Document Title:	**Reevaluating the Deterrent Effect of Capital Punishment: Model and Data Uncertainty**
Author(s):	**Ethan Cohen-Cole ; Steven Durlauf ; Jeffrey Fagan ; Daniel Nagin**
Document No.:	**216548**
Date Received:	**December 2006**
Award Number:	**2005-IJ-CX-0020**

Reevaluating the Deterrent Effect of Capital Punishment: Model and Data Uncertainty

Ethan Cohen-Cole[1]
Steven Durlauf
Jeffrey Fagan
Daniel Nagin

Abstract

While issues of deterrence lie at the heart of criminal justice policy, there are important contexts where the studies of deterrence effects have failed to provide anything close to a scholarly consensus. A principle example of this are laws on capital punishment. Proponents argue that such laws prevent murders because potential criminals fear such strong punishment. Opponents argue that deterrence arguments do not apply in these circumstances and/or that the statistical analyses suffer from grave flaws. Each side can cite many statistical studies in support of its claims. This paper presents a methodology by which one can integrate the various studies into a single coherent analysis. We use a methodology generally called "model averaging" by which one takes weighted averages of a wide set of possible models of deterrence. *Our conclusion is that there is little empirical evidence in favor of the deterrence hypothesis.*

---

[1] Cohen-Cole: Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02210. (617) 973.3294. ethan.cohen-cole@bos.frb.org; Durlauf: Department of Economics, University of Wisconsin. 1180 Observatory Drive, Madison WI, 53706-1393; Fagan: Columbia University School of Law; 435 West 116th Street Room 634, Box D-18 New York NY 10027; Nagin: H. John Heinz III School of Public Policy & Management; Carnegie Mellon University 5000 Forbes Avenue. Pittsburgh, PA 15213-3890. The Department of Justice's National Institute of Justice has provided financial assistance. We are grateful for research assistance provided by Jon Larson. The views expressed in this paper are solely those of the authors and do not reflect official positions of the Federal Reserve Bank of Boston or the Federal Reserve System.

Reevaluating the Deterrent Effect of Capital Punishment: Model and Data Uncertainty

Ethan Cohen-Cole[1]
Steven Durlauf
Jeffrey Fagan
Daniel Nagin

Executive Summary

While issues of deterrence lie at the heart of criminal justice policy, there are important contexts where the studies of deterrence effects have failed to provide anything close to a scholarly consensus. A principle example of this are laws on capital punishment. Proponents argue that such laws prevent murders because potential criminals fear such strong punishment. Opponents argue that deterrence arguments do not apply in these circumstances and/or that the statistical analyses suffer from grave flaws. Each side can cite many statistical studies in support of its claims. Efforts to change the policy landscape are ongoing and policymakers continue to struggle with the interpreting the results of conflicting studies. Thirty-eight states currently have a death-penalty law.

The capital punishment literature has been marked by strongly opposing views. Since Issac Ehrlich's original contributions in (1975) and (1977), the field has produced a range of papers supporting and opposing capital punishment. The fundamental problem that underlies the disparate findings on deterrence effects of death sentencing is that individual studies reflect specific assumptions about the appropriate data, control

[1] Cohen-Cole: Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02210. (617) 973.3294. ethan.cohen-cole@bos.frb.org; Durlauf: Department of Economics, University of Wisconsin. 1180 Observatory Drive, Madison WI, 53706-1393; Fagan: Columbia University School of Law; 435 West 116th Street Room 634, Box D-18 New York NY 10027; Nagin: H. John Heinz III School of Public Policy & Management; Carnegie Mellon University 5000 Forbes Avenue. Pittsburgh, PA 15213-3890. The Department of Justice's National Institute of Justice has provided financial assistance. We are grateful for research assistance provided by Jon Larson. The views expressed in this paper are solely those of the authors and do not reflect official positions of the Federal Reserve Bank of Boston or the Federal Reserve System.

variables, model specification, etc. on the part of the researcher, and can have major effects on the conclusions of a particular data analysis.

The existing research on this topic comes to sufficiently differing conclusions predicated upon one or more underlying assumptions to call into question the ability of any single model to explain the impact of execution laws. Such dependence on the specifics of research design, from data cleaning to aggregation to model choice, forms the basis for the use of averaging techniques. That is, since relatively minor variations in model or variable choice can lead to dramatic changes in conclusions, one suspects that inclusion of the information content of all of these models would lend itself to conclusions upon which policymakers could be more confident.

This paper will describe a method which accounts for model uncertainty and places into a form that is easily interpretable to policymakers. More generally, the structure of model averaging may be understood as follows. Suppose one wishes to produce an estimate of some object of interest δ which measure the effects of a policy. In the context of the capital punishment literature, δ tends to be the coefficient on the execution variable in some deterrence regression. Conventional statistical methods may be thought of as calculating an estimate that is model specific, $\hat{\delta}_m$. In the model averaging approach, one attempts to eliminate conditioning on a specific model. To do this, one specifies a space of possible models *M*. The true model is unknown, so from the perspective of the researcher, each model will have some probability of being true. These model probabilities will depend both on the prior beliefs of the researcher as well as on the relative goodness of fits of the different models given available data *D;* hence each model will have a posterior probability: $\mu\left(m \middle| D\right)$. These posterior probabilities allow us to average the model-specific estimates: $\hat{\delta} = \sum_m \hat{\delta}_m \mu\left(m \middle| D\right)$.

