ORIGINAL PAPER

# What Do Panel Studies Tell Us About a Deterrent Effect of Capital Punishment? A Critique of the Literature

**Aaron Chalfin · Amelia M. Haviland · Steven Raphael**

**Abstract**

*Objectives*   We provide a critical review of empirical research on the deterrent effect of capital punishment that makes use of state and, in some instances, county-level, panel data.
*Methods*   We present the underlying behavioral model that presumably informs the specification of panel data regressions, outline the typical model specification employed, discuss current norms regarding "best-practice" in the analysis of panel data, and engage in a critical review.
*Results*   The connection between the theoretical reasoning underlying general deterrence and the regression models typically specified in this literature is tenuous. Many of the papers purporting to find strong effects of the death penalty on state-level murder rates suffer from basic methodological problems: weak instruments, questionable exclusion restrictions, failure to control for obvious factors, and incorrect calculation of standard errors which in turn has led to faulty statistical inference. The lack of variation in the key underlying explanatory variables and the heavy influence exerted by a few observations in state panel data regressions is a fundamental problem for all panel data studies of this question, leading to overwhelming model uncertainty.
*Conclusions*   We find the recent panel literature on whether there is a deterrent effect of the death penalty to be inconclusive as a whole, and in many cases uninformative. Moreover, we do not see additional methodological tools that are likely to overcome the multiple challenges that face researchers in this domain, including the weak informativeness of the data, a lack of theory on the mechanisms involved, and the likely presence of unobserved confounders.

**Keywords**   Panel studies · Death penalty · Deterrence · Critical research review

A. Chalfin · S. Raphael
Goldman School of Public Policy, UC Berkeley, Berkeley CA, USA

A. M. Haviland (✉)
H. John Heinz III College of Public Policy and Management, Carnegie Mellon University,
Pittsburgh, PA, USA
e-mail: haviland@cmu.edu

 Springer

## Introduction

In 1976, the Supreme Court ruled that a carefully-crafted death penalty statute that meets specific criteria defined by the Court does not violate the Eighth Amendment prohibition against cruel and unusual punishment.[1] The decision effectively ended a 4-year death penalty moratorium created by an earlier 1972 Supreme Court ruling. Since the reinstatement of the death penalty, 7,773 inmates have been sentenced to death in the United States. Of these inmates, roughly 15% have been executed, 39% have been removed from death row due to vacated sentences or convictions, commutation, or a death sentence being struck down on successful appeal, while 5% have died from other causes (Snell 2010).

During this time, the legality of the death penalty and the frequency of death penalty sentences and executions have varied considerably across U.S. states and, to a lesser degree, for individual states over time. A small number of states contribute disproportionately to both the total number of executions as well as the number of inmates held under a death sentence. For example, between 1977 and 2008, a total of 1,185 inmates have been executed. Eight states accounted for 80% of these executions, with Texas alone accounting for 38%.[2] Meanwhile 16 states have not executed a single individual. Similarly, as of December 31, 2008, 3,159 state prison inmates were being held under a sentence of death. Six states (Alabama, California, Florida, Ohio, Pennsylvania, and Texas) accounted for 64% of these inmates (Snell 2010), while 15 states did not have a single inmate on death row.

Certainly, some of this variation is driven by the handful of states that do not impose the death penalty. As of 2009, 36 states authorize the use of the death penalty in criminal sentencing. However, even among those that do, there is considerable heterogeneity in actual outcomes. For example, of the 927 inmates sentenced to death in California between 1977 and 2009, 13 (1%) have been executed, 73 (8%) have died of other causes, while 157 (17%) have had their sentence commuted or overturned. The remaining inmates (684, 74%) are still awaiting execution. By contrast, of the 1,040 inmates sentenced to death in Texas over the same period, 447 (43%) have been executed, 38 (4%) have died of other causes, 224 (22%) have had their sentences overturned or commuted, while 331 (32%) remain on death row (Snell 2010, Table 20). Clearly, being sentenced to death in Texas means something different than being sentenced to death in California.

Over the last decade, empirical research has focused on the differential experiences of states with a death penalty regime to study whether capital punishment deters murder among the public at large. In particular, these studies have exploited the fact that in addition to cross-state differences in actual sentencing policy, there is also variation over time for individual states in the official sentencing regime, in the propensity to seek the death penalty in practice, and in the application of the ultimate punishment. With variation in de facto policy occurring both within states over time as well as between states, researchers have utilized state-level panel data sets to control for geographic and time fixed effects that may otherwise confound inference regarding the relationship between capital punishment and murder rates.

---

[1] See Gregg v. Georgia.

[2] The eight states in descending order are Texas, Virginia, Oklahoma, Florida, Missouri, Georgia, Alabama, and North Carolina. These tabulations are based on the spread sheet in Snell (2010), *Prisoners Executed Under Civil Authority in the United States by Year, Region and Jurisdiction, 1977–2008,* Bureau of Justice Statistics, http://bjs.ojp.usdoj.gov/index.cfm?ty=pbdetail&iid=2079, accessed on March 21, 2011.

While some researchers pursuing this path conclude that there is strong evidence of a deterrent effect of capital punishment on murder (e.g., Dezhbakhsh et al. 2003; Dezhbakhsh and Shepherd 2006; Mocan and Gittings 2003; Zimmerman 2004) there are also several forceful critiques of these findings (e.g., Berk 2005; Donohue and Wolfers 2005, 2009; Kovandzic et al. 2009). In turn, this has led to responses to those critiques (Dezbakhsh and Rubin 2010; Zimmerman 2009; Mocan and Gittings 2010) by the original authors. Across these and additional research papers, the varying results and wide-ranging discussion of these results has lead to uncertainty regarding the degree to which this body of research is substantively informative for policy makers.

In this paper, we provide a critical review of empirical research on the deterrent effect of capital punishment that makes use of state and, in some instances county-level, panel data. We begin with a conceptual discussion of the underlying behavioral model that presumably informs the specification of panel data regressions and outline the typical model specification employed in many of these studies. This is followed by a brief discussion of current norms regarding "best-practice" in the analysis of panel data and with regards to robust statistical inference. Finally, we engage in a critical review of much of the recent panel data research on the deterrent effect of capital punishment.

Regarding the conceptual discussion, we are of the opinion that the connection between the theoretical reasoning underlying general deterrence and the regression models typically specified in this literature is tenuous. Most studies estimate the empirical relationship between murder rates and measures of death penalty enforcement, including the number of death sentences relative to murder convictions, and the number of executions relative to varying lags of past death sentences. Variation in these explanatory variables is driven either by changes in de jure policy, changes in de facto policy, variation in crime rates, and/or changes in the composition of crimes that occur within a stable policy environment. Presumably, a rational offender would only be deterred by changes in policy or practice that alter the risk of actually being put to death, and the expected time to execution conditional on being caught and convicted. It is not clear to us that the typical specification of the murder "cost function" as a linear function of these explanatory variables accurately gauges actual or perceived variation in such risks. This point takes on particular importance as variants in the basic specification lead to quite different results and we do not find strong theoretical reasoning justifying one specification over the other.

Aside from the specification of key explanatory variables, these studies face a number of econometric or statistical challenges that lie at the center of the dispute regarding the interpretation of the results. The first challenge concerns identification of a causal effect of the death penalty using observational data as opposed to data from a randomized experiment. Identifying the deterrent effect of capital punishment requires exogenous variation in the application of the death penalty, whether at the sentencing or execution phase. To the extent that discrete changes in policy are correlated with unobservable time-varying factors (for example, other changes in the sentencing regime or underlying crime trends) or are themselves being reverse caused by homicide trends, inference from regression analysis will be compromised. The extant body of evidence addresses identification by controlling for observables, employing time and place-specific fixed effects, and through the use of instrumental variables. The effectiveness of these identification strategies will be a key area of focus in this review.

The second modeling issue concerns drawing statistical inferences that are robust to non-spherical disturbances. We argue that there is strong consensus regarding the need to address this issue in panel data settings and, while there are many possible fixes, there are several standard approaches that guard against faulty inference.

The third modeling issue concerns the robustness of the results to variants in model specification. On this point, we consider issues with the underlying distribution of the key measures of deterrence that lead to unstable estimation and discuss the implications for interpreting deterrence findings as average effects.

In our assessment, we find the panel data evidence offered regarding a deterrent effect of capital punishment to be inconclusive. There is a weak connection between theory and model specification and subsequent research has demonstrated that results are sensitive to changes in specification. There are many threats to the internal validity of the empirical specification presented in these papers, and the identification strategies proposed to deal with them are based on questionable exclusion restrictions. Perhaps the easiest issue to address is the calculation of standard errors that are robust to within-state serial correlation and cross-state heterosckedasticity. Most, though not all of the papers finding deterrent effects do not adequately address this issue. Moreover, it has been demonstrated that the finding of statistically significant deterrent effects in several of the papers disappear when estimator variance is calculated correctly. Finally, in many cases, the results reported in this literature are driven entirely by a small and highly selected group of states or, in some cases, by a small number of state-years. This suggests that it is inappropriate to interpret model results as national average effects and casts additional doubt on whether the estimated effects can be attributed to capital punishment.

On the whole, we find the current research literature that uses panel data to test for a deterrent effect of capital punishment to be uninformative for a policy or judicial audience. This is to a small degree due to minor but important issues of appropriate model estimation and to a large degree due to unconvincing identification of causal effects and unconvincing theoretical justification for the model specifications employed. Given the difficulty in specifying a theoretically correct relationship between execution risk and murder rates, we argue that future research at the state level should focus on a discrete change in policy (for example, an execution moratorium or the abolition or reinstatement of the death penalty) and employ recent econometric advances in counterfactual estimation and robust inference. Absent an identifiable policy change, it is nearly impossible to assess whether variation in executions is driven by crime trends, overall shifts in composition of a state's judiciary, or changes in de facto policy regarding capital cases and hence whether the variation is relevant to future execution risk. We are of the opinion that additional studies based on state panel data models are unlikely to shed much light on this question as such policy changes are too rare to provide robust evidence.

## Model Specification

Most panel data models relating murder rates to death penalty policy measures are motivated by a straightforward theoretical paradigm. The "rational offender" model as described in Becker (1968) and Ehrlich (1975) posits that offenders or potential offenders weigh the expected costs and benefits when deciding whether or not to commit a crime, and that the likelihood of offending will depend inversely on the expected value of the costs. Stated simply, to the extent that the death penalty increases the "costs" of committing capital murder, capital punishment may reduce murder rates through general deterrence. An alternative theoretical approach places weight on the demonstration value of a relatively rare event (that is to say an execution). Specifically, to the extent that potential offenders are not particularly good at forming accurate expectations regarding the likelihood of being executed, the occasional demonstration that executions occur may

reinforce the notion that a completed death sentence is a real possibility, and in the process, perhaps deter murder. In such a framework, the details of the executions themselves (how long it takes, whether it is botched or the extent of media coverage) may also impact behavior. Thus there are two components of this theory. The rational offender model posits that potential murderers weigh the 'costs' and 'benefits' of committing homicide. The second posits that individuals place weight on specific signals generated by the sanctioning regime, such as an actual execution or a change from the historic execution rate, that provide salient proxies for the likelihood of a the sanction.

Much of the empirical panel data research is informed by the rational offender approach and estimates a model of the form

$$murder_{it} = \alpha_i + \beta_t + \gamma f(Z_{it}) + \delta X_{it} + \varepsilon_{it}$$

where $murder_{it}$ is the number of murders per 100,000 residents in state $i$ in year $t$, $f(Z_{it})$ is an expected cost function of committing a capital homicide that depends on the vector of policy and personal preference variables $Z_{it}$ with corresponding parameter $\gamma$, $X_{it}$ is a vector of control variables with the corresponding parameter vector $\delta$, $\varepsilon_{it}$ is a mean-zero disturbance term, $\alpha_i$ is a state-specific fixed effect and $\beta_t$ is a year-specific fixed effect.

There are several features of this model specification that bear mention. First, the inclusion of state-level fixed effects indicates that the estimated impact on state murder rates of the expected costs associated with committing murder is estimated using only variation in murder rates and expected costs that occur within states over time (i.e., between-state variation in average costs and average murder rates do not contribute to the estimate). Restricting to state variation over time ensures that the models do not attribute changes in the murder rate to the death penalty when they are, in fact, due to state level differences in other dimensions. We note that restricting to state variation over time has considerable consequences for the degree of variation in the capital punishment variables. Second, the inclusion of year fixed effects also removes year-to-year variation in the data that are common to all states included in the panel, ensuring that the models do not attribute changes in the murder rate to the death penalty when they are, in fact, driven by national-level shocks.

Finally, we have defined a general relationship between the expected costs of committing murder and the vector of determinants of costs given in Z via the function $f(.)$ and, for simplicity, a linear relationship between murder rates and the expected cost function. Of course, expected costs of committing capital murder may impact the murder rate non-linearly.

To operationalize this empirical strategy one needs to explicitly choose the functional form of $f(.)$ and articulate the set of covariates to be included in the vector of cost determinants. Functional form issues aside, expected utility theory suggests a number of candidate control variables that could be included in Z. First, the perceived likelihood of being apprehended and convicted should positively contribute to the cost of committing murder. Moreover, the perceived probability distribution pertaining to the adjudication of one's case should distinguish between the likelihood of being sentenced to death and the likelihood of alternative sanctions. Beyond these probabilities, the effect of a specific sentencing outcome on an individual's per-period utility flow as well as the effect on the expected remaining length of one's life should certainly matter. Finally, an individual's time preferences—i.e., the extent to which the potential offender discounts the future—will impact expected costs.

