. Models 55-58 suggest that using year dummies instead of decade dummies and changing the cluster grouping to years will not affect the estimated coefficients of executions and lagged executions, but will make the *de jure* moratorium insignificant. However, using *de facto* moratorium with years as cluster members, executions or executions/population, and with or without Texas has no significant effect on the deterrence estimates—there is a reduction in significance level, but estimates are still significant; see models 69-74. Overall, the results based on the *de facto* moratorium variable seem to be more robust than what Donohue and Wolfers report. It is not clear to what extend their result is an artifact of the unusual way in which they define a binary outcome with two binary variables. Finally, the few models with all insignificant coefficients pertain to cases without Texas and with years as cluster members, year dummies, and mostly with executions/population as measure of deterrence; see models 62-68 and 77-80.

Biased Reporting: The above results suggest that our main findings are generally robust even to the few peculiar specification changes that Donohue and Wolfers suggest. The above results clearly do not tell the same story that Donohue and Wolfers want their readers to believe. One may then wonder "Why Donohue and Wolfers' study suggests that Dezhbakhsh and Shepherd (2006)'s results are fragile?" The answer lies in "selective reporting".

A systematic sensitivity check requires reporting the range and, if possible, the distribution of the resulting estimates and not just a few points chosen either haphazardly or as the result of a data mining exercise. We compared the estimates in Table 1 here with those reported in Donohue and Wolfers' (2005) Tables 2 and 5 or described in the corresponding text to see whether Donohue and Wolfers' work meets this important objectivity criterion. We found that they only report a small subset of their sensitivity results, mainly those with insignificant estimates, and make no mention of many of their estimates that support deterrence.

More importantly, in models where one of the three deterrence coefficients is insignificant, but the other two are significant, Donohue and Wolfers report only the insignificant coefficient to exaggerate the impression of fragility, concealing the other estimates in the same model. For example, consider model 3 in their table 2 (titled "Controlling for Year Fixed Effects"), which corresponds to our model 58 in Table 1.[30] Note that here the estimated coefficients for two of the deterrence variables are highly significant. Yet they neither report these two estimates, and nor do they even mention their significance. Given that all deterrence coefficients are considered key parameters here, concealing their estimates under alternative specifications violates a sacred rule of sensitivity analysis.

In fact, it is a common practice in the sensitivity analysis to report the results related to all key parameters, even when some follow researcher's priors but others do not. For example, Bartley and Cohen (1998) use sensitivity analysis to examine the robustness of Lott and Mustard's (1997) findings about the effect of concealed handgun laws on crime

[30] Their estimate for moratorium variable is -.47 which is close to our estimate of .4 for the moratorium variable. The difference in sign is trivial and due to their designation of 1 and 0 instead of 0 and 1 for the binary variable.

and find the primary effect to be robust but the substitution effect not to be robust across various specifications. They do report both results, however. Quinn (1997), Sjoholm (1996), and Fowles and Loeb (1995) also report results for all key parameters of the model they examine without any exclusion. Unfortunately, Donohue and Wolfers analysis falls short of this inclusivity requirement. This is not an isolated reporting oversight, but rather part of a systematic pattern of reporting as will be exposed further in the following section. In the next section we turn to statistical errors in Donohue and Wolfers' (2005) sensitivity analysis.

IV. Can Invalid Inference Serve as Evidence of Fragility?

Donohue and Wolfers claim that deterrence results based on simultaneous equation models of crime are also suspect due to fragility and other issues related to selection of instrumental variables for these models. Among the studies they examine are Dezhbakhsh, Rubin, and Shepherd (2003) and Zimmerman (2004). Again, we focus on our own study, although some of the issues we raise apply to their criticism of Zimmerman's study as well.[31] Here we show that Donohue and Wolfers obtain some of their evidence in clear violation of the most basic statistical principles. Invalid estimates, and size-inflated inference, made through extensive data mining, are at the heart of their evidence.

Invalid Restrictions and Biased Estimates

One of Donohue and Wolfers' sensitivity checks involves changing the instruments we used to identify our simultaneous equation model. One such instrument is the partisan influence on the justice system. We measure the partisan influence by the Republican

[31] Zimmerman (2006b) has criticized Donohue and Wolfers for mischaracterizing his results and making errors in their analysis.

presidential candidate's percentage of statewide vote in the most recent election. This instrument is intended to capture any political pressure to get tough with criminals, a message popular with Republican candidates.[32] Given that hot issues vary among elections and the correlation between support for republican candidates and capital punishment also varies among elections, it is reasonable to allow the effect of this variable on the justice system to vary with election cycles.

Accordingly, we use six variables each measuring the Republican influence for one of the six elections during our sample period. Donohue and Wolfers, however, thought we had used only one variable to capture partisan influences.[33] So, when replicating our results, they used only one variable to capture the effect for all six elections. The deterrence result superficially switched and they reported it as one of their sensitivity checks. We show here that this accidental sensitivity check is based on biased estimation and invalid inference.

More specifically, in relation to the choice of instruments, Donohue and Wolfers state (page 826) that:

> *"Our point is not that one specification is preferable to the other. Indeed, sorting that out would be a difficult task."*

We disagree; *"sorting that out"* is a simple statistical task that can be performed easily and quickly using many statistical packages. But the implication is immense. Donohue and Wolfers specification assumes that the effect of partisan influences do not vary across elections, forcing it to be estimated with only one parameter. We, on the other hand, allow

---

[32] The influence is exerted through changing the makeup of the court system by appointing new judges or prosecutors that are "tough on crime." This affects the justice system and its desire to convict and execute criminals.

[33] They admit this point in footnote 84. Also, in our description of the partisan influence (PI) we should have stated "six PI variables" instead of "PI" to make the point clear. Nevertheless, as Donohue and Wolfers acknowledge in page 823, they had our computer programs that show we use six PI variables.

the data to determine whether the effect varies or not by estimating six parameters for the six elections. Only if the estimates of these six parameters turn out to be statistically the same, can one assume that the effect does not vary across elections.