Using this methodology, we estimate the deterrent effect of capital punishment. Our finding is that there is little evidence of a deterrent effect.

Reevaluating the Deterrent Effect of Capital Punishment: Model and Data Uncertainty

Ethan Cohen-Cole[1]
Steven Durlauf
Jeffrey Fagan
Daniel Nagin

Section 1 Introduction

While issues of deterrence lie at the heart of criminal justice policy, there are important contexts where the studies of deterrence effects have failed to provide anything close to a scholarly consensus. A principle example of this are laws on capital punishment. Proponents argue that such laws prevent murders because potential criminals fear such strong punishment. Opponents argue that deterrence arguments do not apply in these circumstances and/or that the statistical analyses suffer from grave flaws. Each side can cite many statistical studies in support of its claims. Efforts to change the policy landscape are ongoing and policymakers continue to struggle with the interpreting the results of conflicting studies. Thirty-eight states currently have a death-penalty law.

The fundamental problem that underlies the disparate findings on the deterrent effect of death sentencing is that individual studies reflect specific assumptions about the appropriate data, control variables, model specification, etc. on the part of the researcher. These assumptions can reflect an expression of possible deterrence explanation (e.g. using incarceration rates as a control), and can have major effects on the conclusions of a particular data analysis. However, one is hard pressed to make a compelling argument that inclusion of a given variable over another is the crucial decision in the formation of a deterrence study. This is based on the fact

[1] Cohen-Cole: Federal Reserve Bank of Boston, 600 Atlantic Avenue, Boston, MA 02210. (617) 973.3294. ethan.cohen-cole@bos.frb.org; Durlauf: Department of Economics, University of Wisconsin. 1180 Observatory Drive, Madison WI, 53706-1393; Fagan: Columbia University School of Law; 435 West 116th Street Room 634, Box D-18 New York NY 10027; Nagin: H. John Heinz III School of Public Policy & Management; Carnegie Mellon University 5000 Forbes Avenue. Pittsburgh, PA 15213-3890. The Department of Justice's National Institute of Justice has provided financial assistance. We are grateful for research assistance provided by Jon Larson. The views expressed in this paper are solely those of the authors and do not reflect official positions of the Federal Reserve Bank of Boston or the Federal Reserve System.

that these assumptions are themselves typically not falsifiable. For example, in of itself an assumption that one should include a given variable in a deterrence regression cannot be falsified directly. As a result, two researchers - both of whom may have developed conceptually valid and potentially correct "models" (by which we mean a collection of assumptions) - can reach distinct, and even opposing, conclusions. The resulting lack of certainty about which model is the "true" one, or "model uncertainty", is the technical focus of this research project. In particular, we will utilize model averaging, a method that enables a researcher to take a weighted average of findings across possible models, to develop inferences that are not dependent on the assumption that one of the models is true.

Section 2 of this paper reviews the relevant literature on capital punishment and deterrence effects of execution as well as current work on efforts to handle model uncertainty. Sections 3 and 4 outline the baseline methodology used in this paper. Section 5 discusses the data used in the paper and replicates the results from two current papers that draw on the same data, Dezhbakhsh, Rubin and Shepherd (2003) and Donohue and Wolfers (2005). Section 6 provides details on the econometric implementation of our model averaging methodology and Section 7 concludes.

Section 2 The academic debate about Capital Punishment laws

The capital punishment literature has long been marked by strongly opposing views. The academic debate originated with the contributions of Issac Erhlich (1975, 1777), who used a theoretical model derived from the 'rational choice' models of economic which explained murder as a rational tradeoff between the costs and benefits of illegal behavior (i.e. murder). In particular, using such a framework, one could evaluate how individuals responded to law enforcement actions. Ehrlich's much evaluated result found that the rational response to the death penalty was a decline in the propensity to commit murder.

Since Ehrlich's original contributions in the late 1970s, the field has produced a range of principally empirical papers supporting and refuting his work. The critiques of Ehrlich's work came from the economics and the legal professions. Within economics, the principal and broadest critique that followed of Ehrlich's work was that his estimates were highly sensitive to his particular econometric methods (McAleer and Veall 1989, Leamer 1983, McManus 1985). Each of these used method to evaluate whether the results found were sensitive to the particular assumptions used in setting up the econometric study. In the legal field, Baldus and Cole (1975) and Bowers and Pierce (1975) both argued in the Yale Journal that the results were unfounded.

Though Ehrlich's work has been challenged on many grounds, his principal findings of a deterrent effect within a rational choice model have been supported by many. Our perspective on the literature is that it is characterized principally by studies that ask whether a regressor that proxies for the probability of executions possesses a statistically significant coefficient. Essentially these are all recharacterizations of Ehrlich's original rational choice model interpreted as some type of econometric specification. If the standard of a coefficient's significance is passed, then one sees often some discussion of the magnitude of the coefficient (i.e. the number of murders deterred or caused) and a discussion of alternate models. There are many lines of dispute about the construction of these models and there have been a number of studies on sensitivity adjustments to this basic setup (Isaac Ehrlich and Zhiqiang Liu, 1999, Edward E. Leamer, 1983, Michael McAleer and Michael R Veall, 1989, Walter S McManus, 1985).