The effect of capital punishment on this cost function can occur through several channels. First, to the extent that receiving a death sentence shortens one's expected time until death, the presence of a death penalty statute will increase the expected costs of capital murder. In addition, death row inmates serve time under different conditions than those sentenced to life without parole. The physical and social isolation of death row inmates, as well as having to exist under a cloud of uncertainty regarding whether and when one will be executed, likely diminishes welfare while incarcerated and thus increases costs. As the death penalty is only a small part of the sanction regime for murder, the incremental effect of the death penalty on the cost of being convicted will also depend on the alternative sanctions should one not be sentenced to death. For example, life without parole in a designated high security facility may be more welfare diminishing than a sentence of 20 years to life with no mandatory-minimum security placement. Hence, a complete specification of the cost function would require specification of the discounted present value of one's future welfare stream under a death sentence, the comparable present value under the most likely alternative non-capital sanctions, and some assessment of the conditional probability of each. In this specification stage, additional assumptions are often made that on average, perceived risk of these sanction costs will equate to actual risk. While complications of estimating these actual risks are discussed in detail below, we note here that there is no research on the sanction risk perceptions of potential murderers on which to base these assumptions, and, by definition, there is no mechanism for learning over time through experience with the death penalty to enforce such an assumption.

The existing body of panel data studies often specify the function $f(Z_{it})$ using either an indicator variable for the presence of a death-penalty statute or as a linearly-additive combination of (1) a gauge of murder arrest rates, (2) a gauge of capital convictions relative to murder arrests, and (3) a variable measuring executions in the state-year relative to convictions, either measured contemporaneously or lagged 6 or 7 years.[3] The first specification effectively punts on the questions regarding how the death penalty specifically impacts the expected costs of committing murder. In the second specification, measures (2) and (3) are intended to measure within-state variation over time in the application of the death penalty.

Prior to evaluating specific studies, we believe that there are three primary issues that require some general discussion. First, to what extent do the two general approaches outlined above correctly specify the expected cost function $f(Z_{it})$? Second, is it possible to identify exogenous variation occurring within state over time in the elements of $Z_{it}$ that pertain to the death penalty? Finally, what steps should be taken to ensure that statistical inferences are robust to non-spherical disturbance issues that are common in panel data applications?

Regarding the issue of specifying the cost function, the two specifications that we have identified clearly employ incomplete and indirect gauges of the effect of capital punishment on expected costs. The first specification (whether or not the state has a death penalty statute) makes no effort to characterize variation across or within states over time in the zeal with which the death penalty is applied. Moreover, there is relatively little variation in this variable as few states have switched death penalty regimes since the key 1976 Supreme Court decision. On the plus side, if one were able to identify exogenous variation in the presence of a death penalty statute through either instrumental variables or

---

[3] Other variations also exist, for example, one set of studies (Mocan and Gittings 2003, 2010) includes measures of the likelihood of being removed from death row through a commutation or having a sentence overturned and another (Zimmerman 2006) explores the impact of the type of execution method employed.

appropriate control-function methods, the interpretation of the coefficient is quite straightforward. In particular, assuming identification of a causal effect is achieved, the coefficient on an indicator variable for the presence of a death penalty statute provides an estimate of the (local) average treatment effect of having such a law on the books that operates through unknown mechanisms. While ultimately mediated through a black box with unknown determinants and an unspecified functional form, one can clearly connect policy to outcomes in such a research design.

We find the second approach (controlling for murder arrest rates, capital convictions relative to murders, and executions relative to lagged convictions) problematic due to the incomplete specification of avenues through which the death penalty impacts relative expected costs and the arbitrary specification choices commonly made. To be fair, one could situate each of these measures within the expected cost function that we have specified. Specifically, the murder arrest rate provides a rough measure of the likelihood of capture, while the capital conviction rate provides a gauge of the likelihood of receiving a death sentence. Both of these variables may change over time with changes in de facto enforcement and sentencing policy. However, within-state variation over time in both of these variables is also likely to reflect changes in the volume or nature of crime in addition to changes in policy with regards to resource allocation towards murder investigation and the willingness to pursue capital convictions. Short term changes in factors such as the volume or nature of crime are unlikely to be predictive of the probability of arrest or capital conviction in future periods and hence would not be appropriate measures of expected risk. Given that policy is only one determinant of these outcomes and especially the fact that much of the state variation over time in these variables occurs within states with stable de jure death penalty regimes, it becomes difficult to interpret the meaning of the regression coefficients.

The most problematic variable in this specification is the commonly included measure of executions relative to lagged death sentences, meant to convey information about the probability of execution conditional on receiving a death sentence. In several studies, executions are measured relative to capital convictions lagged 6 or 7 years usually justified by the statement that the typical time to execution among those executed is 6 years. Perhaps the thinking here is that if, on average, it takes 6 years to be executed then the rational offenders should be monitoring this ratio in deciding whether or not to commit murder. We find this specification choice to be puzzling and arbitrary for a number of reasons.

First, basing the lag length of the denominator of this variable on the average time-to-death for those who are actually executed ignores the fact that only 15% of those sentenced to death since 1977 have actually been executed. If one wanted to normalize by the expected value of time to death for those sentenced to death, one would need to account for the 85% of death row inmates with right-censored survival times. As of December 31, 2009 the average death row inmate has been held under sentence of death for approximately 13 years (Snell 2010). Taking these censored spells for death row inmates into account along with the fact that for the 40% of inmates who have been removed from death row for reasons other than execution life expectancy is closer to that of inmates sentenced to life suggests that the time-to-death of the typical inmate with a capital sentence is much greater than 6 years.

Second, the commonly stated stylized fact used to justify the lag-length choice (that the average time to execution is 6 years among the executed) is incorrect for most years analyzed in the typical panel data analysis. Snell (2010, Table 12) presents estimates of the average elapsed time between sentencing and execution for each year from 1984 to 2009.

For the 21 people executed in 1984, the average elapsed time was approximately 6.2 years. Since then, this statistic has increased dramatically. For example, the average for the 52 individuals executed in 2009 stood at approximately 14.1 years. The lowest value for this statistic since 2000 occurred in 2002 (10.6 years).

Third, aside from being incorrect for the United States as a whole, the time-to-execution lag commonly employed most certainly varies across states that have a death penalty statute. As the statistics cited in the introduction to the paper suggest, California has sentenced many inmates to death (approximately 1,000), but has executed relatively few of these inmates (roughly 1.4%). Texas on the other hand, has sentenced slightly more than 1,000 inmates to death and since 1977 and has executed roughly 43% of these inmates. Clearly, the expected time to death for an inmate on California's death row differs from the comparable figure for Texas. To the extent that the same lag is utilized for each state, the coefficient on the execution rate variable will be measured with error that is most likely to be non-classical. In addition, as the models presume that annually updated and state specific changes in murder arrest rates, death penalty conviction rates, and number of executions are relevant for deterrence, it seems inconsistent to assume that a fixed national average lag from early in the time period is the relevant lag.

Perhaps the most important problem concerns the tenuous theoretical connection between this measure of "execution risk" and the expected perceived cost of committing murder. If we define the variable T as the duration between the date of the death sentence and the actual date of death with cumulative distribution function G(T), the execution risk variable as commonly measured seeks to estimate G(7)–G(6). Putting aside the fact that this discrete change in the cumulative density function is measured using synthetic rather than actual cohorts (none of the papers measure actual executions of those particular offenders sentenced to death 6 years ago and our best guess is that with longitudinal data this variable would equal zero in most state-years), there is no theoretical reason to believe that a rational offender considering committing capital murder would pay particular attention to this specific point in the cumulative distribution function. On the other hand, summary measures of the distribution of T, such as E(T) or var(T), would have a stronger theoretical justification for inclusion in the cost function.

Finally, it is not immediately obvious that variation in this variable reflects contemporaneous changes in policy that a rational offender should properly take into account when deciding whether or not to commit capital murder. As with the murder arrest rate and the capital conviction rate, one can certainly envision year-to-year variation in the execution risk variable occurring within a perfectly stable policy environment insofar as the underlying societal determinants of crime vary from year to year. In such an environment, the fact that there are more executions this year relative to last year does not necessarily convey information regarding a change in the expected costs of murder. This problem is exacerbated by the fact that potential murderers would be interested in the potential costs of murder as they would be manifest several or many years into the future, when they would potentially be prosecuted for the murder they are considering committing. It is particularly unclear whether year to year variation in executions or execution rates are relevant to the risk of execution 5, 10, or 15 years in the future.

Besides measurement and functional form issues, these common specifications of the murder cost function are certainly incomplete since none of the studies that we review attempt to articulate, measure, and control for the time-varying severity of the considerably more common alternative sanctions for convicted murderers who do not receive a death sentence. This may be a particularly important shortcoming of this research to the extent that sentencing severity positively co-varies with capital sentencing practices. This brings

us to the more general issue of whether changes in current execution rates or capital conviction rates are either correlated with contemporaneous changes in other sanctions that influence crime rates or are correlated with past changes in sentencing policy that exert current effects on crime.[4] For example, a current increase in executions may reflect a toughening of sentencing policy across the board that exerts a general deterrence and incapacitation effect on murder independently of any general deterrent effect of capital punishment. Similarly, given the long lag between conviction and execution in the United States, a spate of executions today may reflect past increases in enforcement that, in addition to increasing capital convictions, increased convictions of all sorts. Such past enforcement efforts may have long lasting impacts on murder rates to the extent that those apprehended and convicted in the past for other crimes are still incarcerated and are candidate capital offenders. It is also important to note that such factors are likely to be time varying within state and hence not eliminated by the inclusion of year and state fixed effects. With this in mind, the list of covariates as well as the instrumental variables employed to identify causal effects deserve particular attention.

The final issue that is a particular point of contention in this literature pertains to the correct measurement of standard errors. One of the key papers in this debate uses county-level panel data to estimate the effects of various gauges of capital punishment policy measures at the state-year level. The higher level of aggregation in they key explanatory variable suggests that the standard errors tabulated under conventional OLS assumptions will be substantially biased downwards to the extent that there is intra-county correlation in the error terms within state.[5] A second critique levied at several of the other papers in this literature concerns adjusting standard errors for potential serial correlation within cross sectional units and heteroskedasticity across units. Bertrand et al. (2004) demonstrate the tendency for unadjusted panel data difference-in-difference estimators to substantially over-reject the null hypothesis in the presence of serial correlation within cross-sectional clusters. This problem seems to persist both when tabulating standard errors under the assumption of spherical disturbances as well as when employing parametric corrections for serial correlation in the residuals.

Several recent textbooks devoted to the practice of applied econometrics place great emphasis on the importance of taking measures to guard against incorrect inference due to the misspecification of the variance–covariance matrix of the error terms in panel regression models. For example, in their discussion of serial correlation in microdata panels, Cameron and Trivedi (2005) note "the importance of correcting standard errors for serial correlation in errors at the individual level cannot be overemphasized" (p. 708). Likewise, in their treatise on econometric practice, Angrist and Pischke (2009) advocate for calculating conventional standard errors and cluster robust standard errors, and making inference based on the larger of the two. In reviewing the studies below, we will return to these prescriptions.

---

[4] The first possibility would violate the assumption that the regressors in the model above are contemporaneously exogenous—i.e., that the $E[\varepsilon_{it}|Z_{it}, X_{it}] = 0$—while the second possibility would violate the assumption of strict exogeneity—i.e., $E[\varepsilon_{it}|Z_{i1}, Z_{i2}, ..., Z_{iT}, X_{i1}, X_{i2}, ..., X_{iT}] = 0$.

[5] Specifically, if we let $C_i$ be the number of counties in state i define $\rho$ as the intra-state correlation in errors across counties within a given state, Moulton (1986) shows that the ratio of the true variance of the OLS estimator $\hat{\beta}$ to the variance formula assuming spherical disturbances is equal to $1 + [\text{var}(C_i)/\bar{C} + \bar{C} - 1]\rho$ where $\bar{C}$ is the average number of counties per state. Here we have adjusted the "Moulton factor" as described in Angrist and Pischke (2009) for perfect intra-class correlation in the key explanatory variable.

## Existing Panel Data Studies

There are several alternative frameworks that one could use to organize a review of panel data research studies on the death penalty. For example, we could group studies based on methodological approach (OLS, IV), particular specifications of the expected cost function (execution risk measures, overall execution levels, competing risks for those sentenced to death including commutation or being murdered while incarcerated), or general findings (either supporting or failing to support deterrent effects). Alternatively, one could segment the small set of panel data studies by loosely defined research teams. Our reading of this literature is that there are several "research groups" that have contributed multiple papers to the extant evidence and that one can discuss each team's output to illustrate the evolution of how underlying models have been specified. Moreover, there have been several major challenges to the findings of these teams and written responses to the challenges.

In what follows, we organize our review by research group. We begin with the research from teams that draw the strongest conclusions regarding the deterrent effects of capital punishment and then proceed to researchers whose findings and interpretation of the findings are more tentative. Throughout, we interweave discussion of the challenges raised for each paper. In the following section, we discuss published critiques of this research and the responses of these research teams to the published criticisms of their work.

### Dezhbakhsh, Rubin and Shepherd

A series of papers written by various combinations of Hashem Dezhbakhsh, Joanna Shepherd, and Paul Rubin are commonly offered as providing the strongest evidence of a deterrent effect of the death penalty. There are two initial papers (Dezhbakhsh and Shepherd 2006; Dezbakhsh et al. 2003) which we will discuss in this section, followed by a response to critiques of these papers published in 2010 that will be discussed after the critiques are summarized later in this paper.