The validity of Donohue and Wolfers restriction that the six coefficients are the same can be tested using a simple Wald test. As emphasized in most introductory econometric textbooks, if an invalid restriction is imposed on parameters of a regression equation, then the resulting coefficient estimates will be biased.[34] If the constant-over-election restriction that Donohue and Wolfers impose on the partisan variable does not hold, then their coefficient estimates are biased and their inference invalid.

We apply a Wald test of these restrictions to various first stage equations corresponding to the three deterrence probabilities in all models examined in Donohue and Wolfers' Table 7. The test results for the fifteen distinct equations are reported in Table 2. All test statistics are strongly significant with p-values that are near zero. This suggest that the restrictions imposed by Donohue and Wolfers are inconsistent with the data, invalid, and should be rejected. As a result, their instrumental variable estimation is biased. A biased estimate diverges from the true parameter values, and what they interpret as fragility of our estimate in this case is due to their estimation bias. Admitting estimates from clearly misspecified equations into sensitivity analysis is contrary to the norms of appropriate analysis.[35]

This is not the only gross statistical error we have detected in Donohue and Wolfers' sensitivity checks. If fact, a more disguised error occurs in their *"Effects in State-Years in*

[34] This is because an incorrect restriction forces all coefficient estimates to have the same value. But the corresponding coefficients are not equal, so imposing equality causes estimation bias. Note that while imposing a wrong restriction causes estimation bias, failure to impose a correct restriction reduces the accuracy of estimation (efficiency loss); see, e.g., Greene (2003), chapter 8.
[35] See, e.g., McAleer, Pagan, and Volker (1985) and Granger and Uhlig (1990).

Which There In No Death Penalty" reported in Panel B of their Table 8. Here, they estimate our baseline models after restricting the sample to those observations occurring when the state did not have the death penalty. They argue that the results should not show any deterrence effect for these states if our models are correctly specified. But in absence of death penalty laws, probability of execution is measured by zero because it is estimated by number of executions divided by number of convictions. These zero values must be used as the dependent variable in the first stage of the estimation.

How can one run a fist stage regression when all values of the dependent variable are zero, as is the case for execution probability equation? Donohue and Wolfers perform this impossible task by using the entire sample in the first stage of the estimation but dropping the observations for states with death penalty in the second stage. They admit to this inappropriate mid-estimation sample switching in footnote 87. Using the full sample, which also includes all death penalty states, leads to execution probability estimates that are positive instead of zero. So they replace the execution probabilities for non-death penalty states in the crime equation with these non-zero estimates that are biased. Obviously, this leads to biased coefficient estimate of crime equation for the selected non-death penalty sample.[36] This is a trivial exercise that only detects a bias that they have induced themselves through this invalid exercise. There are other statistical errors in their study, although less severe than the ones reported above.[37]

---

[36] It is also not surprising that they find a strong deterrent effect in such cases, because by construction they create biased estimates that suggest executions take place in non-death penalty states.

[37] For example, the difference-in-difference mean analysis reported in their table 1 panels B, C, and D, listed under "Our Innovation", contains an error. What they want to do in Penal D is to perform a matched (or paired) comparison, whereby the crime change for each state is subtracted from the crime change for its matched pairs, and then a one-sample location (mean) test is performed on these differences. What they do instead is a two sample comparison, whereby the overall average obtained for the treatment group is subtracted from the overall average obtained for the control group (averages reported in panel C is subtracted from the corresponding ones in Panel B). Calculation of the standard errors and the t-tests for the

Mining the Data to Determine What Data to Throw Away

Another piece of evidence that Donohue and Wolfers present to support their fragility claim involves estimating our baseline models without either Texas or California. In both cases, the deterrence result changes for some models. As we argued earlier, throwing out data that contain useful information and cannot be classified as statistical outlier is inappropriate. While Donohue and Wolfers had earlier argued that selection of Texas is due to its relatively large number of executions, they offer no reason why they exclude California. California, indeed, falls in the mid range of the cross state distribution of executions and not in the high or low ends.

To shed some light on Donohue and Wolfers' choice here and to put their results in proper perspective, we reexamined our baseline model 4, which according to their results shows sign change for both Texas and California.[38] Accordingly, we estimated this model 50 times, each time excluding one of the fifty states. We found that in 46 cases the deterrence result holds strongly with negative and significant deterrence coefficient estimates. Only in four cases—that involves excluding Texas, California, Virginia, or Illinois—did the results change. Texas and California are the two states that Donohue and Wolfers claim they "arbitrarily" chose for selectivity check.[39] It is highly unlikely that their selection was arbitrary. In fact the chances that these two states were picked randomly is less than one-half of one percent.[40] Rather, their choice seems to be the result of an

---

two approaches are different; one uses a one-sample t-test and the other the two-sample t-test, and the results can be different as well; see, e.g., Sheskin (1997). Correcting this gross error may not alter their finding, but such errors diminish considerably the readers' trust in a statistical study. Donohue and Wolfers also offer no description of the matching process for states and no list of the matched states.

[38] See, Donohue and Wolfers (2005) table 7, panels D and E.

[39] Donohue and Wolfers (2005), top of page 827.

[40] The first pick has a probability of 4/50 and the second pick a probability of 3/49 and the probability of the intersection is .0048.

extensive sample selection search whereby they throw out one state at a time. Instead of reporting the full range of the resulting estimates, however, they only report two of the non-deterrence cases without any mention of the other 46 cases where the deterrence result holds. This is, indeed, data mining, a practice that, ironically, sensitivity analysis is supposed to safeguard against.[41]

We see two serious problems with this data elimination exercise and the way it is reported by Donohue and Wolfers, one is technical and the other ethical. On the technical side, it is inappropriate to throw away data, especially in cases where the data is quite relevant. Deterrence is the result of exercising capital punishment laws rather than having the law without applying it. So states like Texas are quite relevant for the analysis. Moreover, estimating a large number of models in search of a particular outcome leads to statistical significance levels—also known as test size or probability of type-I error—that are much larger than the conventional 1%, 5%, and 10% levels.[42] Therefore, comparing these results with our results that were obtained from a single estimation with the full sample is inappropriate.