Though in principle, the lines of debate are around the end object, whether deterrence is effective, in practice, there is significant dispute about the construction of an appropriate "model". That is, while Ehrlich's concept, and the rational choice framework in general, is instructive in theory, in practice, one can justify conceptually a very wide range of appropriate models. As a result, the field has produce a wide variety of methods employed to support or refute Ehrlich's original conclusion. For the sake of explication, we identify five factors in the capital punishment literature that have been in dispute. These are not claimed to be exhaustive, but are employed in order to illustrate both the range of the debate as well as provide some indication of the connection between a researcher's choice of assumptions and his/her concomitant conclusions The exact nature of these disputed factors is not of particular relevance at this stage. Our Table 1 here serves simply to illustrate that despite the form of statistical machinery brought to bear, there have been opposing conclusions in each case.

Table 1: Examples of Specification Variation in Capital Punishment Literature

Controls	• Mocan and Gittings (2001) use pardon data. • Shepard (2003) studies time on death row.	• Katz et al. (2001) control for prison conditions.[2]
Functional Form	• Ehrlich uses a log-log form (1975, 1977) and varies it in Ehrlich and Liu (1999).	• Passel and Taylor (1977) use a linear form.
Data Stationarity	• Layson (1983, 1985) assumes stationarity.	• Cover and Thistle (1988) assume non-stationarity.
Time period	• Chressanthis (1989) uses data from 1966-1985.	• Grogger (1990) uses California data from 1960-1963.
Data Choice	• Dezhbakhsh et al. (2002) uses county-level panel data.	• Sorenson et al. (1999) uses Texas monthly data from 1984-1997.

As we can see, the choice of functional form, controls variable, etc., and the particular assumption made is non-trivial in the resulting findings. In Ehrlich's original studies he uses a log-log form for his time-series regression and finds evidence of deterrence. Passel and Taylor (1977), using a close copy of Ehrlich's data change the functional form to a linear one and find no deterrence. There are logical arguments to be made for both forms, but one hopes that important policy decisions not rest on a relatively esoteric decision on whether to employ a linear or a logarithmic-transformed series of data. Other issues arise in the use of time-series data – if

[2] Katz et al find that the conditions are a deterrent, but the death penalty is not.

the data are non-stationary, it is well known that standard regression results will be biased if series is not differenced prior to analysis. Layson uses original data (1983, 1985) and Cover and Thistle arguing that the data should be differenced, find a reduced result (1988).

The choice of sample also appears to matter to the conclusions reached. Grogger (1990) argues that inspection of daily homicide information from California in the 1960 produces very different results than the original Ehrlich study. Chressanthis (1989) finds a similar result to Ehrlich using an expanded time period from 1996-1985.

Others supporting Ehrlich's conclusions include Brumm and Cloninger, 1996, Cloninger and Marchesini, 2001, Cloninger, 1977, Dezhbakhsh, Rubin and Shepherd, 2003, Dezhbakhsh and Shepherd, 2006, Ehrlich and Gibbons, 1977, Mocan and Gittings, 2003 Yunker, 1976; Others opposing include Bowers and Pierce, 1975, Donohue and Wolfers 2005, Grogger, 1990, Hoenack and Weiler, 1980, Leamer, 1983, McManus, 1985, Passell and Taylor, 1977, Zimmerman, 2004. A good summary of the recent literature and its policy impact can be founding Fagan 2006.

In our empirical work below, we will explicitly handle disputes over controls and over data choice – these appear to be the most salient factors to the debate. While each of these studies is on its own terms a scholarly contribution, from the perspective of policy evaluation, the disparate approaches and concomitant differences in results leave a policymaker with little clear direction. This is unsurprising since "rational choice/economic" models of crime are very qualitative in the sense that they provide little guidance on the choice of control variables, the functional form for statistical analysis, etc. If one thinks of each study as using a particular model of crime, in which a model refers to the choice of data, time period, control variables, and statistical specification, then the source of the wide range of claims in the capital punishment literature is *model uncertainty*.

While most deterrence studies acknowledge the problem of model uncertainty, in practice, this amounts to considering relatively small adjustments to an initial baseline model. Though in these literatures the range of such modifications can be very wide – even into the thousands of specifications - such an approach may be faulted for two reasons. First, the model "spaces" considered are arbitrary and small in the sense of constituting a local neighborhood of the baseline model within the space of possible models. Second, such methods provide no guidance on how to integrate findings across models. For example, suppose one finds a deterrent effect under one group of model specifications but no effect under another set, what should a policymaker conclude?