Dezhbakhsh and Shepherd (2006) exploit variation in state-level death penalty statutes, some of which is driven by the 1972 and 1976 Supreme Court decisions. The authors conduct a number of tests for deterrent effects. First, the study presents an analysis of national time series for the period 1960–2000 and for the sub-periods 1960–1976 as well as 1977–2000. Second, the authors calculate the percentage change in murder rates pre-post the introduction of a death-penalty moratorium by state as well as pre-post percent changes in murder rates surrounding the lifting of death-penalty moratorium and analyze whether the distribution of these changes (mean, median, proportion of states with a specific sign) differ when moratoria are imposed relative to when moratoria are lifted. Finally, the authors estimate state-level panel regressions for the period 1960–2000 where the dependent variable is the state murder rate and the key explanatory variables are either the number of executions in the state or the number of lagged executions, as well as an indicator variable indicating a state moratorium.

To summarize their findings, national level time series regression analysis demonstrates significant negative correlations between the national murder rate and the number of executions (both contemporaneous and lagged one period) as well as a positive association with years when there was a national death-penalty moratorium (1972–1976). With respect to the pre-post moratorium analysis, the authors find an increase in murder rates following the introduction of a moratorium in 33 of the 45 states. They also find that in state years surrounding the lifting of a death-penalty moratorium, murder rates, on average, decline. Finally, the state-level panel regressions find significant positive effects of the death

penalty moratoria on homicide rates and negative coefficients on the number of executions and the number of executions lagged.

While the authors interpret this collage of empirical results as strong evidence of a deterrent effect of the death penalty, there are several problems with each component of the empirical analysis that we believe precludes such a conclusion. We distinguish here between two categories of problems, with the first and more important being plausible identification of an effect of capital punishment and the second being appropriate specification of the uncertainty of the estimates. Beginning with the national time series evidence, the main challenge to identification of an effect of capital punishment is the possible confounding effect of national trends in murder that are unrelated to capital punishment. On this point, we note that the stated positive effect of the national moratoria is driven by relatively high murder rates in the years between the 1972 *Furman* and 1976 *Gregg* decisions. A visual inspection of the national murder rate time series reveals that the rate began to climb in the early 1960s, was already at a historic high by 1970 and increased a bit further in the years surrounding the Furman decision. The high murder rate in the aftermath of the national moratorium clearly reflects the end-stages of a secular increase in the U.S. murder rate with a starting year for this transition that long predates 1972. This pre-existing trend provides sufficient challenge to the identification of an effect of the moratorium.

Aside from the analysis of the effects of the national moratorium, the authors also draw conclusions from the apparent negative relationship between murder rates and execution totals. However, there is reason to believe that the national level murder rate time series is non-stationary, calling into question any inference that one might draw from a simple national-level regression. In Table 1, we present results from various specifications of the augmented Dickey-Fuller unit root test using the two alternative data series for national murder rates that are used in this literature: the series constructed from vital statistics and the series based on FBI Uniform Crime Reporting data. In all specifications, we cannot reject the null that the homicide rate follows a random walk. As is well known, hypothesis tests arising from regressions with non-stationary dependent variables are prone to over-rejecting the null hypothesis.[6] Hence, we place little weight on these national-level results.

Turning to the analysis of pre-post moratoria changes in state level homicide rates, the strongest evidence of a deterrent effect comes from the increases in homicide associated with the introduction of a moratorium. However, we believe that this pattern is less supportive of a deterrent effect than the authors argue due to identification concerns. Most of the state-years where a moratorium is introduced correspond in time to the 1972 *Furman* decision. As we have already noted, the national time series clearly reveals homicide rates that are trending upward. One plausible interpretation of this evidence is that the authors have documented the continuation of an existing trend in homicide rates occurring both at the state and national level. Indeed, this point is made quite forcefully in Donohue and Wolfers (2005). In a reanalysis of these data, the authors first reproduce the state-level distribution of the percent change in homicide associated with moratoria reported in Dezhbakhsh and Shepherd (2006). However, they go one step further and show that contemporary changes in murder rates in states not experiencing a policy change parallel those that do. Specifically, in their reanalysis, the authors construct comparison distributions for the percentage change in state murder rates for states that do not experience the

---

[6] The authors report results from their own augmented Dickey-Fuller test and note that their test also fails to reject the null of a unit root. They argue, however, that the *P* value of 0.16 is "marginal" and … Given that these unit root test are notorious for low power (inability to reject an incorrect unit root null), a low *P* value makes the unit root problem an unlikely scenario here (p. 518 footnote 9).

**Table 1** *P* values from augmented Dickey-Fuller tests for non-stationarity in the national level homicide rate time series for 1960–2000

| | Homicide rate based on vital statistics data | Homicide rate based on FBI data |
|---|---|---|
| *Panel A: No trend* | | |
| # of included lags | | |
| Zero | 0.502 | 0.709 |
| One | 0.210 | 0.199 |
| Two | 0.179 | 0.168 |
| Three | 0.198 | 0.182 |
| Four | 0.158 | 0.231 |
| Five | 0.182 | 0.146 |
| *Panel B: Allowing for a trend* | | |
| # of included lags | | |
| Zero | 0.994 | 0.996 |
| One | 0.887 | 0.867 |
| Two | 0.871 | 0.877 |
| Three | 0.934 | 0.912 |
| Four | 0.908 | 0.930 |
| Five | 0.898 | 0.873 |

Figures in the table are approximate *P* values for tests of the null hypothesis of a unit root

introduction of a moratorium. Donohue and Wolfers find changes in murder rates for these "control" states that are nearly identical to those introducing a moratorium. They also find similar results when they reanalyze the effects of death–penalty reinstatements. Notably, Dezhbakhsh and Shepherd (2006) make no effort to construct a comparison group that would permit some assessment of the counterfactual distribution of homicide changes under no moratorium.

Finally, the authors estimate panel data regression models, weighted by state population, in which the murder rate is specified as a function of the number of executions (both contemporaneous and once lagged), an indicator variable that captures whether a state has a death penalty regime, and a host of other covariates.[7] These panel regression estimates suffer from a number of methodological problems that again limit the strength of this evidence. First, the authors intimate that their panel data regressions adjust for year fixed effects, yet in replicating their results Donohue and Wolfers discovered that the authors have instead included decade fixed effects. The result is that national-level murder shocks within the 1970s, the decade in which the national murder rate peaked, are uncontrolled in the regression. Second, there are concerns regarding whether the standard errors in their panel data regression are correctly estimated. The authors note adjusting for possible "clustering effects" but do not indicate the manner in which they cluster the standard errors. Donohue and Wolfers find that clustering on the state to address correlation over time within state yields standard errors that are nearly three times the size of those reported in an earlier working paper draft by Dezbakhsh et al. (2003).[8]

---

[7] In addition to state and time fixed effects, the covariates included in the models are: per capita real income, unemployment rate, police employment, percent minority, percent 15–19 year old, and percent 20–24 year old.

[8] Dezbakhsh and Rubin (2010) reply to Donohue and Wolfers' criticism regarding their failure to cluster standard errors at the state level by noting that the standard correction is only asymptotically valid and that, with fifty clusters, the traditional standard errors might be preferred. However, Bertrand et al. (2003)

A final issue concerns the manner in which the authors specify the expected cost function that we discuss in the previous section. The authors enter the count of executions on the right hand side of the regression model. This certainly is a poor choice if the intention is to gauge the relative frequency with which convicted murderers are put to death in a specific state. One possible justification for such a specification would be if executions have a "demonstration effect" whereby the execution or news coverage of it temporarily deters homicidal activity (and this temporary deterrence is not offset by a later increase, and the timing of executions within the year doesn't confound the estimates using an annual unit of analysis). An implication of this however is that a single execution in a very large state would deter more murders than an execution in a small state. Functionally, such specification would also require that potential offenders in large states draw the same inference that potential offenders in small states would draw from a single execution regarding the likelihood that they will be punished similarly. This seems like a particularly restrictive functional form to impose absent a strong theoretical reason for doing so.[9] Moreover, experimentation with various normalizations of the number of executions in Donohue and Wolfers leads to drastically different results. The specification of the cost function in this paper is also clearly incomplete as it does not include any information on state-year changes in other much more frequent sanctions for murder such as the likelihood of life without parole or the average length of prison sentences. This incomplete and restrictive specification of the cost function raises strong concerns about identification and related model uncertainty. In conjunction, we are of the opinion that this cluster of concerns limits what one can infer from the results of this particular analysis.

Given the apparent sensitivity of the results to small changes in specification, we performed further exploration of the underlying data used in this paper with an eye towards identifying the source of this model specification sensitivity. In particular, we explored whether there are issues related to the distribution of the data underlying the model that could explain the swings in coefficients of interest under changes in specification. Figure 1 presents a concise summary of this analysis. The figure presents a partial regression plot of residual annual state murder rates against residual annual number of state executions after accounting for state fixed effects, year fixed effects, and a commonly-employed set of control variables. As can be seen, there is a large degree of variation in annual state homicide rates remaining (although more than 80% of the variation in this outcome can be attributed to state fixed effects)—this is the variable displayed on the vertical axis of the graph. However, there is little remaining variation in executions, with almost all data points clustered closely around zero on the horizontal axis. The coefficient estimated by Dezhbakhsh and Shepherd (2006) is, roughly speaking, the simple linear regression coefficient between these two sets of residuals.

Figure 1 demonstrates that this coefficient estimate is heavily influenced by the rare data points that fall outside this data cloud. In fact, the simple lowess (dashed line) of the data suggests no effect in the bulk of the data and a decreasing trend in the sparse data regions outside the data cloud that is determined entirely by the handful of points designated by triangles. These data points are all observations from Texas, with the particularly influential data point in the lower right being Texas in 1997. With the removal of this

---

Footnote 8 continued

establish that clustered standard errors work quite well with 50 clusters, generating $P$ values on tests that are remarkably accurate—on the order of 6% with a $P$ value of 0.05.

[9] In their 2007 reply, Dezhbakhsh and Rubin assert that the effects of a single execution may vary considerably among states with different sized populations.
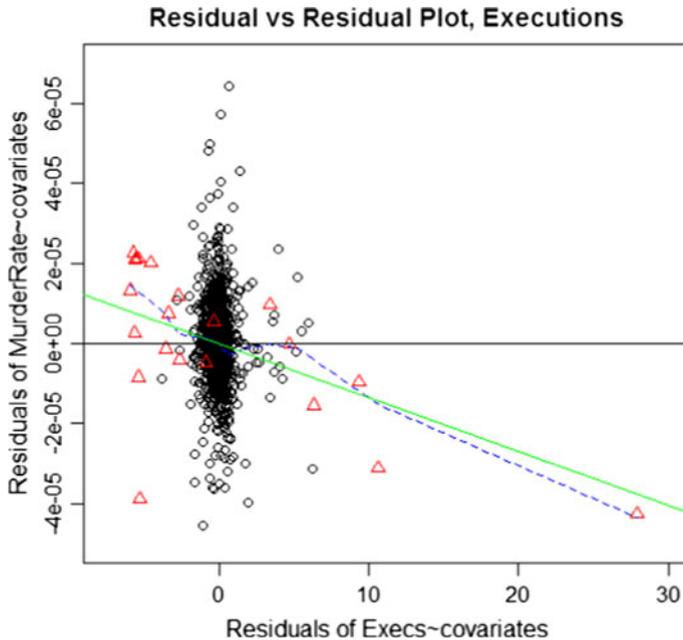
**Fig. 1** Illustration of Influential Data Points

single data point the slope remains negative albeit with much reduced statistical significance (when no adjustment is made to standard errors in parallel to the original analysis). The removal of all Texas data points yields a slope that is very close to zero and clearly non-significant.

Thus, while the panel model suggests that the execution effect estimated is an average effect over many states and years, it in fact is entirely determined by a small handful of data points all from one state. This carries two implications for the main findings in the paper under review. First, the estimated effect of executions is not an average effect but a Texas effect; it generalizes to other states only by assumption, not by empirical evidence. Second, this reanalysis raises additional concerns about identification in this case—were there other issues that may have been affecting homicide rates in Texas, particularly in 1997? It also carries implications more broadly for model sensitivity. With the vast majority of the data in a vertical cloud with zero slope, the coefficient of interest will be determined by whatever small set of data points fall outside of this cloud—and which data points have this characteristic is likely to vary across specifications leading to wildly differing estimates. In contrast, any set of models that keeps the set of points outside the data cloud relatively fixed will appear to be robust. This exploration suggests that through use of a linear specification with little variation in the explanatory variable of interest, except for a handful of outliers, many of the models employed in this literature may be over-fitting the data.

The second oft-cited paper by this research team is that of Dezhbakhsh et al. (2003).[10] This study analyzes a county-level panel data set and provides a more careful specification

---

[10] Dezhbakhsh et al. (2003), "Does Capital Punishment Have a Deterrent Effect? New Evidence From Post-Moratorium Data," *American Law and Economics Review*, 5: 344–376.

of the expected cost of committing murder. Specifically, the authors assume that the expected cost function is an additive linear function of the risk of arrest, the risk of receiving a death sentence conditional on arrest, and a measure described as the risk of being executed conditional on receiving a death sentence. The measure of execution risk is operationalized as the number of executions in the current year normalized by the number of death penalty sentences 6 years prior. This specification is copied with some variation in several other papers in this literature. The key innovation of the paper is to estimate the model using two-stage least squares where state-level police payroll, judicial expenditures, Republican vote shares in presidential elections, and prison admissions are used as instruments for the three variables that capture the expected costs of committing murder. The authors identify their key variable of interest as the effect of the state execution risk measure on county-level homicide rates.