Furthermore, we believe that concealing the extent of the selection search (data mining), and characterizing the selection as arbitrary is inappropriate and misleading.[43] Assume, for the sake of argument, that Donohue and Wolfers had stated that they estimated fifty models by dropping one state at a time and found a deterrent effect for capital

[41] It is not difficult to change the magnitude or sign of an estimate through *ex post* manipulation of the sample, even in simple statistical inference. For example, consider estimation of the mean parameter using a small sample from a normal distribution. The resulting mean estimate is unbiased, but one can always try various combinations of data points to find a subset of the sample that its elimination changes the mean estimate significantly. The larger the variance of the distribution, the easier it is to change the mean estimate by changing sample, as the observations are more disperse and their elimination exerts more influence on the resulting mean estimate. Obviously, such manipulation is easier in more complex models.
[42] Lovell's (1983) simulation of a data miner's selection of regressors with various search strategies reveal actual significance levels of close to 50% where the nominal significance levels were only 5%.
[43] See, e.g. Lovell (1983), page 11.

punishment in 46 cases. Such full disclosure would give the reader a vastly different impression than their present claim that arbitrary selection of these two states reversed the deterrence results. The selective reporting of two estimates while veiling the full frequency distribution obtained from fifty estimates is hardly expected of an inquiry that is solely intended to set the records straight by revealing the uses and abuses of empirical evidence.

Invalid Instrumental Variable Test and Erroneous Interpretation

Donohue and Wolfers claim they have used Hausman specification test to show that the instruments in Dezhbakhsh, Rubin, and Shepherd (2003) are not valid.[44] Their claim, however, is based on misunderstanding the simultaneous equation model in our study, invalid application of Hausman test, and use of visual comparison instead of a statistical metric to make inference. Each of these is a serious error as we explain next.

We have four equations with four endogenous variables that include crime rate as well as the three probabilities of execution, conviction, and arrest:

$$M=f(P_a, P_{c|a}, P_{e|c}, Z, TD, u_1) \quad (1)$$

$$Pa=g(M, PE, TD, u_2) \quad (2)$$

$$Pc|a=h(M, JE, PI, PA, TD, u_3) \quad (3)$$

$$Pe|c=k(M, JE, PI, TD, u_4) \quad (4)$$

where M is murder rate, P is probability (conditional probability) with a, c, and e denoting arrest, conviction, and execution (respectively), Z is a set of demographic/economic variables, PE is police payroll expenditure, JE is expenditure on judicial and legal system, PI is six partisan influence variables as measured by the Republican presidential candidate's

[44] In its general form, Hausman specification test involves a statistical comparison of two estimators of the same set of parameter(s). Under the null hypothesis of no misspecification, these two sets of estimates are statistically equal. In the present context the null hypothesis is that the instruments are valid.

percentage of the statewide vote in the most recent six elections (discussed above), PA is prison admission, TD is a set of time dummies that capture national trends, and u's are regression errors.[45]

We estimate these equations via the two stage least squares (2SLS) method; see Dezhbakhsh, Rubin, and Shepherd (2003), pages 356-360.[46] This is equivalent to instrumental variable estimation of these equations where the instruments are the predicted (fitted) values of the endogenous variables (M and P's) from the first stage estimation of the reduced form equations. These relatively efficient instruments are linear combinations of *all* exogenous variables in the system, and not just a subset of them.

Donohue and Wolfers misunderstood the identification of our simultaneous equations, and consequently made a mistake in identification of our crime equation when obtaining various estimates for Hausman test comparison. The confusion and the resulting mistake have been carried over to the analysis reported in their tables 8 and 9. In particular, they consider *only* nine of the exogenous variables as instruments, and exclude some of these variables from the instrument list to obtain alternative estimates for comparison; see panels C and D of their table 8. However, they obtain their comparison estimates incorrectly. As acknowledged in notes (b) and (c) in their table 8, they use the excluded instruments[47] as control variables in the crime equation when estimating the models reported in this table. But our crime equation (1) does not include any of these instruments as control. In fact, it

---

[45] The subscripts i and t are dropped and the linear equations are written in general form for expositional ease.

[46] The first stage involves estimating the reduced forms of equations 1-4 and the second stage involves re-estimating these equation while replacing the endogenous variables on the right hand side with their predicted values from the first stage. Also, appropriate correction is made to the standard errors to account for the fact that the residuals in the 2SLS estimation is the difference between the left hand side variable and the linear combination of the right hand side variables and not their predicted values, as obtained when least squares is applied in stage 2 of estimation; see, e.g., Davidson and MacKinnon (1993), section 7.5

[47] They obtain alternative estimates by (i) excluding PE, JE, and PA, or (ii) excluding the six PI variables from the set of instruments.

does not make sense to assume that crime rate is directly affected by variables such as partisan influence or judicial expenditures, rather than the indirect affect captured through P_a, $P_{c|a}$, $P_{e|c}$. So Donohue and Wolfers have added inappropriate variables to equation (1), instead of simply dropping some of the exogenous variables from equations (2)-(4) as a way of revising the instrument list. What they find, therefore, is the misspecification they have introduced into the crime equation rather than invalid instruments.[48]

Notwithstanding the above error, as serious as it might be, Donohue and Wolfers do not even report any statistics—using actual coefficient estimates, their corresponding variances, and p-values—for their Hausman test.[49] They instead use eyeball detection to compare lives-saved estimates, which are functions of inappropriate estimates of the actual coefficients they obtain through changing the specification of the crime equation. This violates the standard statistical practice of using formal metrics based on an actual coefficient estimate rather than visual comparisons of a function of this estimate to make statistical inference.[50] The above visual comparison is their second test of our instruments. The first test was the biased mid-sample data switching exercised that we discussed and rejected earlier.