Literature on prior efforts to handle model uncertainty

Some authors have attempted to address this issue in the past. We discuss a few such efforts here. First, the topic was addressed with a technique called Extreme Bounds Analysis (Edward E. Leamer, 1983). Applying Extreme Bounds Analysis on a coefficient estimate involves estimating a set of alternate model specifications and seeing how the coefficient estimate changes. If the sign of the coefficient is not constant, i.e. it "flips" across specifications, one concludes that the evidence is "fragile". This strategy suffers from two problems. One, the conclusion of fragility can be influenced by the choice of coefficient that one considers; that is, if one were to choose another coefficient the results for the original choice may change. Second, the procedure fails to fully integrate information across the complete set of models that are analyzed – as it does not account for goodness of fit differences across models. Moreover, extreme bounds analysis concludes evidence is fragile even when, out of 1000 regressions, 999 produce a positive coefficient estimate and 1 produces a negative estimate. Brock, Durlauf and West (2003) in fact show that extreme bound analysis implies a special and extreme form of risk aversion if one uses it to guide policy decisions.

Second, McManus (1985) used a precursor of sorts to the method advocated in our paper. McManus' paper applied a Bayesian-style analysis to look at the importance of a researcher's prior views on the results of a deterrence study of capital punishment. He specified five distinct views of the world and posited a method by which such views would be implemented in a deterrence study. The intuition behind his study is straightforward and appropriate; he found that even by specifying a very small number of models, the results can be quite varied.[3] This method had two problems. One, the number of models in his study is quite limited, and thus subject to a similar type of critique that we have leveled at the remainder of the deterrence literature. Though, in his defense, McManus is quite disciplined about including varied models. Two, in the past 20 years, the statistical machinery to integrate models in a Bayesian context, now often called model averaging, has been greatly developed.

[3] McManus called the models in his papers "beliefs". Since we use the term beliefs to discuss priors on variable inclusion, discussed below, we match terminology by using "models" here.

A growing literature has emerged in the use of this method known as model averaging. The basic concept is not far removed from McManus. It seeks to avoid researcher bias in the determination of variable choice, specification choice, etc. by including a large set of possibilities into a single analysis. The earliest discussion of model averaging is Leamer (1978) but the approach seems to have been dormant until the middle 1990's; Draper (1995), Raftery (1995), and Raftery, Madigan and Hoeting (1997) apparently initiated recent interest. Useful introductions are available in Wasserman (1996) and Hoeting, Clyde, Madigan and Raftery (1999). Model averaging has been advocated and employed in Brock and Durlauf (2001a), Brock, Durlauf and West (2003), Fernandez, Ley and Steel (2001), Sala-i-Martin, Doppelhofer, and Miller (2004) and Masanjala and Papageorgiou (2004b). We continue in the next section to explain the nature and implementation of model averaging.

Section 3 Model Uncertainty and Model Averaging

As the prior section illustrates, the existing research on this topic comes to sufficiently differing conclusions predicated upon one or more underlying assumptions to call into question the ability of any single model to explain the impact of deterrence laws. Such dependence on the specifics of research design, from data cleaning to aggregation to model choice, forms the basis for the use of averaging techniques. That is, since relatively minor variations in model or variable choice can lead to dramatic changes in conclusions, one suspects that inclusion of the information content of all of these models would lend itself to conclusions upon which policymakers could be more confident. This section will describe a method which accounts for model uncertainty and places into a form that is easily interpretable to policymakers.

This project intends to account explicitly for model uncertainty in the analysis of deterrence laws. We will follow the model averaging literature as our mechanism for dealing with model uncertainty. We will adapt the general framework that has been developed in the statistics literature; however, we will use standard frequentist estimators.[4] This frequentist approach to model averaging is described in Sala-i-Martin, Doppelhofer, and Miller (2004) and Brock, Durlauf, and West (2003). The basic idea of model averaging is straightforward. Consider an object of interest – in this case the difference in crime rates under alternate laws – and take a

[4] Within the statistics literature, model averaging is usually done in Bayesian contexts. A full discussion of the difference between Bayesian and Frequentist approaches is beyond this paper.

weighted average of the findings across all possible models that can explain the effect. Each model's effect is weighted based on its ability to explain the data.

To see how this procedure` works intuitively, consider the question of modeling time trends in deterrence regressions. One of the sources of the different findings in Dezhbakhsh, Rubin and Shepherd and Donohue and Wolfers is the choice of data to use. Donohue and Wolfers argue that excluding a single state (CA or TX) from the Dezhbakhsh, Rubin and Shepherd data is a major source of the difference in findings. One can think of this disagreement as a reflecting a simple form of model uncertainty in that the model space has only two elements: a set of data that includes CA and one that does not. (There are of course many other differences between the two papers, but we ignore those for expositional reasons). How would we propose adjudicating the disagreement? We would argue that one should average the estimates from the two studies by taking a weighted average of the results from each, where the weights are posterior model probabilities.[5] Dezhbakhsh, Rubin and Shepherd can be interpreted as placing a prior probability of 1 on the model with data that includes CA whereas Donohue and Wolfers place a prior probability of 1 on the model without CA. By setting these two cases alongside one another, it is possible to evaluate the relative probability that each can describe the data used.