While there are a number of issues regarding the manner in which the authors characterize the execution risk (a factor that we discussed at length in "Model Specification"), there are several fundamental problems with this analysis that limit the need for further discussion. First, some researchers have raised concerns with the quality of county-level crime data from the FBI's Uniform Crime Reports. In particular, Maltz and Targonski (2002) find that these data have major gaps and make use of imputation algorithms that are inadequate, inconsistent and prone to either double counting or undercounting of crimes depending upon the jurisdiction.[11]

Second, the dependent variable in this paper is measured at the county level while the key explanatory variable (execution risk) varies by state-year. At first blush, this might appear to be an advantage with over 3,000 counties in the United States and only 50 cross-sectional units in the typical state-level panel. However, given the unit of variation of the key explanatory variable, the authors have fewer degrees of freedom than is implied by the number of counties multiplied by the number of years. The authors make no attempt to adjust their standard errors to reflect the higher level of aggregation of their death penalty variables despite the well-known result that the OLS standard errors are, in this case, downward biased (Moulton 1990).

Finally, a more fundamental problem concerns their choice of instruments used to identify the effect of execution risk on murder rates. These instruments are police expenditures, judicial expenditures, Republican vote share, and prison admissions.[12] For these instruments to be valid, they must impact murder rates only indirectly; to be specific, only through their effects on the mean risk of being executed. Certainly, policing, judicial expenditures, and prison admissions (either directly or through their correlation with lagged values of prison admissions) may impact murder rates through many alternative channels (incapacitation of the criminally active, deterrence, greater social order, a smoother functioning court system etc). Prison admissions, through incapacitating criminally active people, will impact crime directly beyond the indirect effects operating through murder arrest rates, conviction rates, and execution rates.

The multitude of alternative paths through which the instruments may be impacting homicide and crime in general render the results of this analysis highly suspect. As this

---

[11] Maltz and Targonski (2002) further document the fact that reporting errors in these data are not only inconsistent, they are highly non-random, with certain jurisdictions and certain states reporting more consistently than others. The authors provide a detailed analysis of these data and conclude that "… county-level crime date cannot be used with any degree of confidence."

[12] As Donohue and Wolfers (2005) note, the results are highly sensitive to the operationalization of the Republican vote share in the first stage regression in a manner that again suggests to us that overfitting may be of particular concern.

entire analysis rests on the validity of the instrumental variables, again we find this analysis to be thoroughly unconvincing.[13] Donohue and Wolfers (2005) suggest a falsification test to check the validity of the exclusion restriction underlying the analysis of Dezhbakhsh, Rubin, and Shepherd. Specifically, they restrict the sample to include only those states that did not have an operational death penalty statute. They re-specify the models employed by Dezhbakhsh, Rubin, and Shepherd to estimate "the effect of exogenously executing prisoners in states that have no death penalty" and find significant deterrence effects in five of six specifications. As the authors note, the most obvious interpretation of these data is that the instruments have a direct effect on the homicide rate independent of its effect through changes in the number of executions.

Zimmerman

Paul R. Zimmerman has authored three articles that reach mixed conclusions regarding the deterrent effect of capital punishment. The first two articles are research papers on different but related questions surrounding capital punishment while the third article was written in response to the critique by Donohue and Wolfers (2005) of the earlier research papers. In this section, we discuss the first two papers.

Zimmerman (2004) estimates state panel data regressions covering the period 1977–1998 with a specification of the expected cost function that is somewhat similar to that in Dezhbakhsh et al. (2003). Specifically, Zimmerman models state murder rates as a function of the murder arrest rate (murder arrests divided by total murders), the likelihood of receiving a death sentence conditional on being arrested for murder (death sentences divided by murder arrests), and a measure described as the probability of execution conditional on conviction. A departure from Dezhbakhsh et al. (2003) is that instead of normalizing executions by death sentences lagged 6 years, Zimmerman defines the execution rate as the number of contemporaneous executions divided by the number of contemporaneous death sentences. Zimmerman then estimates model specifications with the three cost function variables measured in the same time period as the murder rate and models where the three cost functions variables are lagged one period. The models employ a fairly standard set of socio-demographic and crime-related, time-varying covariates[14] and are weighted by state population.

The author presents two sets of results: estimation results employing ordinary least squares and 2SLS results specifying instruments for each of the three key explanatory variables. With three endogenous variables, the author needs at least three instruments, and specifies a set of eight, resulting in an overidentified model. The instruments include (1) the proportion of a state's murders in which the assailant and victim are strangers, measured contemporaneously and once-lagged, (2) the proportion of a state's murders that are non-felony, measured contemporaneously and once-lagged, (3) the proportion of murders by non-white offenders, measured contemporaneously and once-lagged, (4) a one-period lag of an indicator variable that captures whether there were any releases from death row due

---

[13] Invalid instruments are especially problematic when those instruments are only weakly predictive of the endogenous covariates. However, contrary to current standard practice, the authors do not report results from their first stage regressions.

[14] In addition to state and year fixed effects and state specific linear time trends, the covariates in the model are: per-capita prisoners, per-capita police, per-capita robberies, per-capita assaults, percent unemployed, per-capita income, proportion in metro areas, in poverty, Black, and in age groups 18–24, 25–44, 45–64, 65+.

to a vacated sentence, and (5) a one-period lag of an indicator variable that captures whether there was a botched execution.

According to the text of this article and the written specifications of the first-stage regressions, the proportion of murders that are committed by strangers and the lag of that variable are used as an instrument for murder arrest only. The variables measuring the proportion non-felony and the proportion committed by non-white offenders are used as instruments for death sentences per murder arrest, while the lagged indicator variables for releases from death row and botched executions are used as instruments for the execution rate. However, the printed equations indicate that the murder rate (the dependent variable) is included as an explanatory variable in each of the first-stage models. We presumed this to be a typographical error and decided to seek additional information from the underlying computer code to see the exact specification of the first stage.[15] Indeed the dependent variable was not entered on the right hand side of the first-stage regressions. However, in contrast to the text, all eight instrumental variables are included in each first-stage model.

Both the OLS and 2SLS regressions include complete sets of state and year fixed effects as well as state-specific linear time trends. No effort is made to adjust the standard errors for possible serial correlation of the error term within state, although in a follow-up paper that we will discuss later, Zimmerman (2009) re-estimates these models with parametric adjustments for auto-correlation.

The OLS regression yields no evidence of an impact of the execution risk on murder rates though the murder arrest rate is consistently statistically significant with a sizeable negative coefficient. The 2SLS results yield statistically significant and much larger negative effects of the murder arrest rate and a significant negative coefficient on the execution risk variable. Zimmerman also experiments with a log–log specification for the first stage model and reports finding no evidence of a significant effect of the execution risk or arrest risk variables using this specification.

The 2SLS results supportive of a deterrent effect suffer from many of the same problems that have been raised in response to the work of Dezhbakhsh et al. (2003). We identify five such problems and then discuss each in turn. First, the key execution risk variable is constructed in a manner that is difficult to interpret and does not clearly relate to the actual risk of execution for an individual convicted at a given point in time or to the expected amount of time until execution. However, we note, although the author does not, that as little is known about the perceptions of sanction risk by potential murderers, risk estimates based upon recent information on death sentences may be just as plausible as the other specifications presented. Second, a key coefficient of interest in one of the first stage regressions, the effect of a botched execution on execution rates in the following year, is of opposite sign to what Zimmerman hypothesizes that it should be, yet this is reported nowhere in the published paper. Likewise, the first stage relationship (in a prior working paper version of the paper but not reported in the final version of the paper contrary to current standard practice) between the endogenous regressors and the instruments is very weak and does not pass standard econometric tests of instrument relevance. Third, mechanical relationships between each of the types of murder (felony versus non-felony, murders committed by white versus non-white offenders) potentially threaten the validity of the instruments in these models. Finally, the significance of the 2SLS results is not robust to standard adjustments for auto-correlated disturbances.

---

[15] Justin Wolfers who re-estimated the results in this paper provided us with Zimmerman's data and corresponding Stata code.

Beginning with the key explanatory variable, it is not clear how a change in the ratio of executions to new death sentences would impact the expected cost of committing capital murder. If one were to multiply this ratio by the inverse of the proportion of the state population on death row, then the ratio could be interpreted as the ratio of the transition probability off death row though execution to the transition probability onto death row from the general population. However, even this ratio wouldn't have a meaningful interpretation. Absent a clear articulation of the theoretical reasoning that would justify measuring the execution risk in this manner, it is difficult to interpret what the 2SLS result is telling us.

Regarding the instruments, an inspection of the actual first stage results reveals several patterns that would lead one to question the exclusion restrictions that the author is imposing. In the first-stage model for the execution risk variable, the instrument with the largest $t$ statistic (over 3.5 in both 2SLS models estimated in the paper) is the lagged variable indicating a botched execution. The lag of the variable indicating commutations or vacated death sentences is never significant. The author states that the botched execution indicator is one of the key variables identifying execution risk, and argues that a botched execution last period should reduce future execution risk either through judicial action or reluctance on the part of key policy makers to risk another botched execution. To the contrary, the variable measuring botched execution exerts a significant *positive* effect on the execution risk. Given that the observed effect is opposite to the author's initial hypothesis, the underlying quasi-experiment driving the key finding in this paper is certainly not what the author suggests.

A related problem is that of instrument relevance. As Zimmerman notes, in order for the instruments employed to be valid they must be both excludable from the structural equation and relevant. In this context, relevance refers to the explanatory power of the instruments in the first stage regression. The $F$-statistics on the excluded instruments reported in an earlier version of the paper were very small (never greater than 5). With three endogenous regressors and eight excluded instruments, the corresponding Stock-Yogo critical value for this $F$ test is 15.18 (Stock and Yogo 2005). Thus, the instruments employed in this analysis should be viewed as weak. In 2SLS estimation, two issues arise in the presence of weak instruments. First, in the event that the instruments are even weakly correlated with the errors in the structural equation, 2SLS estimators perform especially poorly since the bias of 2SLS is scaled by the inverse of the explanatory power of the instrument in the first stage regression (Bound et al. 1995).[16] Second, even if the exclusion restriction is satisfied, weak instruments typically lead to over-rejection of hypothesis tests in the second stage regression.

A third problem with the 2SLS models in Zimmerman (2004) involves the validity of the remainder of the instruments. The remainder of the instruments can be described as the proportion or share of murders of different types: those committed by strangers, those committed by non-whites, and non-felony murders. There are two concerns with this set of instruments. The first is that if some of these categories of murder are more variable than others, then the share of murders of that type will be directly correlated with the total homicide rate. For instance, if the rate of murders (to population) committed by strangers and non-whites is fairly stable but the non-felony murder rate varies considerably over time, then variations in the homicide rate will largely be due to variations in the non-felony

---

[16] Though Zimmerman (2009) drops two particularly problematic instruments from his original analysis and presents tests of over identifying restrictions that suggest instrument exogeneity, such tests are not able to provide an explicit test of the exclusion restriction.

murder rate and increases in the share of murders that are non-felony will be directly correlated with increases in the murder rate. The second concern is that variation in the sanction risk probabilities due to shifts in the share of each of these types of murder doesn't affect an individual potential offender's risk of sanctions. An individual potential offender either is a stranger to their potential victim or not, either is non-white or not, and either potentially commits a felony or nonfelony murder. The sanction risk associated with each of these states is not varying, just the proportion of murders of these types, and thus not the sanction risk for an individual offender. Consequently, these variables only operate as successful instruments if potential offenders adjust their perception of sanction risks in response to the portion of variation in total sanction probability that is due to changes in the composition of murders but without knowing that this is the source of those variations. If the source of the variations is known to potential offenders then it should not change their perception of their sanction risk.

A final concern raised by Donohue and Wolfers (2005) and readdressed in Zimmerman (2009) pertains to the robustness of the statistical inference. In his later analysis of the same data, Zimmerman (2009) tests for autocorrelation in the murder rate error terms within states and finds evidence, as one might expect that the errors are not independent of one another. As we discussed above, auto-correlated errors in panel data models tend to bias standard error estimates tabulated under the assumption of homoscedastic independent and identically distributed error terms towards zero. In replicating Zimmerman's results, Donohue and Wolfers (2005) find that clustering the standard errors in the 2SLS model by state leads to an appreciable increase in standard errors and resulting insignificant coefficient on the execution risk variable.[17]

As with the prior sets of papers, we find the evidence for deterrence from this paper to be uninformative. Again, this is due in small part to model estimation issues such as adjustment of standard errors and in large part to substantial identification challenges arising from unconvincing instruments and problematic specification of capital punishment variables.

Zimmerman (2006) uses a similar framework as the OLS specification implemented in Zimmerman (2004) but replaces the execution risk variable with a vector of variables giving the risk of being executed by each execution method. These are specified with two lag structures, the numerator is the number of executions by a specific method carried out in the current year with either current or once-lagged denominators of the number of death penalty sentences. Further specifications vary the denominator, using all murders instead of those sentenced to the death penalty and filling in missing values of the deterrence variables as described in Zimmerman (2004; missing values are due to zero death sentences in the relevant year for the denominator of the execution risk variables). The once-lagged murder rate is included as a covariate in the models to account for autocorrelation in the error term over time within state. The author reports that only executions by electrocution are associated with reductions in the homicide rate.

We note that it is consistent with Zimmerman (2004) that the most common method of execution, lethal injection, is not found to have a significant relationship with homicide when using the same concurrent lag specification. The rarity of the other execution methods, including the one found to have a significant deterrent effect in these models, adds to concerns about identification beyond those given in the discussion of the more

---

[17] Zimmerman (2009) does not contest this point, noting that t-statistics on the key explanatory variable fall from 2.24–0.74 to 2.94–0.81 in the two primary specifications employed in the paper.

general model with the same specification used in the 2006 paper and would lead to an exaggeration of the issues demonstrated in Fig. 1.