[48] It is not clear whether the estimates in their Table 9, that also correspond to what they call instruments, are obtained by including the demographic/economic and trend variables in their estimation. If not then these results are also invalid.

[49] See Hausman (1978) or many standard econometric textbooks; e.g., Greene (2003, chapter 5), Johnston and DiNardo (1997, chapter 8), and Davidson and MacKinnon (1993, chapter 11).

[50] Even if their comparison estimates were correctly obtained, visual comparison would not be appropriate for a Hausman test, because the variance of the difference between alternative estimates is not the difference in the corresponding variances, so reporting two estimates and the corresponding standard errors is not very telling, no matter how different they are.

VI. Concluding Remarks

Donohue and Wolfers' Only Undisputable Finding: Dewald, Thursby, and Anderson (1986) report the results of a two year study funded by the National Science Foundation to examine replication of empirical research in economics . The finding of the study, which is known as the J*MCB* project, was alarming: many empirical results could not be replicated due to poor documentation and inadvertent errors that seemed to be a commonplace rather than a rarity. This finding prompted some economic journals to change their editorial policy, requiring submission of data or full documentation of empirical analysis. That was twenty years ago, but how different is the state of affairs today?

Donohue and Wolfers' study has one undisputable contribution. It shows that most of the capital punishment deterrence results they examine can be easily replicated.[51] This finding, that is unfortunately lost in the mirage of fragility fishing, has important implications for the death penalty literature and, more importantly, for the economics profession. If the replication success is, indeed, equally pervasive in other areas of economic inquiry, then we have come a long way since the *JMCB* project. If the replication ease is still a rarity, then the death penalty literature deserves special credit that must be noted by policy makers.

Obviously, more work is still needed to provide appropriate protocol and guidelines to help "de-con" the practice of sensitivity analysis and promote proper and impartial sensitivity checks on published work. We reiterate that even in a simple inference like estimation of the mean of a random sample, one can search through the data points to identify sub-samples that produce mean estimates which are vastly different from the

[51] See, e.g., Donohue and Wolfers (2005), pages 811, 816, 823, and 827 for replication acknowledgements.

unbiased and efficient estimate based on the entire sample. Such counter evidence is hardly credible. In more complex models, fishing for counter evidence is even easier. One can search through various specification, valid or invalid, to obtain a different result. Reporting such results might not be disturbing, but concealing the scope of the search for counter evidence and hiding many results that validates the original finding is.

Summary: Donohue and Wolfers examine the robustness of a few of the recent studies on the deterrent effect of capital punishment. They claim based on their results that *"there is little evidence to convince believers in the deterrent hypothesis"*.[52] The obituary of "the deterrence findings" that they write is based on a selective sample, flawed analysis, biased reporting, and violation of guidelines for objective sensitivity analysis (see Section II).

For example, among the 21 studies that examine the deterrence hypothesis, they examine only four studies for robustness.[53] In fact, six of their nine tables relate to only two studies.[54] It is not clear whether such limited scope is the outcome of sample selectivity or reporting selectivity. In either case, the narrow scope of Donohue and Wolfers' inquiry does not warrant such a grandiose claim. More importantly, their sensitivity analysis of our two studies that they extensively cover is replete with statistical errors, biased reporting, concealing results, and misleading conclusions.

Some of these errors could have been easily avoided had they followed normal scientific procedures by seeking comments from the concerned authors. But they only sent

[52] Donohue and Wolfers (2005), page 844.

[53] See the introduction of this paper for a list of these 21 studies. Donohue and Wolfers have been aware of all but perhaps one (Fagan, Zimring, and Geller, 2006) of these studies which have all been cited in Dezhbakhsh and Shepherd (2006) or Shepherd's congressional testimony that Donohue and Wolfers cite in footnote 11 of their 2006 study. The four studies they examine include Dezhbakhsh, Rubin, and Shepherd (2003), Dezhbakhsh and Shepherd (2006), Katz, Levitt, and Shustorovich (2003), and Mocan and Gittings (2003). They also briefly touch on three others studies by Cloninger and Marchesini (2001 and 2005) and Zimmerman (2004), without a detailed sensitivity analysis.

[54] Six of the nine tables reporting their sensitivity checks are devoted to two studies, Dezhbakhsh and Shepherd (2006) and Dezhbakhsh, Rubin, and Shepherd (2003).

us their paper when it was about to go to print.[55] More importantly, they chose to publish their econometric study in a student-refereed law review rather than a peer-refereed economics journal. Law students have no particular expertise in econometrics to identify the aforementioned errors. Obviously, the flaws in the paper could have been detected by expert refereeing, if they had submitted their paper to a peer reviewed outlet.[56] We did ask *the Stanford Law Review* the right to reply, but were not granted this right.

Finally, most of Donohue and Wolfers' results are, ironically, obtained through extensive data mining, a practice that sensitivity analysis is supposed to safeguard against. Moreover, contrary to the inclusivity norms of sensitivity analysis, they report some estimates of the key parameters, while remaining silent about other estimates. Their selectivity strongly favors anti-deterrence. We believe that Donohue and Wolfers would have benefited from listening to their own words:

> *"…potential dangers [are] awaiting those who might wish to short-circuit the full process of scientific inquiry and validation…"*[57]

May our collective wisdom advance science and enlighten policy.

[55] Donohue and Wolfers e-mailed us their paper Friday night December 2, 2005, asking for input by Monday December 5, 2005—the date Stanford Law Review had given them to make further revisions, as they stated in their e-mail. Despite various prior e-mails where they sought our help with data and computer codes, they never communicated to us the exact purpose of their study or their findings until that Friday evening. Moreover, the paper was already in *the Stanford Law Review* format, indicating that they had prepared it much earlier.

[56] We find it quite amusing that Wolfers in an interview with Tanner (2007) refers to deterrence findings as "flimsy [results] that appeared in second-tier journals." The studies he refers to have gone through rigorous peer review process, often blind review, and appeared in economics peer reviewed journals, while his finding with Donohue, that he uses to call the deterrence studies flimsy, has never been peer reviewed and appears only in a student edited/refereed journal.