More generally, the structure of model averaging may be understood as follows. Suppose one wishes to produce an estimate of some object of interest δ which measure the effects of a policy. In the context of the capital punishment literature, δ tends to be the coefficient on the execution variable in some deterrence regression. Conventional statistical methods may be thought of as calculating an estimate that is model specific, $\hat{\delta}_m$. In the model averaging approach, one attempts to eliminate conditioning on a specific model. To do this, one specifies a set or space of possible models *M*. The true model is of course unknown, so from the perspective of the researcher, each model will have some probability of being true.[6] These probabilities depend on the relative goodness of fit of the different models given available data *D* as well as the prior beliefs of the researcher (something we discuss below); hence each model will have a posterior probability: $\mu(m|D)$. This can be read as the probability of a given model (*m*)

[5] This can be thought of simply as the conditional probability that a given model describes the data. We discuss this at greater length below.

[6] If one model is true then this method produces a probability of 1 for that model and zero for all others.

conditional on data available, *D*. These posterior probabilities allow us to average the model-specific estimates: $\hat{\delta} = \sum_m \hat{\delta}_m \mu(m|D)$

The estimate $\hat{\delta}$ thus accounts for the information contained in each specific model about δ and weights this information according to the likelihood the model is the correct one. As suggested above, in the case that a single model is true, it will receive a weight of 1. Brock et al. (2003) argue that the strategy of constructing posterior probabilities that are not model-dependent is the appropriate one when the objective of the statistical exercise is to evaluate alternative policy questions such as whether to implement capital punishment in a state. Notice that this approach does not identify the "best" model; instead, it studies the effect of the policy, i.e. the parameter δ. Thus, while the exercise could in theory find a single model with weight of one, in practice, a finding that a give model, *m**, within some space *M*, has the highest conditional probability of describing the data is not a recommendation to select that model. Instead, it merely identifies that model as being the one that has the largest contribution to $\hat{\delta}$.

Notice that averaging across models means that a key role is played by the posterior model probabilities. Using Bayes rule, the posterior probability may be rewritten as

$$\mu(m|D) = \frac{\mu(D|m)\mu(m)}{\mu(D)} \propto \mu(D|m)\mu(m). \quad (1)$$

The calculation of posterior model probabilities thus depends on two terms. The first, $\mu(D|m)$ is the probability of data given a model. Raftery (1996) has developed a proof to illustrate that the probability of a model given a dataset can be calculated using the ratio of a given model's likelihood to the sum of the likelihood of all models in the space *M*.[7] This derivation allows us to use the likelihood of a given model for $\mu(D|m)$. The second term, $\mu(m)$, is the prior probability assigned to model m. Hence, computing posterior model probabilities requires specifying prior beliefs on the probabilities of the elements of the model space M (see below).

Calculating Variance of $\hat{\delta}$

[7] A full discussion of the derivation of

Applying the concepts above, one can compute the uncertainty, i.e. variance, associated with an estimated policy effect when one avoids conditioning on knowing the true model. This variance is written as:

$$var\left(\hat{\delta}\right) = \sum_{m \in M} \mu\left(m \middle| D\right) var\left(\hat{\delta}_m\right) + \sum_{m \in M} \mu\left(m \middle| D\right)\left(\hat{\delta} - \hat{\delta}_m\right)^2 \qquad (2)$$

This formula illustrates how model uncertainty affects the overall uncertainty one should associate with given parameter estimates. The variance of $\hat{\delta}$ consists of two separate parts. The first, $\sum_{m \in M} \mu\left(m \middle| D\right) var\left(\hat{\delta}_m\right)$, is a weighted average of the variances for each model and is effectively the same construction as the estimate itself. The second term however, reflects the variance across models in M; this reflects that the models are themselves different. This term, $\sum_{m \in M} \mu\left(m \middle| D\right)\left(\hat{\delta} - \hat{\delta}_m\right)^2$, is not determined by the model-specific variance calculations and in some sense captures how model uncertainty increases the variance associated with a parameter estimate relative to conventional calculations. To see why this second term is interesting, suppose that $var\left(\hat{\delta}_m\right)$ is constant across models. In general, one should not conclude that the overall variance is equal to this same value. So long as there is any variation in $\hat{\delta}_m$ across models, then $var\left(\hat{\delta}_m\right) < var\left(\hat{\delta}\right)$; the cross model variations in the mean increase the uncertainty (as measured by the variance) that exists with respect to δ.

The importance of this last section is that model averaging not only permits policymakers to account for differences in predictions of the direction of effect of capital punishment, but it also allows policymakers to have a greater understanding of the errors in these predictions of the effects of changes in these policies. Once such additional variance has been accounted for, a finding of a significant effect (positive or negative) of deterrence laws allows a policymaker to have that much more confidence about her decisions.

Section 4 Uncertainty over Data

As we discussed in the section above, it appears that policy conclusions from this literature depend in part on the construction of datasets used. This is important in this context as much of the debate over the effect of capital punishment has included differences over the choice of data construction (see Table 2, Panels D and E for example).

As such, an innovation in this paper is a generalization of the model averaging methodology from section 4. In the language of the Section above, the coefficients, $\hat{\delta}$, calculated above would need to be the weighted average of all models in the space *M* as well as across data choices *D*. As such, we can adapt the results above such that we integrate over the space of data possibilities as well.