## Mocan and Gittings

There are two papers by Mocan and Gittings (2003, 2010) that estimate a more extensive variant of the model specified in Dezhbakhsh et al. (2003) but without instruments. Specifically, the authors estimate state panel data regressions where the dependent variable is the state murder rate and the key independent variables are the murder arrest rates, the death penalty conviction rates, and measures described as execution risk and the risk of being removed from death row for reasons other than execution. There are several key departures from the papers that we have reviewed thus far. First, Mocan and Gittings cluster the standard errors by state in all models, addressing a key weakness that pervades much of the extant literature. Second, this team makes an effort to specify some other components of the expected cost function that may confound variation in the execution risk. Third, the authors add a more extensive set of time varying control variables to the specification.[18] Finally, the authors estimate associations not only of executions but also of commutations and removals from death row with homicide rates.

The specification of the key cost-function variables follows a somewhat complex schema. The murder arrest rate is measured as the ratio of the number of murder arrests to total number of murders in a given year (with the ratio lagged one period in all specifications in order to minimize endogeneity bias). The ratio of prisoners per violent crime is also included and is meant to further control for the arrest and conviction risk. To capture changes in the expected cost of a murder that arise under a capital punishment regime the authors employ three additional deterrence variables. First the authors include a variable measuring the ratio of death sentences in year $t - 1$ to murder arrests in year $t - 3$. The 2-year lag is justified by the average observed empirical lag between arrest and sentencing in capital cases. Second, the authors include the ratio of the number of executions in $t - 1$ to death sentences in year $t - 7$.[19] They justify this lag length based on the statement that the average time to execution among the executed is 6 years. Third, the authors include the ratio of the number of death sentence commutations in year $t - 1$ to death sentences in year $t - 6$. In some specifications, the authors substitute overall removals from death row through avenues other than execution for commutation in the numerator of this variable. The shorter time lag for the removal risk (relative to the definition of the execution risk variable) is justified by the shorter average time to removal relative to time to execution. As in the prior papers reviewed here, no measures of the risks associated with other more common sanctions for murder are included.

Similar to the earlier papers, the authors estimate panel data regressions by weighted least squares where the observations are weighted by each state's share of the U.S.

---

[18] In addition to state and year fixed effects, the authors also include on the right hand side of their models standard covariates: the unemployment rate, real per capita income, the proportion of the state's population in each of several age groups, the proportion of the state's population that is black, the proportion of the state's population that reside in urban areas; and additional covariates: the infant mortality rate, the legal drinking age, the governor's party affiliation, and the number of prisoners per violent crime. Also included is a dummy equal to one in Oklahoma in 1995 in order to capture the impact of the Oklahoma City bombing.

[19] Mocan and Gittings re-weight the data such that executions performed late in a given calendar year receive more weight in the subsequent calendar year and vice versa. They do not state what this implies about the mechanism they assume to be operating between their execution intensity measure and a potential offenders cost function.

population. The authors find significant negative associations between the execution risk variable and murder rates, and significant positive associations of the removal risk and murder rates. In summarizing the numerical magnitude of the results, the authors report that each additional execution is associated with five fewer homicides, that each additional commutation is associated with five more homicides, and that each additional removal from death row is associated with one additional homicide. In regression models that also include an indicator variable for the existence of a capital punishment statute, they find that a death penalty law has an independent negative effect on murder, above and beyond the effect of an execution.[20] Finally, as a robustness check, the authors demonstrate that the execution, commutation and removal rates are not associated with the subsequent crime rates for robbery, burglary, assault and motor vehicle theft which they interpret as evidence that these rates may be considered exogenous in their effects on crime rates for homicide.

Mocan and Gittings (2003) are consistently careful in their model estimation and present clear descriptions of their regression models. That said, the lack of any measures of the risk of other sanctions for murder is a glaring omission that threatens the identification of effects of the death penalty. In addition, we have raised a number of issues with the specification of the execution risk in this manner in "Model Specification" that generate additional substantial concerns regarding the appropriate interpretation of the related coefficient even if one were willing to assume that variation in this measure were exogenous. As Donohue and Wolfers (2005) note, perhaps the most obvious critique of the models employed by Mocan and Gittings concerns the complex temporal structure of the explanatory variables of interest—the probability of an execution, a commutation and a removal are estimated as the ratio of executions to death sentences issued five or 6 years earlier. The authors argue that these ratios are the most accurate measures of risk and state their belief that potential murderers will be not just rational in their decision making but also accurate in their risk perceptions and hence propose that these measures are good proxies for the risk perceptions of potential murderers. The authors state that the choice of 6 years is made for two reasons: (1) because the average time from a death sentence until execution is reported by Bedau (1997) to be 6 years for prisoners who are actually executed and (2) in order to maintain consistency with prior research by Dezhbakhsh, Rubin and Shepherd.

In response, Donohue and Wolfers suggest that a better measure of the deterrence probability would use a 1-year lag under the supposition of Zimmerman (2004) that offenders are likely to utilize the most recent information available to inform their behavior. We find no greater theoretical justification for a one year lag than a six year lag given the lack of information on risk perceptions of potential murderers but note that the results are sensitive to the lag employed. When the model presented in Mocan and Gittings (2003) is re-specified using 1 year lags, the coefficients on executions and removals become insignificant.[21]

Mocan and Gittings (2010) provide a multi-faceted and detailed defense of their choice of lags, arguing that the once lagged versions of the deterrence variables suggested by Zimmerman are uninterpretable and have no meaning since individuals sentenced to death are almost never executed within 1 year. Among the robustness checks included in this follow-up paper, the authors employ alternative normalizations of executions using either 4 or 5 year lags of death sentences and find that doing so does not alter their conclusions.

---

[20] The coefficient on the indicator variable for a death penalty regime suggests that the existence of a capital punishment statute is associated with approximately 64 fewer murders.

[21] Using once lagged data, the models are estimated on data that span 1978–1997.

While we will discuss their 2010 response in greater detail in "Responses to These Challenges", here we discuss this particular set of robustness checks.

Given that the average time to execution for those receiving a death sentence is clearly closer to 6 years than to 1 year, there is some theoretical justification for employing a longer lag, although we again note that it is unknown how potential murderers construct their perceptions of sanction risks. However, the choice of a 6-year lag remains problematic for several reasons. First, we have pointed out that the chosen lag length to execution is based on a statistic calculated on a select sample of death row inmates—i.e., those 15% who have actually been executed. This grossly under-estimates the average time to death for those who will eventually be executed, or released, given the large number of inmates sitting on death row for a decade or more. Second, as noted earlier, the 6-year figure is an incorrect measure of the national average of time until execution for those executed in the time period studied. According to the figures presented in Snell (2010), the raw average of time to execution for executions occurring between 1984 and 2009 is 10.06 years. Calculating this average using the number of executions in each year as weights yields the higher value of 10.86 years. The comparable figures calculated over the sub-period 1984–1997 (corresponding roughly to the time period analyzed by Mocan and Gittings) are 8.6 and 9.4 years respectively. We replicated the results of Mocan & Gittings, utilizing 7, 8, 9 and 10 year lags of the sentencing variable to normalize state execution totals. Using 7 lags, the coefficient on executions remains negative and significant ($P < 0.02$). However, for 8, 9 and 10 year lags, the coefficient is not significantly different from zero.[22] Having raised this concern over the lag length, both theoretically and empirically, and the incomplete specification of the cost function more generally, we reserve further discussion of the degree to which the Mocan and Gittings results are informative to the critiques and response to the critiques that follow.

Katz, Levitt, and Shustorovich

In an analysis of state-level panel data for the period 1950–1990, Katz et al. (2003) estimate the impacts of overall prison death rates and deaths by execution on murder rates, violent crime rates, and property crime rates. The authors argue from the outset that it is difficult to believe that the additional risk through execution during the post-1977 period is sufficient to have a measurable deterrent effect on a group of generally highly present-oriented people. They make this argument based on the relative rarity of an execution, the small fraction of convicted murders who are punished by execution, and the relatively high mortality rate through other causes faced by those who are perhaps the most likely to commit criminal homicide in the United States.

Their principal model estimates involve fixed effect panel regressions of crime rates on inmate deaths for all causes normalized by the state prison population and executions also normalized by the prison population; the latter being a key difference in specification

---

[22] Given the theoretical ambiguity underlying the choice of a lag structure, a natural alternative that is more flexible and agnostic with regard to functional form is to include each of the relevant lags of the execution rate in the same regression model and test the joint significance of these variables in explaining changes in the murder rate. We re-estimate (2) first including execution variables that include one through six lags and next including just the execution variables in which the denominator is lagged by 4, 5 and 6 years. In the first model using all six lags, the $F$ statistic on a joint test of significance of each of the incarnations of the lagged execution rate variable was 1.09 ($P = 0.38$). Using just four, five and six lags, the $F$-statistic was 2.62 ($P = 0.06$), suggesting the presence of a relationship when attention is restricted to this set of lags, albeit one that is not significant at conventional levels.

relative to the papers we have reviewed thus far. The models include fixed effects for states and years. All models are estimated by weighted least squares with state population share as weights and the authors tabulate standard errors clustered at the state-decade level.

The authors find fairly stable parameter estimates for the prison death rate's association with homicide but unstable, sometimes positive, sometimes negative estimates of the execution rate's association with homicide. The evidence with regards to violent crime, however, yields more consistent results. The effect of prison death rates is negative and statistically significant in most models, while the execution rate is unstable across specification. Similar findings are observed for property crime. Finally, the authors explore model results where they allow for several lags of the prison death rate and prison execution rate. The results here parallel the findings from the models without additional lags.

Donohue and Wolfers (2005) reproduce the original results in Katz, Levitt and Shustorovich and provide additional point estimates with the data extended from 1950 to 2000 and an even larger data set covering the period from 1934 to 2000. Their reproduction of the original model specifications similarly shows execution rate effects that are highly unstable across specification (with negative significant effects of the execution rate in three of the eight model specifications estimated and insignificant effects in the remainder). When the data are extended through 2000, none of the execution effects are significant while model estimates on the longer time period from 1934 to 2000 yield several coefficients indicative of a *positive* significant effect of the execution rate on murder rates.

The specification of the execution risk in this paper has been criticized by Mocan and Gittings (2010) as not accurately reflecting the risk of execution for an individual who is considering committing murder. Certainly, as the overwhelming majority of inmates in prison are not on death row, an execution risk normalized by the overall prison population does not provide a gauge of the execution risk faced by the average prison inmate. Again we note that it is unknown how potential murderers perceive execution risks. We will defer a more complete discussion of alternative normalizations until later in this paper.

## Challenges to Initial Studies

There have been at least three major published challenges to some or all of the body of research reviewed above. All three reanalyze the data on which these studies based their findings, and each makes different yet related points. First, Donohue and Wolfers (2005) provide an extensive review and reanalysis of every paper that we have discussed above. Second, Berk (2005) reanalyzes the data utilized by Mocan and Gittings (2003) and makes a fundamental point about the extreme sensitivity of the estimated relationship between homicides and execution to literally a handful of observations (in this context, specific state years). Finally, Kovandzic et al. (2009) re-estimate many of the models in the papers cited above—those estimated without instruments—with an expanded time series and with a richer set of covariates. In addition, they provide a critique of the rational offender model of criminal offending, which motivates the search for a capital punishment deterrent effect, based on qualitative and quantitative research by criminologists on the determinants of criminal behavior. Finally, we discuss a fourth paper by Fagan et al (2006) which critiques the identification strategies of the earlier papers and proposes a different identification strategy. In this section, we review each of these papers in turn.

Donohue and Wolfers

At the heart of the review conducted by Donohue and Wolfers is an extensive reanalysis of the data sets and estimators presented in the extant body of recent panel data papers. We have already discussed many of the findings from this review article as they pertain to the individual studies discussed above when we considered them among the most salient critiques of those studies. Hence, we will not repeat those specifics here. Instead, here we extract some of the general points made in Donohue and Wolfers as they raise objections that apply to several of the papers in this body of research.

First, Donohue and Wolfers emphasize a basic tool in identifying a treatment effect—that of attempting to construct counterfactual comparison paths, in this case for states that experience some form of "treatment" such as the abolition or reinstatement of the death penalty. For instance, the authors demonstrate that the relatively high U.S. homicide rates during the 4-year death penalty moratorium between the 1972 *Furman* and 1976 *Greggs* decision were also observed in Canada where there was no such contemporaneous variation in sentencing policy. In fact, they demonstrate that while homicide rate levels in Canada are lower than they are in the U.S., the overall temporal trends in homicide rates are quite similar between the two countries, suggesting that one should exercise extreme caution before drawing inferences regarding general deterrence based on movements in either univariate time series. A similar point is made in their reproduction of the state-level analysis in Dezhbakhsh and Shepherd (2006), where the authors analyze pre-post changes in homicide associated with death penalty moratoria or reinstatement. Donohue and Wolfers demonstrate that contemporaneous changes in states with no such policy variation are extremely similar in both sign and magnitude to changes observed among states with a change in death sentence policy. This finding calls into question causal inferences that are predicated on pre-post analyses exclusively within states that change policy regimes. Indeed an alternative conclusion that would rationalize this pattern in the data is that the observed relationship between homicide and executions is merely an artifact of a secular increase in violent crimes that occurs prior to and during the 1970s.

Second, echoing a point made by Katz, Levitt and Shustorovich, Donohue and Wolfers emphasize the relative rarity of an execution and the low level of variance across years in any explanatory variable constructed from state-year level execution series relative to the variation observed in state-level homicide rates. The precision of the estimated coefficient on the execution risk variable, however measured, depends directly on the degree of residual variation in the variable after accounting for state and year fixed effects and whatever other control variables are included in the model. The great sensitivity in deterrent effects to changes in specification suggests that model uncertainty exacerbated by the outlier prone distribution of execution measures dominates sampling variability in this setting. This concern is the focus of the work discussed below by Berk.