[57] Donohue and Wolfers (2005), page 845.

Appendix

The most important task in sensitivity analysis is to choose the direction of inquiry over a large dimensional assumption space. This makes the task hard for those who want to conduct a thorough analysis and easy for those who want to fish for counter evidence no matter how inappropriate or irrelevant it might be. The choices should, indeed, be made while keeping in mind that not all estimates are created equal. The following example helps illustrate this point. Consider the regression equation

$$Y_i = \beta X1_i + \gamma\, X2_i + u_i\,, \tag{1}$$

where u's are regression errors and β and γ are unknown vector of parameters. Assume researcher A estimates the above regression equation and obtains an estimate of β which we denote by $\hat{\beta}_A$. Under standard assumptions, the sampling distribution of $\hat{\beta}_A$ is centered on β. So while not every $\hat{\beta}_A$ obtained from some sample is equal to β, these values are close to β on average. Now assume researcher B decides to drop X2 from the regression and reestimate β. Denote the resulting estimate by $\hat{\beta}_B$. The distribution of $\hat{\beta}_B$ is centered on $\beta+\Delta(\rho)$ where $\Delta(\rho)$ is a function of the cross moments of X1 and X2. If X1 and X2 are orthogonal then $\Delta=0$, otherwise $\Delta\neq0$. A nonzero Δ will be the source of the difference between the estimates reported by the two researchers.[58]

What does the difference between the two estimates imply? If there is a reason to believe that X2 does, indeed, belong to equation 1, then $\hat{\beta}_B$ is a biased estimate and should not be included in the sensitivity analysis. In fact, if the data are time series, dropping X2 will most likely cause serial correlation in errors of the resulting equation. The Durbin-

---

In fact, Δ is proportional to the simple correlation between X1 and X2.

Watson statistic will then turn significant, signaling specification problem and prompting the researcher to exclude $\hat{\beta}_B$ from the analysis as suggested by Granger and Uhlig (1990). For example, if equation (1) is a sales equation for a product in a market with some competition, X1 is price of the product, and X2 is price of a rival product, then dropping X2 from the equation and estimating β_1 results in a biased estimate. The difference between this biased estimate $\hat{\beta}_B$ and the original estimate $\hat{\beta}_A$ does not indicate that $\hat{\beta}_A$ is fragile, it rather reflects a bad move on the part of researcher B. The emphasis of the Leamer's method on prior selection of free parameters, that are precluded from elimination, is to avoid such mistake. The specification checks of McAleer, Pagan, and Volker (1985), diagnostic checks of Granger and Uhlig (1990), and theory-based restrictions of Ehrlich and Liu (1999) are also in a similar spirit.

References

Albert, C. J. 1999. "Challenging Deterrence: New Insight on Capital Punishment Derived from Panel Data." 60 *University of Pittsburgh Law Review* 321-371.

Bailey, William C. 1998. "Deterrence, Brutalization, and the Death Penalty: Another Examination of Oklahoma's Return to Capital Punishment." 36 *Criminology* 711-733.

Bartley, William A. and Mark A. Cohen. 1998. "The Effects of Concealed Handgun Laws: An Extreme Bounds Analysis." 36 *Economic Inquiry* 258-265.

Berk, Richard. 2005. "New Claims About Executions and General Deterrence: Déjà Vu All Over Again?." 2 Journal of Empirical Legal Studies 303-330.

Black, T. and T. Orsagh. 1978. "New Evidence on the Efficacy of Sanctions as a Deterrent to Homicide," 58 *Social Science Quarterly* 616-631.

Bowers, W.J. and J.L. Pierce. 1975. "The Illusion of Deterrence in Isaac Ehrlich's work on Capital Punishment," 85 *Yale Law Journal* 187-208.

Brumm, Harold J. and Dale O. Cloninger. 1996. "Perceived Risk of Punishment and the Commission of Homicides: A Covariance Structure Analysis," 31 *Journal of Economic Behavior and Organization* 1-11.

Cameron, Samuel. 1994. "A Review of the Econometric Evidence on the Effects of Capital Punishment," 23 *Journal of Socio-Economics* 197-214.

Chatfield, Chris. 1995. "Model Uncertainty, Data Mining and Statistical Inference." 158 *Journal of the Royal Statistical Society*, *Series A* 419-466.

Chressanthis, George A. 1989. "Capital Punishment and the Deterrent Effect Revisited: Recent Time-Series Econometric Evidence," 18 *Journal of Behavioral Economics* 81-97.

Cloninger, Dale, O. 1977. "Deterrence and the Death Penalty: A Cross-Sectional Analysis," 6 *Journal of Behavioral Economics* 87-107.

Cloninger, Dale O. and Roberto Marchesini. 2001. "Execution and Deterrence: A Quasi-Controlled Group Experiment," 35 *Applied Economics* 569-576.

Cloninger, Dale O. and Roberto Marchesini. 2005. Execution Moratoriums, Commutations and Deterrence: The Case of Illinois. Economics Working paper Archives, Working Paper No. 0507002, also in 2006. 38 *Applied Economics*.

Cloninger, Dale O. and Roberto Marchesini. 2007. Reflections on a Critique, *Applied Economic Letters* (forthcoming).

Cover, James Peery and Paul D. Thistle. 1988. "Time Series, Homicide, and the Deterrent Effect of Capital Punishment," 54 *Southern Economic Journal* 615-22.

Cook, R.D. 1977. "*Detection of Influential Observations in Linear Regression. 19 Technometrics* 15-18.

Cook R.D. and S. Weisberg. 1982. *Residuals and Influence in Regression*. Chapman and Hall :New York.

Davidson, Russell and MacKinnon, James G. 1993. *Estimation and Inference in Econometrics*. Oxford University Press: New York.

Dezhbakhsh, Hashem, Paul H. Rubin, and Joanna M. Shepherd. 2003. "Does Capital Punishment Have a Deterrent Effect? New Evidence from Post Moratorium Panel Data." 5 *American Law and Economics Review* 344-376.