We can generalize the above $\hat{\delta} = \sum_m \hat{\delta}_m \mu\left(m \middle| D\right)$ to allow for difference in data as well.

$$\hat{\delta} = \sum_{d \in D} \sum_{m \in M} \hat{\delta}_m \mu\left(m \middle| d\right) \mu\left(d\right) \tag{3}$$

where the $\mu\left(d\right)$ are the prior probabilities associated with each possible dataset. This allows one to average over all combinations of data and models to produce a single weighted parameter, $\hat{\delta}$.

There are some technical differences in the application of data uncertainty. While a full discussion of these is beyond the scope of this paper, there are two issues at hand. The first is whether in the calculation of posterior probabilities conditional on a set of data *d*, one assumes the existence of a correct data set. In our setup here, we effectively interpret a model-dataset pair as a type of model, thus making no assumption on the validity of the data themselves. Second, in order to calculate $\hat{\delta}$, one needs that the probability $\mu\left(m \middle| d_1\right)$ can be compared to $\mu\left(m \middle| d_2\right)$, where $d_1, d_2 \in D$. Effectively, this means that we need for likelihood to be comparable across data. We sidestep the issue by calculating our the posterior probabilities within each data before comparing across them – this means that the comparison across datasets is done at the level of relative likelihoods.

With this caveat, the implementation of uncertainty over data is effectively the same as implementation over models alone.

Section 5 Replicating Dezhbakhsh, Rubin and Shepherd and Donohue and Wolfers.

One of the prominent papers arguing for the presence of a deterrent effect of capital punishment is Dezhbakhsh, Rubin and Shepherd (2003). They use county-level data for the post-moratorium period (1977-1996), at the time the most detailed and disaggregate dataset to have been used. The data is publicly available from the Department of Justice's Bureau of Justice Statistics, the FBI's Uniform Crime Reports, and the Bureau of the Census. Donohue and Wolfers (2005) replicate many of the results in Dezhbakhsh et al.'s, and we use data provided by Dezhbakhsh, Rubin and Shepherd in this section to replicate and discuss the results from both papers. Both papers estimate some version of a deterrence regression drawn from Ehrlich (1977). That is, the murder rate is a function of three principal deterrence variables: the probability of arrest, the probability of receiving a death sentence conditional of being arrested, and the probability of being executed conditional on receiving a death sentence. Of course, all of the specifications have various demographic and economic control variables added. They include controls for the aggravated assault rate, the robbery rate, the population proportion of 10-19 year olds, 20-29 year old, demographic percentages of blacks, percentage of non-black minorities, percentage of males, the percentage of NRA members, real per capita income, real per capita income maintenance payments, real per capita unemployment insurance payments, and the population density.

Thus, the principle regression is:

$$\begin{aligned}\frac{Murders_{c,s,t}}{pop_{c,s,t}} &= \delta_0 + \delta_1 \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} + \delta_2 \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} + \delta_3 \frac{Executions_{s,t}}{DeathSentences_{s,t-6}} \\ &+ \gamma_1 \frac{Assaults_{c,s,t}}{Population_{c,s,t}} + \gamma_2 \frac{Robberies_{c,s,t}}{Population_{c,s,t}} + \gamma_3 Demographics_{c,s,t} + \\ &+ \gamma_5 economy_{c,s,t} + \gamma_6 \frac{NRAmembers_{s,t}}{population_{s,t}} + \sum\nolimits_c \gamma_{7,t} county_c + \sum\nolimits_t \gamma_{8,t} time_t + \eta_{s,t} + \varepsilon_{c,s,t}\end{aligned} \tag{4}$$

They use first stage regressions to estimate their variables of interest as follows:

$$\frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} = \psi_0 + \psi_1 \frac{Murders_{c,s,t}}{Pop_{c,s,t}} + \psi_2 PolicePayroll + \sum\nolimits_t \psi_{3,t} time_t + \varepsilon'_{c,s,t} \tag{5}$$

$$\begin{aligned}\frac{DeathSentences_{s,t}}{Arrests_{s,t}} &= \theta_0 + \theta_1 \frac{Murders_{c,s,t}}{pop_{c,s,t}} + \theta_2 JudicialExpense + \theta_3 PartisanInfluence \\ &+ \theta_4 Admissions + \sum\nolimits_t \theta_{5,t} time_t + \varepsilon''_{c,s,t}\end{aligned} \tag{6}$$

$$\begin{aligned}\frac{Executions_{s,t}}{DeathSentences_{s,t}} &= \phi_0 + \phi_1 \frac{Murders_{c,s,t}}{pop_{c,s,t}} + \phi_2 JudicialExpense + \phi_3 PartisanInfluence \\ &+ \sum\nolimits_t \phi_{4,t} time_t + \varepsilon'''_{c,s,t}\end{aligned} \tag{7}$$

In these cases, we have followed the notation from Donohue and Wolfers. The variable $pop_{c,s,t}$ indicates the population in county *c*, state *s*, and time *t*, divided by 100,000. Partisan influence in the Republican presidential candidate's vote share in the most recent election, and Admissions is the prison admission rate. Note that some of the key variables are estimated at the state level (the subscript *c* is omitted in these cases).[8] Additional information is available in the original text.