Third, most of the published papers finding evidence of a general deterrent effect do not adequately account for potential serial correlation within states in the error terms of the regression model, with research by Mocan and Gittings being the key exception. The downward bias to the OLS standard error formula in the presence of positive serial correlation is well known, and the implications for inference (in particular, the impact on the tendency to over-reject the null) have been well established (Bertrand et al. 2004). One of the key contributions of the Donohue and Wolfers reanalysis is to demonstrate how many of the significant results become insignificant when standard errors are clustered at the state level. In fact, Donohue and Wolfers (2005) effectively ended the practice of failing to adjust OLS standard errors for serial correlation that was previously a mainstay in this

literature. However, debate has since ensued about the proper methodology for making such adjustments, in particular about adding models of the error structure based upon additional parametric assumptions versus 'cluster robust' standard error adjustment based on sandwich estimators. Our take on this issue, as noted earlier, is that in the technical literature (i.e., recent standard econometric textbooks) there is clear and well justified recommended practice on this issue, and hence it is inaccurate to frame this is an active debate.

Fourth, Donohue and Wolfers make the general point that the execution risk, as specified in this collection of papers, often has little or no theoretical justification and that changes in the manner in which this risk is constructed yield very large differences in research findings. For example, normalizing executions by the state-level population, by the number of prison inmates, or by lagged homicides, all yield insignificant results despite a significant negative coefficient on execution levels.[23] Similarly, normalizing executions by contemporary death sentences yields a statistically insignificant coefficient while normalizing by death sentences lagged 6 years yields a significant deterrent effect. All of these results exist in published papers (for example, a comparison of OLS results in Mocan and Gittings (2003) and Zimmerman (2004) demonstrate the sensitivity to which year of death sentences is employed in the denominator). However, Donohue and Wolfers' juxtaposition of these findings in one place illustrated a fragility of this relationship over a range of model specifications. In the face of this fragility, there would need to be a quite strong theoretical and/or empirical basis for selecting one specification over another. Notably when an $F$ test is employed as a more theoretically agnostic test of the joint significance of a relevant subset of lags of the execution variable, the results are nearly always insignificant.

Finally, in a detailed discussion of the two papers that employ instrumental variables, the authors illustrate the tenuousness of the assumed exclusion restrictions and the counter-intuitive implications of the estimated first stage results. Of note, the first-stage results are not reported in the two papers with conclusions that rest heavily on the validity of questionable exclusion restrictions.

We find the Donohue and Wolfers (2005) study to be useful in laying out the issues just described. Our two concerns with this paper are relatively minor. The first is stylistic and perhaps difficult to avoid given the breadth of their reanalysis. While this paper suggests many reasons why the literature it reviews may be less conclusive than it claims, it does not separate out the relative importance or nature of each concern raised. As a result, it does not clarify whether the problems they identify are of theory or of practice and hence the extent to which they can be remedied. Our second concern is technical and relates to the randomization test described on p. 833. The error in this randomization test is detailed in Kennedy (1995, p. 90) and relates to use of an inappropriate null hypothesis that all of the coefficients are zero instead of that just the coefficient of interest is zero. However, the correct randomization test for this model would also need to be adapted to address the two stage least squares estimation.

---

[23] In their response to Donohue and Wolfers, Mocan and Gittings (2010) deflate executions by the state population as well as lagged homicide rates but using a different lag for these denominators. Using a 6 year lag, the coefficient on the execution rate remains significant though the coefficients on commutations and removals are no longer significant.

Berk

Berk (2005) presents a particularly revealing analysis of state-level annual data on executions and homicides. The contribution of this work is a detailed assessment of the extent to which the underlying *data* used in the panel studies can support the existence of deterrence regardless of the validity of the identifying model assumptions. The available data may be regarded as the entire population or as a single sample from an underlying data generating mechanism or super-population with the crucial observation that it is the only sample that is observable. This brings particular attention to the attributes of these data for the purposes of estimating deterrence effects of executions. While Berk analyzes the panel data covering the period from 1977 to 1997 employed by Mocan and Gittings (2003), his points raise questions for all of the papers in this research area, as they are all to some extent analyzing the same data set. Berk focuses on the bivariate relationships between homicide (either measured as state-level totals or as homicides per 1,000 residents) and the raw number of executions lagged 1 year. There are 21 years of data in the panel analyzed with 1,050 state-year observations, and Berk focuses on the 1,000 observations for homicide from 1978 through 1997 matched to executions in 1977–1996.

Berk begins by analyzing the frequency distribution of both the homicide data and the execution data. While the homicide data exhibits strong evidence of skewness, this appears to be a particularly important problem for executions. In particular, 859 of the 1,000 state-year observations have zero executions. Of the remaining 141 observations, 78 have one execution per year while only 11 observations have five or more executions per year (eight of which are Texas state-years).[24] With both a mean and median value of zero (in fact an 85th percentile of zero) and a very small number of observations with even modest numbers of executions, Berk raises concerns regarding the possibility that this handful of observations may be exerting excessive influence on the findings from linear panel regression models.

To assess this possibility, Berk analyzes the bivariate relationships between the number of homicides and the number of executions as well as the relationship between the homicide rate and the number of executions. The results are presented graphically and employ a flexible functional form (estimated by cubic splines) along with estimated 95% confidence intervals. Simple bivariate results yield homicide levels that are increasing at a decreasing rate in the number of lagged executions and homicide rates that increase in the number of executions through the value of five and then decline slightly for high values.

To explore multivariate relationships, Berk presents similar plots of homicide totals or homicide rates by executions after netting out state fixed effects, year fixed effects, and both state and year fixed effects.[25] When state fixed effects are accounted for, a negative relationship between homicide and lagged execution counts emerges, however again only for execution totals beyond the value of five. For state-years with zero through five executions (roughly 99% of the sample) there is no relationship between homicide totals or homicide rates and the execution variable.

These latter results then lead to an analysis of the influence of the 11 state-year observations with more than five executions. Omitting these 11 observations from the

---

[24] This comes from our tabulations of Berk's Figure 14.

[25] We note that our empirical work shows that the state and time fixed effects dominate these models, accounting for 89% of the variation in the outcome, and the remaining covariates typically included account for less than 2% of the variation. The sanction risk variables typically employed account for less than 0.5% of the variation in homicide rates.

analysis yields flexibly-fit functions between the homicide measures and executions that suggest that there is no underlying relationship between the two variables. Most tellingly, including the 11 observations and fitting a linear function to the data yields a marginally significant negative relationship, driven entirely by these highly influential observations. In the authors own words, "In short, by forcing a linear relationship and capitalizing on a very few extreme values, a negative relationship can be made to materialize. But the linear form is inconsistent with the data, and the 11 extreme values highly atypical" (Berk 2005, page 319).

As we have already mentioned, many of these influential observations are state-years for Texas (8 of the 11). Based on this fact, Berk explores the sensitivity of the observed relationship between the homicide outcomes and execution counts to omitting Texas. Perhaps the most telling test is the following. Using the 950 observations omitting Texas, Berk randomly matched each state-year's homicide rate to the execution totals from a different state year and then estimates the relationship between homicide rates and executions for these 49 states. This test yields no evidence of a relationship, which is what one would expect since by construction the two series are unrelated. Berk then adds the Texas observations and re-estimates the relationship. Adding Texas to the shuffled data for the other 49 states yields a negative linear relationship between the homicide rate and the number of executions, when we know by construction that for 950 of the 1,000 observations there is no relationship between these two variables.

Berk does not explore the relationship between the homicide rate and the various transformations of the execution total employed in the panel data studies. However, he makes the general point that regardless of the denominator, the extreme skewing of the numerator in any constructed execution risk variable renders all such estimates extremely sensitive to outlier observations.

We find this analysis informative and, for one family of specifications, conclusive on the degree to which the data can inform the question of a deterrent effect. For models using execution counts, this presentation suggests that there is too little variation in the state annual execution counts to obtain meaningful average effects. We agree with Berk that it also raises pertinent questions for those employing specifications of execution risk with various denominators. First is whether similar patterns of sensitivity to the linearity assumption are due to the influence of a small set of outlying observations are underlying the model results. No matter what normalization is employed, 85% of the execution risk values will be zero (or missing due to denominators with a value of zero) as this is the case for the execution counts. The only question is how the other 15% of the values are distributed. Second is whether the researcher is comfortable with the vast majority of the variation in the execution risk, within the 15% of the data that is nonzero, coming from the normalizing denominator (and covariates). This puts the weight of the identification of a deterrence effect of execution risk on the denominator, the relevant at risk population, as opposed to the numerator, the number of executions. Where we disagree with Berk is in the author's statement that from this analysis one can conclude that the data cannot support analysis of a deterrence effect for any specification of execution risk. We also would have found a discussion of the homogeneous treatment assumption helpful in clarifying the impacts of the distribution of the available data.

Kovandzic, Vieritis and Paquette-Boots

Kovandzic et al. (2009) present a reanalysis of many of the models estimated in the papers that we review but are unique in using a longer panel of data that documents homicides and

executions through 2006. In addition, they employ an expanded set of control variables, paying particular attention to variables omitted in prior research that are known to be predictive of homicide. Regarding their specification of the execution risk, they explore a fairly large set of possibilities including an indicator variable for a death penalty statute, the execution count, executions per prisoner and executions per capita, per death sentence lagged 1 year and lagged 6 years, and per homicide committed. Their set of control variables are perhaps the most extensive of this literature and include the a typical set of demographic characteristics, plus controls for three-strikes laws, right-to-carry laws, prisoners per capita, prison death rates, police per capita, and in some specifications, a state-year level gauge of the severity of the crack-cocaine epidemic. These additional covariates are well supported in the criminology literature as predictive of crime in general or homicide in particular. In all models, standard errors are clustered at the state level and in most models observations are weighted by share of population. Notably, one difference in their specification of the death penalty risk variables from most of the rest of the literature is that they do not specify a measure of the risk of arrest given homicide nor of receiving a death sentence conditional on arrest for homicide. Instead, they include prisoners per population as an overall measure of arrest and conviction rates (per population instead of per crime). They also include prison death rate, deaths from all causes except execution, to capture the risk of death by execution relative to death from other causes given conviction. Neither of these alternate measures is specific to those committing homicide.

To summarize their empirical results, they find no evidence of a linear relationship between execution risk (defined in any of the ways listed above) and state homicide rates. They subject these findings to a number of specification choices, including restricting the data to the 1977–2000 period, not weighting the regressions by state population, calculating heteroscedastic-robust standard errors without clustering, controlling for the crack epidemic, and a number of other checks. Their finding of a null effect is robust across these specifications.

Interestingly, nearly all of the papers that we have reviewed have been written by economists (with the exception of the work by Berk, a statistician). All are motivated by an underlying rational choice theoretical paradigm whereby individuals weigh the perceived costs and benefits of specific actions and make decisions accordingly. Kovandzic et al. (2009) offer a discussion assessing whether such a model is a realistic description of the decision-making governing extremely violent acts, based on research by criminologists. This discussion raises a number of points worthy of additional theoretical and empirical development including

- many criminally active individuals are uninformed as to the consequences of their actions and the likelihood that they will be caught (they cite research suggesting that the criminally active have poorer assessment of these probabilities than the non-criminally active),
- many of those on death row have long histories of violence and many have prior felony incarcerations (i.e., these individuals do not seem to be the marginal offender sensitive to deterrence),
- many violent acts are committed by people who are under the influence of drugs or alcohol and/or in intense emotional states ('hot states') when one is the least inhibited and perhaps the most present oriented,
- many capital murders are committed during the commission of a separate felony, and hence capital punishment may not be at the forefront of one's thought when a homicidal act is carried out.

This rich theoretical discussion highlights the complicated nature of criminal decision making and also puts into sharp relief the simplistic cost-benefits approach that undergirds much economic research on the determinants of crime.

We find this analysis useful in mapping out a region of possible model specifications in which, with appropriate model estimation and inference, there is no evidence of a deterrence effect. While we have various concerns about the particular specifications employed here, as discussed throughout this review, it is nonetheless useful to have on hand this evidence of a region of the model space in which there are few or no significant effects. We found the one case in which the authors detail the specific factors that caused their model results to differ from those in another publication using a similar specification to be particularly clarifying. Our only specific critique is that it would have been useful to have similar analyses in other instances where their results using similar specifications differed from those presented elsewhere in the literature. In particular, we are interested in the impact of different specifications of the risk of arrest, conviction, and death due to different causes. We also note that similar to several other sets of authors, these authors also assume that their results are not biased by omitted variables, such as measures of the risks of other sanctions for murder, or by potential reverse causality. In this as in the other cases, we do not find this to be a credible assumption.

Fagan, Zimring, and Geller (2006)

Fagan et al. (2006) use similar modeling techniques but a substantively different strategy to identify a deterrent effect of capital punishment in comparison to the prior papers. The authors classify individual homicides (based on statutory information and homicide data in the *Supplementary Homicide Reports* portion of the Uniform Crime Reports) as capital eligible or not capital eligible. Their identification strategy is then to identify differential changes in capital eligible as opposed to non-capital eligible homicides in response to capital punishment. Fagan et al. assert that other studies "suffer from an important and avoidable aggregation error: they examine the relationship between death penalty variables and total non-negligent homicide rates, despite the fact that three-fourths of all such killings do not meet the statutory criteria to be eligible for the death penalty. … Since the risk of an execution is more than fifty times greater in the death penalty states for the "death eligible" cases, the variations in these cases but not the others should produce the distinctive fingerprints of death penalty policy deterrence…" (Fagan et al. 2006, p. 1859). The identification strategy rests on a hypothesis that potential murderers will know a priori whether the homicide they are considering or at risk of committing would be capital eligible or not. Further, the authors propose that in response to increased death penalty risk, some potential murderers will choose not to commit capital eligible murder while those considering non-capital eligible murder will not be affected. In addition, they propose that besides death penalty sanction risks, these two types of homicide will follow similar patterns over time and hence the non-capital homicide trends in the same states and years are what capital eligible homicides should be compared with.