Dezhbakhsh, Hashem and Joanna M. Shepherd. 2006. "The Deterrent Effect of Capital Punishment: Evidence from a Judicial Experiment." 14 *Economic Inquiry* 512-535.

Dewald, William. G., Jerry G. Thursby and Richard G. Anderson. 1986. "Replication in Empirical Economics: The Journal of Money Credit and Banking Project." 76 *American Economic Review* 587-603.

Donohue, J. and Justin Wolfers. 2005. "Uses and Abuses of Empirical Evidence in The Death Penalty Debate." 58 *Stanford Law Review* 791-845.

Ekelund, Robert B., John D. Jackson, Rand W. Ressler, and Robert D. Tollison. 2006. "Marginal Deterrence and Multiple Murders." 72 *Southern Economic Journal* 521-541

Ehrlich, Isaac. 1975. "The Deterrent Effect of Capital Punishment: A Question of Life and Death," 65 *American Economic Review* 397-417.

Ehrlich, Isaac. 1977. "Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence," 85 *Journal of Political Economy* 741-788.

Ehrlich, Isaac and Joel Gibbons. 1977. "On the Measurement of the Deterrent Effect of Capital Punishment and the Theory of Deterrence," 6 *Journal of Legal Studies* 35-50.

Ehrlich, Isaac and Zhiqiang Liu. 1999. "Sensitivity Analysis of the Deterrence Hypothesis: Lets Keep the Econ in Econometrics," 42 *Journal of Law and Economics* 455-488.

Eysenck, H. 1970. *Crime and Personality*. London: Paladin.

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. 2006. "Capital Punishment and Capital Murder: Market Share and the Deterrent Effects of the Death Penalty." Forthcoming.

Falls, Carrie. 2005. "Specification Testing in Panel Data Models Estimated by Fixed Effects with Instrumental variables." Working Paper. Michigan State University.

Fowles, Richard and Peter D. Loeb. 1995. "Effects of Policy-Related Variables on Traffic fatalities: An Extreme Bounds Analysis Using Time-Series Data." 62 *Southern Economic Journal* 359-366.

Granger, Cliv W.J. and Harald F. Uhlig. 1990. Reasonable Extreme Bounds Analysis." 44 *Journal of Econometrics* 159-170.

Greene, William H. 2003. *Econometric Analysis*. 5th Edition. Prentice Hall: Upper Saddle River, NJ.

Grogger, Jeffrey. 1990. “The Deterrent Effect of Capital Punishment: An Analysis of Daily Homicide Counts,” 85 *Journal of the American Statistical Association* 295-303.

Hand, David J. 1998. “Data Mining: Statistics and More.” 52 *American Statistician*. 112-118.

Hausman, Jerry A. 1978. Specification Tests in Econometrics. 46 *Econometrica* 1251-1271.

Hausman, Jerry A. 1983. “Specification and Estimation of Simultaneous Equation Models.” In Z Griliches and M.D. Intriligator (Eds.) *Handbook of Econometrics*, Volume I, North Holland: Amsterdam, 391-448.

Hendry, David. 1980. “Econometrics-Alchemy or Science?.” 47 *Economica* 387-406.

Hoenack, Stephen A. and William C. Weiler. 1980. “A Structural Model of Murder Behavior and the Criminal Justice System.” 70 *American Economic Review* 327-41.

Jevons, William S. 1869. “The Depreciation of Gold.” 32 *Journal of the Royal Statistical Society* 445-449.

Johnston, Jack and John DiNardo. 1997. *Econometric Methods* 4th Edition McGraw-Hill: New York.

Katz, L., S. D. Levitt, and E. Shustorovich. 2003. “Prison Conditions, Capital Punishment, and Deterrence.” 5 *American Law and Economics Review* 318-343

Layson, Stephen. 1985. “Homicide and Deterrence: A Reexamination of the United States Time-Series Evidence,” 52 *Southern Economic Journal* 68-89.

Leamer, Edward and Herman Leonard. 1983. “Reporting the Fragility of Regression Estimates.” 65 *Review of Economics and Statistics* 306-317.

Leamer, Edward. 1994. *Sturdy Econometrics*. Brookfield, Vermont: Edward Elgar Publishing.

Leamer, Edward. 1983. “Let’s Take the Con out of Econometrics.” 73 *American Economic Review* 31-43.

Leamer, Edward. 1985. “Sensitivity Analysis Would Help,” 75 *American Economic Review* 308-313.

Liu, Zhiqiang. 2004. “Capital Punishment and the Deterrence Hypothesis: Some New Insights and Empirical Evidence.” 30 *Eastern Economic Journal* 237-258.

Lott, John R. Jr. and David B. Mustard. 1997. “Crime, Deterrence and Right-to-Carry Concealed Handguns,” 26 *Journal of Legal Studies* 1-69.

Lott, John R., Jr. and William M. Landes. 2000. “Multiple Victim Public Shootings.” Olin Law and Economics Working Paper No. 73. University of Chicago.

Lovell, Michael C. 1983. “Data Mining.” 65 *Review of Economics and Statistics* 1-12.

McAleer, Michael and Michael R. Veall. 1989. “How Fragile are Fragile Inferences? A Re-evaluation of the Deterrent Effect of Capital Punishment,” 71 *Review of Economics and Statistics* 99-106.

McAleer, Michael, Adrian R. Pagan, and Paul A. Volker. 1985. “What Will Take the Con out of Econometrics.” 75 *American Economic Review* 293-307.

McKee, D.L., and M.L. Sesnowitz. 1977. "On the Deterrent Effect of Capital Punishment," 6 *Journal of Behavioral Economics* 217-224.

McManus, W. 1985. "Estimates of the Deterrent Effect of Capital Punishment: The Importance of the Researcher's Prior Beliefs," 93 *Journal of Political Economy* 417-425.

Mocan H. Naci and R. Kaj Gittings. 2003. "Pardons, Executions, and Homicides." 46 *Journal of Law and Economics* 453-478.