Tables 3 and 4 in Dezhbakhsh, Rubin and Shepherd (p 362-363) present the results from Equation 3 given six different versions of the variable $\frac{Executions}{DeathSentences}$. Their measures of execution probabilities in the six columns are as follows:

Columns 1 and 4: $$\frac{Executions_{s,t}}{DeathSentences_{s,t-6}} \tag{8}$$

Columns 2 and 5: $$\frac{Executions_{s,t+6}}{DeathSentences_{s,t}} \tag{9}$$

Columns 3 and 6: $$\frac{\sum_{t=-3}^{3} Executions_{s,t}}{\sum_{t=-9}^{-4} DeathSentences_{s,t}} \tag{10}$$

Columns 1-3 omit observations in which there are no death sentences. Columns 4-6 use a method to use the probability of the most recent year which had a death sentence. Donohue and Wolfers' Table 7 (p 824) repeats the results. While both papers provide a wide variety of other information, we focus on this Table as it provides a useful case to illustrate our points.

Table 3 below reproduces the Dezhbakhsh, Rubin and Shepherd results. No computation has been done here, the information has simply been transferred for comparability. The basic point made is that all of the coefficients are negative in sign and most are significant at the 5% level. When translated into a calculated number of lives saved per execution, the estimates range from 19 to 36 persons.

Table 4 below replicates the Donohue and Wolfers innovations. Donohue and Wolfers have graciously provided their data and code to us, and we use these to produce Table 4 in its entirety. They have made minor modifications to the Dezhbakhsh, Rubin and Shepherd

[8] It is worthwhile at this stage to point out that DRS use a combination of county and state effects to predict county-level murder rates. We will not discuss the merits (or difficulties) of this type of estimation strategy other than to comment that it will not impact the model averaging exercise that we are conducting.

construction to illustrate the sensitivity of the original results. The three innovations they made were to include only a single voting variable instead of six in the first stage regressions (see Appendix for full specification), and to drop Texas or California from the analysis. Each of these three analyses causes the signs to reverse. That is, each execution is predicted to *increase* the murder rate.

Section 6 Model Averaging Results

In this section, we present results from the model averaging methods describing in the above two sections. As mentioned, Table 3 and Table 4 below are the second stage regression results from Donohue and Wolfers. The appendix contains information for all variables as well as detail on each regression specification.

After reproducing the basic result from Dezhbakhsh, Rubin and Shepherd and Donohue and Wolfers, we use the theory from Section 3 and Section 4 above to estimate the weighted average coefficients in a variety of cases. The mechanism here is to provide the highest weight to those models most likely to explain the data. While Table 4 illustrates the apparent sensitivity of the Dezhbakhsh, Rubin and Shepherd estimates appear to specification, this is not necessarily an indication that their conclusions are incorrect. If Donohue and Wolfers picked unlikely models or data, one should be skeptical of their associated conclusions.

We will estimate the results in a number of ways. Table 5 allows for averaging across models and data in different combinations. To begin, we treat the Dezhbakhsh, Rubin and Shepherd first-stage regressions as given and estimate model averaged coefficients for the six regressions in panel A of the Table 3. That is, we re-estimate their first stage regressions, assuming that these specifications are ‘correct’. Using the predicted variables

These six specify different methods of proxying for expectations of criminals regarding the various deterrence variables. Each dependent variable is the murder rate per 100,000 population. The six different dependent variables are defined slightly differently by Dezhbakhsh, Rubin and Shepherd, but are all various proxies for the way that individuals form expectations of deterrence variables.

Our estimates in Table 5 take the Dezhbakhsh, Rubin and Shepherd data, their choice of candidate variables, as given. Once the first stage regressions are complete, we define the model space as all possible combinations of 12 of their most important variables. Thus, for each of the 6 deterrence proxies specified, we run 2^12=4096 regressions of the form in equations 3-6. The variables that we consider are the aggravated assault rate, the robbery rate, the population proportion of 10-19 year olds, 20-29 year old, demographic percentages of blacks, percentage of non-black minorities, percentage of males, the percentage of NRA members, real per capita income, real per capita income maintenance payments, real per capita unemployment insurance payments, and the population density. For each column 1-6, we report the model and data averaged coefficients on the probability of arrest, the probability of death sentence, and the probability of execution. All 18 of the key model average coefficients in the 6 columns are insignificant. The signs are positive in most cases, which would reject the Dezhbakhsh, Rubin and Shepherd conclusions, but the standard errors are very large. The estimated lives saved trade off are all negative and large. We also report a calculation for the implied trade off in terms of net lives saves from each execution. A positive number in this row suggests that an execution produces the social benefit of savings lives by deterring future murders. Each of the six here are negative, but also with very large standard errors.