The authors begin with qualitative comparisons of the raw rates of the two types of homicides over time (and the share of all homicides that are capital eligible) nationally, and pooled across all states with and all states without a death penalty statue (in each year from 1976 to 2003). From these figures they conclude that the patterns over time are similar across each of these geographic sets: that the rates of capital eligible homicides vary relatively little over time while the rates of non-capital eligible homicides are reduced by roughly 50% and consequently the share of homicides that are capital eligible increases

over time. Based on the authors' hypothesis that deterrence would cause different patterns over time in death penalty versus non-death penalty states in the share of homicides that are capital-eligible, they conclude that there is no evidence of deterrence from this comparison. Fagan, Zimring, and Geller make further qualitative conclusions of no deterrence based on the lack of a clear differential reduction in capital eligible homicide in states with a death penalty following years in which (nationally) there were a greater number of executions. This exercise is repeated for the state of Texas alone and Harris County in Texas as the authors state "if executions show a distinctive impact on death-eligible killings anywhere, Texas should be the place (p. 1813)." Descriptively, they report similar findings from these state and county specific data summaries.

These qualitative comparisons are followed by the results of two types of regression models again using national data from 1976 to 2003, and then data from counties in Texas from 1978 to 2001. The first type of regression is similar to the fixed effects regressions described at the outset of this paper where the unit of analysis is the state/year and the death penalty sanction risk variables are presence of a death penalty statute, the number of executions in each of the two prior years, and a moving average of the number of death sentences over the prior 3 years. In contrast to prior papers, the outcome is the rate of capital eligible homicides and one of the control variables is the rate of non-capital eligible homicides which operationalizes their identification strategy of testing for differential effects of capital sanctions on capital eligible homicides. Similar to other specifications, the models include state and individual year fixed effects, as well as a number of covariates.[26] The authors also include one other measure of the sanction regime which is defined as the log of the ratio of the state prison population to the number of felony crimes (lagged by 1 year and logged). Finally, they include the robbery rate in order to control for the potential supply of robbery-homicide incidents that comprise a significant portion of the capital eligible homicides.

The panel regression model results indicate no statistically significant effects of death penalty statues, executions, or death sentences on capital eligible homicides after controlling for non-capital eligible homicides which is consistently the strongest predictor. These results hold in the Texas only county/year fixed effect regression models. The results also hold in the second type of regression they run which implement state specific random intercepts and slopes in addition to year fixed effects. The death penalty measures are then included as main effects and interacted with a linear specification of year—allowing the death penalty measures to have different effects over time. Interactions with a linear specification of time are also included for the incarceration variable, the robbery rate variable, and all the controls.

We find this to be an interesting identification strategy and a sensible homicide classification method but have a number of concerns about the specification and interpretation of the models. While the identification strategy has some appeal it is also limited as even under ideal conditions, the lack of deterrence of this particular type does not tell us that capital punishment does not deter, just that it does not deter in this particular manner. In particular, the assumption that potential murderers know a priori whether the murder they are considering is capital eligible may not hold. We also question whether non-capital eligible homicides should be expected to follow similar patterns over time as capital eligible homicides and thus be useful by comparison as there are a number of attributes that

---

[26] Covariates are: poverty rate, percent in urban areas, percent Black, percent age 15–24, percent 65 or over, ratio of persons 35 and older to 14 and younger, Gini index of inequality, an indicator for the Oklahoma City bombing in 1995, and the logged population size.

differ about these two types of crime that could make them respond differently to a number of observed and unobserved conditions.

Our primary critique is that similar to the other papers reviewed here, the authors do not include measures of the rest of the sanction regime for murder (capital eligible or not) in their specification. As discussed earlier, the lack of a fuller specification of the sanction regime could bias estimates of deterrence in either direction and in the case of this paper could obscure any deterrence effect of capital sanctions that would otherwise be present. In addition, they use a particular specification of what they consider to be the salient aspects of capital sanctions—in particular information from only the last 2 or 3 years on executions and death sentences along with population based scaling factors and log specifications that are hotly debated in this literature.

In addition, there are a number of issues of clarity or modeling that are problematic. These include: inconsistent descriptions of the model and variable specifications, a description of a selection model for having a death penalty statute that suggests incorrect use of this methodology, apparent errors in reporting of either statistical significance or standard errors, incomplete information on the specification of the fixed effects models (were they weighted by state population share, how was the error term specified, etc.), incomplete information on the specification of the random effects models, inaccuracies in the interpretation of the random effects model results, incomplete information on the specification of the zero-inflation factor in the Poisson specification of the Texas specific random effects models, and insufficient information about how Figures 5 and 6 (Fagan et al. 2006, p. 1851, 1857) were created leading to seemingly inadequate accounting of uncertainty in these figures. We also note that the authors did not report the variance component results for the random effect models, thus keeping some of the potentially most informative results from readers. In short, while this paper has some interesting ideas we do not find the data summaries and qualitative conclusions compelling evidence and do not find the incompletely specified, unclear, and perhaps inaccurate regression results compelling evidence.

## Responses to These Challenges

Three of the research teams whose studies we review in "Existing Panel Data Studies" have published responses to the challenges to their research findings. All focus primarily on the critiques offered by Donohue and Wolfers (2005).[27] All of the responses were written prior to the publication of Kovandzic et al. (2009) and hence do not address their research findings or their critique of the rational offender model informing this body of research. Finally, Donohue and Wolfers (2009) provide an updated review of this research, raising many of the issues discussed in their earlier paper while incorporating reviews of more recent studies. The first response by Dezbakhsh and Rubin (2010) focuses solely on the reanalysis in Donohue and Wolfers (2005). The 2007 working paper version of their recently published paper intimates that Donohue and Wolfers engaged in a data mining exercise in pursuit of a specific but unspoken agenda. In the authors' own words:

> The data miners' abuses have been the subject of much criticism, leading to a higher scrutiny before publication. The doubt casters' mischief, however, needs more scrutiny, because the practice is just as damaging. Indeed the relative ease of

---

[27] One paper cites Berk (2005) in a footnote but does not address his critique.

obtaining irrelevant or invalid evidence of fragility, … 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 the model over and over till a different inference emerges is much easier than collecting your own data, building an econometric model, developing a statistics strategy, and making an inference (page 9 in working paper version.)

The authors go on to contest many of the issues raised by Donohue and Wolfers including (1) the contention that one should explore the results with and without Texas (an issue raised quite forcefully by Berk 2005), (2) whether one needs to adjust the standard errors to account for the possibility of serially-correlated error terms within state and, if so, what procedure should be employed, (3) how the execution risk should be characterized, and (4) whether decade fixed effects are sufficient in panel models rather than the standard (and far more granular) year fixed effects. The authors strongly imply that these issues are cherry picked by Donohue and Wolfers in order to highlight a select set of results that show little impact of deterrent. They then present 80 separate model estimates, some of which reproduce the basic results in Dezhbakhsh and Shepherd (2006), and some of which reproduce the models in Donohue and Wolfers (2005) and note that, in the majority of the models, the coefficients on the execution risk variables are statistically significant and negative. It should be mentioned that the overwhelming majority of these model specifications suffer from the issues raised by Donohue and Wolfers, in particular issues that we argue constitute errors in model estimation and inference by current econometric/statistical standards—e.g., inadequate control for time effects, standard errors clustered by year rather than state, etc.

We find their response lacking for several reasons. First, we find their characterization of the Donohue and Wolfers critique as "an abuse of sensitivity analysis" to be largely unfounded. The key issues that these latter authors raise are essentially the same standard questions raised every day in research seminars, in consultations between graduate students and their faculty advisers, and in peer reviewing of empirical research. The idea that adjusting for year fixed effects in a state panel data set or that adjusting standard errors for potential within-cluster autocorrelation are unusual contentions is, at best, inconsistent with current standards of empirical research and, at worst, absurd. Moreover, in light of the unusually large contribution of key states to variation in their execution risk measure (Texas in particular); questioning whether estimation results are sensitive to dropping such observations is standard and sound practice.

Second, the additional results presented in this paper basically reproduce what we already know from their original analysis and the reanalysis by Donohue and Woflers. In particular, when the models employ individual year fixed effects and standard errors are clustered at the state level, the linear coefficient on lagged executions is negative and significant in the models that include Texas (models 15–21) and insignificant in the models that do not (models 22–28). Figure 1 of this paper demonstrates why this is the case.

In the response by Zimmerman (2009), the author makes a more substantive effort to address the issues raised by Donohue and Wolfers and points out that there is less disagreement between Zimmerman (2004) and the general tenor of results in their review than one might expect. Zimmerman readily acknowledges that clustering standard errors by state yields insignificant results, that his OLS regression finds little evidence of a deterrent effect, and that the significant effect in 2SLS disappears when the model is re-specified in log–log form. Zimmerman does, however, contest the correction that Donohue and

Wolfers employ to adjust the standard errors for serial correlation: namely, calculating the standard errors using the empirical variance-covariance matrix for the error term estimated via 2SLS. Zimmerman argues that there are many possible parametric fixes for this problem that explicitly model the serial correlation within the panel (either constraining the autocorrelation parameters to being constant across state or allowing these parameters to vary across the cross-sectional units of the panel). Zimmerman employs a number of these methods and re-estimates the 2SLS models by feasible generalized least squares (FGLS). Employing a number of alternative parameterizations, Zimmerman finds that his results are still statistically significant and imply sizable deterrent effects of capital punishment.

To be sure, FGLS has been shown to be a lower variance estimator and thus generate efficiency gains relative to the clustered variance estimator. However, this is true only under the assumption that one has a consistent estimate of the variance matrix for the model error term. If this matrix is mis-specified, the standard errors from FGLS may continue to be biased downwards (as is demonstrated by the tendency for parametric solutions to still over-reject the null hypothesis in Bertrand et al. 2004). For this reason, recommended practice is to apply cluster robust standard error adjustment in addition to the parametric modeling of the error structure suggested by Zimmerman.

However, aside from the issue of the correct measure of variance for the estimator, we offer several reasons above to question the identification assumptions behind the 2SLS results in Zimmerman's initial analysis. First, the key explanatory variable identifying execution risk has a coefficient in the first stage regression with a sign that contradicts the hypothesized experiment that the author outlines. For the remaining instruments (proportion of homicides committed by nonwhites, the proportion of homicides committed by strangers), there is no clear theory justifying the exclusion restrictions, a problem that is compounded by the use of weak instruments. Second, the explanatory power of the excluded instruments in the first stage regressions is extremely low, rendering the resulting estimates subject to the multitude of critiques associated with 2SLS estimation using weak instruments. While the additional work presented by Zimmerman is a constructive and engaging response to critics, it does not change our assessment of the findings from the original analysis.

Mocan and Gittings (2010) provide a detailed response to Donohue and Wolfers (2005), presenting a number of additional empirical results using their original data set. As other sets of authors have done, Mocan and Gittings respond to the variable results found under different specifications by arguing that their preferred specification, which results in a significant deterrent effect associated with increased execution risk, is correct and that specifications with different results are, in various ways, inferior. In this paper they begin with the same data and their preferred specification from their 2003 paper, where execution risk is operationalized as recent number of executions (pro-rated throughout the year, as previously discussed) divided by the number of death penalty sentences 6 years prior. They then deviate from that specification along several dimensions—by shortening the lag in this denominator to 5 or 4 years, by varying the denominator within this lag range to include denominators suggested by other authors, and by removing, one at a time, a set of potentially high influence states. The variations they present often retain a negative and significant execution risk effect and more often retain a positive and significant effect of either commutation risk or risk of removal from death row.

With regard to omitting potentially high influence states, the authors re-run their primary specification first omitting high population death penalty states; Texas, next California and then both states simultaneously from the data. The coefficients remain mostly significant and in the expected directions. Mocan and Gittings also omit one-by-one from

the analysis Virginia, Arkansas and Louisiana, those states with the highest overall execution probabilities and find that the results are robust to these omissions. In our re-analysis of the data, we largely replicate these results. However, we note that omitting the three states with high execution probabilities simultaneously from the analysis yields a coefficient estimate on executions that is insignificant, regardless of the number of lags on the execution variable that are employed. This suggests that the strength of the deterrent effect of executions depends on the three states that are most likely to carry out executions, a finding which raises the possibility that the impact of execution risk may be non-linear or otherwise specific to these three states. Of course, it also is not surprising that statistical significance is lost when removing the most informative states from the model (where informativeness is based upon variation in execution risk). Thus, while the analysis of Mocan and Gittings is robust to the exclusion of Texas, assuming one accepts the modeling assumptions there remains doubt as to whether they have uncovered a "national effect" of execution.

In addition to the effects of executions, Mocan and Gittings also test the robustness of their findings with respect to the effect of commutations and removals from death row. Indeed, though this review paper is primarily concerned with evaluating the effect of the rate of *executions*, that the authors find significant positive coefficients on the commutation and removal variables in a variety of specifications is compelling. The authors conclude that potential offenders are responsive to the execution risk along two different margins. However, a detailed review of Mocan and Gittings' most recent results reveals that the estimated coefficients on commutations and removals are not as stable as their discussion suggests. In particular, the authors do not discuss that in the vast majority of the model specifications presented, the coefficient on *either* removals or commutations is positive and significant but it is rarely the case that both are significant.[28] We also find it notable that when the authors use a time-varying national estimate of the lag of the denominator of the commutation and removal risk variables (which is likely superior to the static lag in the sense that these are measured with less error), the coefficients on commutations and removals both become insignificant.

More generally, we have seen in other papers discussed above, that it is possible to find a large set of potentially reasonable model specifications that obtain similar results. A useful comparison here is to Kovandzic et al. (2009) who use many of the same execution risk variations, take a variety of stances on inclusion of covariates, and show overwhelmingly non-significant results with both positive and negative point estimates. The key dimensions on which these two papers differ in most specifications are the lag used, the length of the data series used, and the specification of the series of risk probabilities. Mocan and Gittings (2010) return to their 1976–1997 data, whereas Kovandzic et al. extend the data series up through 2006 (or 2000 in some specifications).