Mocan, Naci H. and R. Kaj Gittings. 2006. "The Impact Of Incentives On Human Behavior: Can We Make It Disappear? The Case Of The Death Penalty," National Bureau of Economic Research, Working Paper 12631, http://www.nber.org/papers/w12631

Passell, Peter and John B. Taylor. 1977. "The Deterrent Effect of Capital Punishment: Another View," 67 *American Economic Review* 445-51.

Quinn, D. 1997. "The Correlates of Change in International Financial Regulation." 91 *American Political Science Review* 531-551.

Sellin, J. T. 1959. *The Death Penalty.* Philadelphia, PA: American Law Institute.

Shepherd, J. M. 2004. "Murders of Passion, Execution Delays, and the Deterrence of Capital Punishment." 33 *Journal of Legal Studies* 283-321.

Sims, Christopher. 1980. "Macroeconomics and Reality." 48 *Econometrica* 1-48.

Sjoholm, Fredrik. 1996. "International Transfer of Knowledge: The Role of International Trade and Geographic Proximity." 132 Weltwirtschaftliches Archiv 97-115.

Sorenson, Jon, Robert Wrinkle, Victoria Brewer, and James Marquart. 1999. "Capital Punishment and Deterrence: Examining the Effect of Executions on Murder in Texas." 45 *Crime & Delinquency* 481-493.

Stigler, Stephen M. 1986. *The History of Statistics, The Measurement of Uncertainty before 1900*. Harvard University Press: Cambridge, MA.

Stolzenberg, Lisa and Stewart J. D'Alessio. 2004. "Criminology: Capital Punishment , Execution Publicity and Murder in Houston, Texas." 94 *Journal of Crime Law and Criminology* 351-379.

Sunstein, C.R. and A. Vermeule. 2005. "Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs. 58 *Stanford Law Review* 703-750.

Tanner, Robert. June 10 2007. "Studies Say Death Penalty Deters Crime". Associated Press.

Welsch R.E. 1982. Influence Functions and Regression Diagnostics. In R.L. Launer and A.F. Siegel's Modern Data Analysis Academic Press: New York 149-169.

Wooldridge. Jeffrey M. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Wooldridge. Jeffrey M. 2003. *Introductory Econometrics: A Modern Approach*. Thomson Publishing: U.S.A.

Yunker, James A. 1976. "Is the Death Penalty a Deterrent to Homicide? Some Time Series Evidence, " 5 *Journal of Behavioral Economics* 45-81.

Yunker, James A. 2002. "A New Statistical Analysis of Capital Punishment Incorporating U.S. Postmoratorium Data." 82 *Social Science Quarterly* 297-311.

Zimmerman, P. R. 2004. "State Executions, Deterrence, and the Incidence of Murder." 7 *Journal of Applied Economics* 163-193.

Zimmerman, P. R. 2006a. "Estimates of the Deterrent Effect of Alternative Execution Methods in the United States: 1978-2000." *American Journal of Economics and Sociology,* Forthcoming, *available at* http://papers.ssrn.com/sol3/papers.cfm?abstract_id=355783.

Zimmerman, P. R. 2006b. "On the Use and "Abuse" of Empirical Evidence in the Death Penalty Debate," Manuscript.

Table 1: Sensitivity Results for Dezhbakhsh and Shepherd(2006), Based on Donohue and Wolfers (2006) Suggestions

Variable	1	2	3	4	5	6	7	8	9	10	11	12	13	14
Executions	-0.1***			-0.1***				-0.2***			-0.1**			
	(10.58)			(3.19)				(4.53)			(2.21)			
Executions Lagged		-0.2***		-0.1***					-0.2***		-0.1***			
		(6.59)		(2.95)					(5.1)		(2.8)			
Moratorium Indicator			0.9***	0.9***			0.8***			1***	0.9***			1***
			(4.25)	(3.86)			(3.84)			(4.77)	(4.44)			(4.49)
Executions per 100,000 residents					-12.3***		-9.6***					-5.8***		-4.5***
					(5.47)		(4.06)					(3.14)		(2.99)
Lagged Executions per 100,000 residents						-13.3***	-8.7***						-7.2***	-5.1***
						(5.51)	(3.31)						(3.62)	(2.77)
	Further Controls													
Texas Included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No
Type of Time Dummies?	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade
Moratorium Measure?	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure
Cluster Members?	States	States	States	States	States	States	States	States	States	States	States	States	States	States

Notes: See end of table.

Table 1 (continued): Sensitivity Results for Dezhbakhsh and Shepherd(2006), Based on Donohue and Wolfers (2005) Suggestions

Variable	15	16	17	18	19	20	21	22	23	24	25	26	27	28
Executions	-0.1***			-0.1***				-0.1**			0			
	(10.84)			(2.81)				(1.94)			(0.73)			
Executions Lagged		-0.1***		-0.1***					-0.1***		-0.1			
		(7.15)		(2.4)					(2.44)		(1.18)			
Moratorium Indicator			0.3*	0.4*			0.3*			0.5**	0.5**			0.5**
			(1.39)	(1.47)			(1.35)			(1.92)	(1.84)			(1.84)
Executions per 100,000 residents					-6.7***		-5.9***					0.6		0.3
					(3.63)		(2.57)					(0.57)		(0.18)
Lagged Executions per 100,000 residents						-7***	-4.4**						0	-0.1
						(3.69)	(1.89)						(0.01)	(0.08)
	Various Specifications													
Texas Included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No
Type of Time Dummies?	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Moratorium Measure?	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure
Cluster Members?	States	States	States	States	States	States	States	States	States	States	States	States	States	States

Notes: See end of table.