We offer two specific challenges to Mocan and Gittings' preferred specification on theoretical grounds, and highlight three sets of related empirical analyses, in order to suggest that Mocan and Gittings (2010) does not encompass the full set of reasonable models and that specifying this set more broadly leads again to inconsistent deterrent results. The key arguments in Mocan and Gittings (2010) for their specification concern the denominator in the execution risk measure and its lag. On the theoretical side, the goal is to generate a reasonable estimate of the state-specific, time varying, risk of execution as perceived by potential murderers. Mocan and Gittings suggest using what they pose as the

---

[28] It is generally the case that the coefficient on commutations tends to be significant when the four-year lag is employed while the coefficient on removals tends to be significant when five lags are employed.

actual risk of execution, since under rational expectations, perceived risk should be actual risk. While we note again that there is no mechanism that would align perceptions with reality over time (as say there is for business risks) we accept that this is a reasonable hypothesis to empirically test. It then follows that risk should be measured among those eligible to be executed, i.e., those who have received a death sentence. The specification then gets interesting in deciding which people who have received a death sentence to include. One reasonable option that we were glad to see here for the first time is to use all those on death row in the recent time period—this is quite literally those who are at risk for being executed. However, this is not the stated preferred specification because not all people on death row have the same probability of being executed in a given year. For example, those who enter death row in the current year have extremely low probability of being executed within that same year.

Hence, Mocan and Gittings prefer to use those people who received a death sentence 6 years prior, as they state this is the average number of years spent on death row for those who are executed. We discussed this specification at some length in "Model Specification" and concluded that we found it flawed. Very briefly, we find this problematic, as the risk is conditional on receiving a death sentence, not on being executed, suggesting that average time to execution should be estimated forward in time from a cohort of those receiving the death penalty instead of backward in time from a cohort being executed. As we have previously discussed, average time to execution among those sentenced to death is much longer than 6 years. Thus, our first general point is that if a single fixed lag length is going to be used, that longer lags, of say 8, 9, 10 years or longer to get closer to the average, would also be reasonable, perhaps more so.

Our second point is that it is inconsistent with the rest of the risk specification to use a single fixed lag length. While Mocan and Gittings acknowledge that national average lags (calculated using their method, conditional on execution) vary over time and they show results using year-specific national average lags, they do not address state-specific lags. This assumes that potential offenders make decisions using a stable estimate of the lag structure, obtained by averaging over states, or states and years, but that the rest of the execution risk components need to be updated each year with state-specific numerators and denominators. This combination of stability and immediacy in risk perceptions seems unlikely. In addition, for both the models using the constant 6-year lag as well as the models using the national average year-specific lags in death sentences to normalize executions, it is clear that such normalizations are infusing the execution risk variable with non-classical measurement error. To see this, consider California and Texas. In comparison to California, Texas executes those sentenced to death relatively swiftly. Applying the same lag length normalization to both states likely over-estimates the likelihood of being executed in 6 years in California and under-estimates the likelihood of being executed in 6 years in Texas.[29] Hence, the measurement error in this variable is positively correlated with the degree to which the state expedites executions. Certainly, states that swiftly carry out executions may differ from states that do not along a number of dimensions, including the ideological predisposition of state judges, the safeguards included in the appeals process, the overall harshness of criminal sentencing and so forth. Allowing the normalization to vary by year does not address this problem. Thus, we suggest that using state-year specific lags, and

---

[29] Of course, if one were to use actual executions out of specific sentencing cohorts (defined by year), this would not be a problem. Our understanding however, is that the papers that have employed these normalizations simply divide executions by lagged death sentences, with no assurances that those executed are actually from the cohort used to normalize the execution risk variable.

presumably one would want to do this for the lags of each of the death penalty related risk variables, would also be reasonable, perhaps more so. A related alternative is to run a joint significance test on a range of theoretically relevant lags of the denominator of the execution rate. Our re-analysis of the data discussed in "Existing Panel Data Studies" indicates that this method is unlikely to result in a significant deterrent effect.

Empirically, we offer three related sets of results. First, as we noted earlier, using longer lag lengths (8, 9, or 10 years) that are actually closer to the average time to execution for those eventually executed yields insignificant effects. To be fair, this difference may be driven by the fact that each additional year of lag results in one less year of data and hence lower power. However, the same point holds for why the 4, 5, and 6 year lag significant results may differ from the 1, 2, and 3 year lag insignificant results, as dropping years can either bring out or suppress significance (as demonstrated by Fig. 1). Nonetheless, given the available empirical evidence on the prevailing time until execution, we do not see any strong argument for favoring 4–6 year lag lengths and have established that conditional on using a single fixed lag, using our preferred longer lag leads to insignificant results.

Second, one of the model specifications implemented by Kovandzic et al, (2009) is very close to Mocan and Gittings preferred one. And, interestingly, under a number of variants of this model, they show consistently negative and quite insignificant execution risk coefficients. This is a stark, and potentially surprising, contrast to the consistently negative and significant results reported by Mocan and Gittings (2010). While we have not identified the exact reason for these divergent findings, the options appear to be that Kovandzic et al. employ a longer time series (although one of their specifications that is still insignificant goes only through 2000), a different set of conditional risk measures prior to execution risk, and they have utilized slightly different covariates. This establishes that a similar specification, except for minor variations in covariates, on a longer time series yields consistently insignificant results.

Third, we note here the results of a 2009 Ethan-Cole et al. paper that sought to reconcile the seemingly disparate results in Donohue and Wolfers (2005) and Dezhbakhsh et al. (2003) through Bayesian model averaging. The authors ran models covering the entire model space between the preferred specifications of the two papers and found that the posterior probability of a negative coefficient was about 70%. Thus many of the models they ran had negative coefficients, and a much smaller but substantial subset had negative and significant coefficients. On the other hand, the remainder of the models had positive coefficients, some of which were also statistically significant (under a frequentist framework). While we find the exclusion restrictions of the instrumental variables used in all those models to be untenable, our point here is that over other more reasonable portions of the model space, a wide variety of results may be found. And in addition, standard regression diagnostic tools, such as those presented in Fig. 1, may be useful to track down the sources of such model sensitivity. In our analyses, we have found significant negative coefficients to be reliant upon a small set of outlying observations—if model variants all roughly keep those same observations, results will be consistent and appear to be robust. If model variants affect those outlying observations, significant results (whether negative or positive) are lost.

Thus, while we find the carefully implemented presentation by Mocan and Gittings (2010) to be of great interest in that it maps out the results in one somewhat reasonable segment of the model space, our and Kovandzic et al.'s starkly different results in largely the same or what we find to be more reasonable parts of the model space lead us to regard their conclusions with considerable skepticism. Given that one is willing to make the assumptions required for the coefficient estimates to measure a causal effect, we do not

dispute that they have identified a portion of the model space in which results are largely supportive of a deterrent effect of increased execution risk. Instead, we point out that this is not the whole story and that other similarly reasonable or perhaps more reasonable portions of the model space offer no support for a deterrent effect of increased execution risk. On the assumption required for a causal interpretation, we find the lack of inclusion of measures of the risk of more common sanctions for homicide alone to be sufficient to deem the exogeneity assumption to be not credible. This sensitivity to model specification and challenges to constructing measures of death penalty related sanction risks combined with the lack of credibility for making causal attributions in this context lead us to conclude that the Mocan and Gittings (2010) results are not conclusive.

## Conclusion

Our review of the panel data studies testing for a deterrent effect of the death penalty on state level murder rates can be summarized with three key points. First, we believe that the empirical research in these papers is under-theorized and difficult to interpret. No single paper has a clear articulation of the connection between their empirical model specifications and underlying theory whereby the death penalty impacts the expected costs of committing crime. Many of the specification choices are ad-hoc and loosely related to deterrence theory based on the idea that potential offenders rationally consider the costs and benefits of committing a murder.

Second, many of the papers purporting to find strong effects of the death penalty on state-level murder rates suffer from basic methodological problems, including but not limited to weak instruments, questionable exclusion restrictions, failure to control for obvious factors such as sufficiently granular time effects, and incorrect calculation of standard errors which in turn has led to faulty statistical inference. While there are some fairly careful analyses (the papers by Mocan and Gittings in particular) many of the studies suffer from such basic problems that it is hard to place much weight on their findings.

Third, the lack of variation in the key underlying explanatory variables and the heavy influence exerted by a few observations in state panel data regressions seems to be a fundamental problem for all panel data studies of this question, independent of the issues we raise regarding theory and model specification. The fact that the key explanatory variable does not vary for over 85% of the observations in the typical panel data set raises serious questions regarding whether one wants to place much stock in results generated by a handful of state-year observations.

Together, these three points lead us to find the recent panel literature on whether there is a deterrent effect of the death penalty to be inconclusive as a whole and in many cases uninformative. Moreover, we do not see additional methodological tools that are likely to overcome the multiple challenges that face researchers in this domain, including the weak informativeness of the data, a lack of theory on the mechanisms involved, and the likely presence of unobserved confounders.

We do feel that there are fewer challenges for research using discrete changes in policy potentially combined with recent methodological innovations for identifying appropriate counterfactual comparison groups. In particular, the synthetic control methods and robust inference techniques proposed by Abadie and Gardeazabal (2003) and Abadie et al. (2010) may prove fruitful in estimating the overall deterrent effects associated with single-state changes in policy, such as the recent abolition of the death penalty in Illinois. While such analyses may not permit precise statements regarding the number of lives saved per

execution, the clear source of policy variation and a robust research design would at least permit an assessment of the total effect of such policies after filtered through an admittedly very black box. However, precision and identification of a causal effect would still be strongly affected by the small number of such changes in policy and the threat of unobserved confounders.

# References

Abadie A, Gardeazabal J (2003) The economic costs of conflict: a case study of the Basque Country. Am Econ Rev 93(1):113–132

Abadie A, Diamond A, Hainmueller J (2010) Synthetic control methods for comparative case studies: estimating the effect of California's tobacco control program. J Am Stat Assoc 105(490):493–505

Angrist J, Pischke J (2009) Mostly harmless econometrics. Princeton University Press, Princeton

Becker G (1968) Crime and punishment: an economic approach. J Polit Econ 76:169–217

Berk R (2005) New claims about executions and general deterrence: deja vu all over again? J Empir Legal Stud 2:303–330

Bound J, Jaeger DA, Baker RM (1995) Problems with instrumental variables estimation when the correlation between the instruments and the endogenous explanatory variable is weak. J Am Stat Assoc 90:443–450

Cameron AC, Trivedi PK (2005) Microeconometrics: methods and applications. Cambridge University Press, New York

Dezhbakhsh H, Rubin PH, Shepherd JM (2003) Does capital punishment have a deterrent effect? New evidence from postmoratorium panel data. Am Law Econ Rev 5(2):344–376

Dezhbakhsh H, Rubin PH (2010) From the 'Econometrics of Capital Punishment' to the 'Capital Punishment of Econometrics': On the Use and Abuse of Sensitivity Analysis. Appl Econ (forthcoming). Available in advance here: http://pdfserve.informaworld.com/420130_770849120_928446184.pdf

Dezhbakhsh H, Shepherd JM (2006) The deterrent effect of capital punishment: evidence from a 'judicial experiment'. Econ Inq 44(3):512–535

Donohue JJ, Wolfers J (2005) Uses and abuses of empirical evidence in the death penalty debate. Stanford Law Rev 58(3):791–845

Donohue JJ, Wolfers J (2009) Estimating the impact of the death penalty on murder. Am Law Econ Rev 11(2):249–309

Ehrlich I (1975) Capital punishment and deterrence: a question of life or death. Am Econ Rev 65(3):397–417

Fagan J, Zimring FE, Geller A (2006) Capital punishment and capital murder: market share and the deterrent effects of the death penalty. Texas Law Rev 84(7):1803–1867

Katz L, Levitt SD, Shustorovich E (2003) Prison conditions, capital punishment, and deterrence. Am Law Econ Rev 5(2):318–343

Kovandzic TV, Vieraitis LM et al (2009) Does the death penalty save lives? Criminol Public Policy 8(4):803–843

Maltz MD, Targonski J (2002) A note on the use of county-level UCR data. J Quant Criminol 18(3):297–318

Mocan HN, Gittings RK (2003) Getting off death row: commuted sentences and the deterrent effect of capital punishment. J Law Econ 46(4):453–478

Mocan HN, Gittings RK (2010) The Impact of Incentives on Human Behavior: Can We Make it Disappear?: The Case of the Death Penalty. In: Di Tella R, Edwards S, Schargrodsky E (eds) The economics of crime; lessons for and from Latin America. University of Chicago Press, Chicago

Moulton B (1990) An illustration of a pitfall in estimating the effects of aggregate variables on micro units. Rev Econ Stat 72:334–338

Moulton BR (1986) Random group effects and the precision of regression estimates. J Econometrics 32(3):385–397

Snell T (2010) Capital punishment in the United States. Bureau of Justice Statistics

Stock J, Yogo H (2005) "Testing for weak instruments in linear IV regression. In: Identification and inference for econometric models: essays in Honor of Thomas Rothenberg. Cambridge University Press, Cambridge, pp 109–120

Zimmerman PR (2009) Statistical variability and the deterrent effect of the death penalty. Am Law Econ Rev 11(2):370–398

Zimmerman PR (2004) State executions, deterrence, and the incidence of murder. J Appl Econ 7:163–193

Zimmerman PR (2006) Estimates of the deterrent effect of alternative execution methods in the United States: 1978–2000. Am J Econ Sociol 65(4):909–941