Table 1 (continued): Sensitivity Results for Dezhbakhsh and Shepherd(2006), Based on Donohue and Wolfers (2005) Suggestions

Variable	29	30	31	32	33	34	35	36	37	38	39	40
Executions		-0.1***			-0.1**			-0.1***			0	
		(3.08)			(1.86)			(2.79)			(0.61)	
Executions Lagged		-0.1***			-0.1***			-0.1***			0	
		(2.9)			(2.67)			(2.36)			(1.1)	
Moratorium Indicator	0.7***	0.6***	0.6***	0.7***	0.6***	0.7***	0.2**	0.2*	0.2*	0.3***	0.3**	0.3***
	(4.29)	(3.61)	(3.57)	(4.6)	(4.08)	(4.23)	(2.07)	(1.38)	(1.39)	(2.4)	(2.11)	(2.35)
Executions per 100,000 residents			-9.1***			-3.8***			-5.7***			0.6
			(3.85)			(2.45)			(2.51)			(0.4)
Lagged Executions per 100,000 residents			-8.3***			-4.6***			-4.2**			0.2
			(3.21)			(2.53)			(1.83)			(0.12)
	Various Specifications											
Texas Included?	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No	No
Type of Time Dummies?	Decade	Decade	Decade	Decade	Decade	Decade	Year	Year	Year	Year	Year	Year
Moratorium Measure?	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto
Cluster Members?	States	States	States	States	States	States	States	States	States	States	States	States

Notes: See end of table.

Table 1 (continued): Sensitivity Results for Dezhbakhsh and Shepherd(2006), Based on Donohue and Wolfers (2005) Suggestions

Variable	41	42	43	44	45	46	47	48	49	50	51	52	53	54
Executions	-0.1***			-0.1***				-0.2***			-0.1***			
	(8.81)			(8.16)				(3.85)			(2.45)			
Executions Lagged		-0.2***		-0.1***					-0.2***		-0.1***			
		(8.37)		(6.87)					(3.48)		(2.71)			
Moratorium Indicator			0.9**	0.9**			0.8**			1**	0.9**			1**
			(1.91)	(1.85)			(1.81)			(2.05)	(2.01)			(2.02)
Executions per 100,000 residents					-12.3***		-9.6***					-5.8***		-4.5***
					(2.38)		(2.49)					(3.06)		(2.98)
Lagged Executions per 100,000 residents						-13.3***	-8.7***						-7.2***	-5.1***
						(2.56)	(2.57)						(3.31)	(2.66)
	Various Specifications													
Texas Included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No
Type of Time Dummies?	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade	Decade
Moratorium Measure?	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure
Cluster Members?	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years

Notes: See end of table.

Table 1 (continued): Sensitivity Results for Dezhbakhsh and Shepherd(2006), Based on Donohue and Wolfers (2005) Suggestions

Variable	55	56	57	58	59	60	61	62	63	64	65	66	67	68
Executions	-0.1***			-0.1***				-0.1*			0			
	(6.56)			(6.68)				(1.35)			(1.14)			
Executions Lagged		-0.1***		-0.1***					-0.1		-0.1			
		(6.05)		(6.15)					(1.25)		(1.16)			
Moratorium Indicator			0.3	0.4			0.3			0.5	0.5			0.5
			(0.67)	(0.76)			(0.68)			(0.95)	(0.94)			(0.95)
Executions per 100,000 residents					-6.7		-5.9*					0.6		0.3
					(1.19)		(1.34)					(0.31)		(0.15)
Lagged Executions per 100,000 residents						-7	-4.4						-0.0	-0.1
						(1.23)	(1.23)						(0.0)	(0.07)
	Various Specifications													
Texas Included?	Yes	Yes	Yes	Yes	Yes	Yes	Yes	No	No	No	No	No	No	No
Type of Time Dummies?	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year	Year
Moratorium Measure?	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure	De Jure
Cluster Members?	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years

Notes: See end of table.

Table 1 (continued): Sensitivity Results for Dezhbakhsh and Shepherd(2006), Based on Donohue and Wolfers (2005) Suggestions

Variable	69	70	71	72	73	74	75	76	77	78	79	80
Executions		-0.1***			-0.1**			-0.1***			0	
		(7.34)			(1.98)			(6.11)			(0.8)	
Executions Lagged		-0.1***			-0.1***			-0.1***			0	
		(6.96)			(2.67)			(5.47)			(1.03)	
Moratorium Indicator	0.7**	0.6**	0.6*	0.7**	0.6**	0.7**	0.2	0.2	0.2	0.3	0.3	0.3
	(1.99)	(1.65)	(1.58)	(2.11)	(1.75)	(1.9)	(0.71)	(0.49)	(0.45)	(0.91)	(0.75)	(0.87)
Executions per 100,000 residents			-9.1**			-3.8***			-5.7			0.6
			(2.27)			(2.32)			(1.26)			(0.31)
Lagged Executions per 100,000 residents			-8.3***			-4.6***			-4.2			0.2
			(2.42)			(2.42)			(1.14)			(0.1)
	Various Specifications											
Texas Included?	Yes	Yes	Yes	No	No	No	Yes	Yes	Yes	No	No	No
Type of Time Dummies?	Decade	Decade	Decade	Decade	Decade	Decade	Year	Year	Year	Year	Year	Year
Moratorium Measure?	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto	De Facto
Cluster Members?	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years	Years

Notes: The table presents estimated coefficients and absolute t-values (in parentheses). ***, **, and * denote significance at 1%, 5% and 10%, respectively.

Table 2: Wald Test for Equality of Partisan Influences Across Elections

	Probability of Arrest Equation	Probability of Death Row Sentence Equation	Probability of Execution Equation
Model 1	4.16	42.49	99.17
	(0.0009)	(0.00)	(0.00)
Model 2	4.16	26.72	220.45
	(0.0009)	(0.00)	(0.00)
Model 3	4.16	83.30	364.91
	(0.0009)	(0.00)	(0.00)
Model 4	5.07	42.49	118.45
	(0.0001)	(0.00)	(0.00)
Model 5	5.07	26.72	518.77
	(0.0001)	(0.00)	(0.00)
Model 6	5.07	83.30	438.42
	(0.0001)	(0.00)	(0.00)

Notes: The table presents the F-statistic for the Wald tests on equality of the partisan influence variables in the first-stage regressions. The P-values are in parenthesis.