Our next step is to incorporate some of the critiques of Donohue and Wolfers into our model and data averaging techniques.[9] First, we illustrate the results of averaging over only a number of different data possibilities. Donohue and Wolfers suggest three simple data modifications that they show to have a large impact on the Dezhbakhsh, Rubin and Shepherd results (see Table 4). We test this hypothesis by specifying four sets of data: the original Dezhbakhsh, Rubin and Shepherd data, as well as three additional datasets as suggested by Donohue and Wolfers. The three modifications are 1) using a single voting variables, 2) dropping Texas, and 3) dropping California. In the first case, the voting variables are the percentage of statewide votes going to the republication presidential candidate in each of six elections. This produces six distinct variables. In the cases where we produce results for the first change in data specification, we use on the vote percentage from the most recent election. We assign a prior probability to the Dezhbakhsh, Rubin and Shepherd data of ½ and a probability of

[9] A couple of technical notes are relevant at this stage. First, the development of model averaging techniques to date has been limited to certain types of regression. To use instrumental variables regression in a model averaged context, there is some outstanding debate about whether one should “select” an optimal first stage model or use model averaged coefficients for use the second stage. To be consistent with the remainder of the paper, we use model averaged coefficients in Table 8 to produce our fully model averaged results. As well, to be able to interpret variance estimates in a model averaged context, one must use standard OLS techniques.

1/6 to each of the three Donohue and Wolfers models, such that the combined probability of the Donohue and Wolfers models is equal to ½ as well. We use the baseline specification from Dezhbakhsh, Rubin and Shepherd for both the first and the second stage regressions. Table 6 reports the same 18 coefficients as in Table 5, and repeats the net lives analysis. Again, the results show very large standard errors in all cases.

The subsequent Table 7 combines the methods used in Table 5 and Table 6 such that the averaging takes place over both the second stage as well as over data. We continue the priors assignment used for the two previous tables. Each of the 4096 models receive equal weight and the data possibilities are weighted at ½, 1/6, 1/6, 1/6. As in Table 5 and Table 6, the results again show very large standard errors.

Finally, Table 8 shows MA coefficients where we allow for a large model space in the first stage regressions as well. In this case, we specify that for each of the 14 first stage regressions, we allow for 2^12 unique possible models. In the first stage regressions, we iterate over either seven or twelve variables in most cases. The variables are police expenditure, judicial expenditure, assault rate, robbery rate, state NRA membership, prison admissions, and either one or six voting variables as discussed above. As Dezhbakhsh, Rubin and Shepherd and Donohue and Wolfers have done, we use the predicted values of the variables from the first stage regressions in our second stage. In our case, instead of using the OLS coefficients in each case, we averaged coefficients for each of the 14 specifications. These new predicted variables are then used in the second stage regressions. As in Tables 4 and 6, the second stage regressions are also estimated using the averaging methods described. Again the results are that none of the 18 coefficients are significant.

The summary table below provides a quick reference for the exercises conducted.

Table 2: ***Summary of Averaging Exercises***

Table	***First Stage Models***	***Second Stage Models***	***Data***
Table 5	From Dezhbakhsh, Rubin and Shepherd	MA coefficients	From Dezhbakhsh, Rubin and Shepherd
Table 6	From Dezhbakhsh, Rubin and Shepherd	From Dezhbakhsh, Rubin and Shepherd	MA coefficients
Table 7	From Dezhbakhsh, Rubin and Shepherd	MA coefficients	MA coefficients
Table 8	MA coefficients	MA coefficients	MA coefficients

Our tentative conclusion is that there is no statistical evidence of a link between capital punishment and deterrence. The existing results that claim such a link appear to be based on a perhaps inadvertent selection of individual, and highly unlikely, models. In the course of this evaluation, we have been unable to locate a reasonably-sized model space that produced even a single significant negative coefficient on one of the three deterrence variables.

Additional Evaluation

This section will provide some additional insight into the results above. We have shown model-averaged coefficients in Table 5-8 that fail to support the link between deterrence and capital punishment. These are the synthesis of thousands of potential specifications, presentation of which would of course be impossible in the space of this paper. To illuminate some of features of these specifications we show the following.

Table 9 shows the most likely handful of models, along with the coefficient estimate for the execution probability deterrence variable (δ_3, from equation 1), and the posterior model probability for that model. These suggest that the weighted average results are positive due to the higher probability associated with the dataset that includes a single voting variable rather than 6. In fact, the first 100 of the highest probability models come from that dataset.

A couple of charts use useful here as well. Figure 1 shows an example of distribution of coefficients from one of the models. Notice that it is bimodal, with one mode below zero and another above. One can thus easily imagine how two authors with different starting point for research could construct studies with fundamentally different results. While the mass of observations to the left of the zero point is significantly smaller than that to the right, it is sufficiently large for a researcher to find a non-trivial number of local variations to support robustness claims. Figure 2 shows the same information decomposed into the four datasets used in the analysis. Each of the four show the same bimodal distribution.

Figures 3 and 4 show histograms from data used to produce Table 8 – the fully averaged exercise. Figure 3 is a histogram of coefficients on the execution variable (row 3, column 1), and Figure 4 is a histogram of the corresponding number of estimated lives saved. Notice the bimodal nature of both distributions. This provides some additional insight into the disagreement found in the literature. In Figure 4, we also mark the Dezhbakhsh, Rubin and Shepherd and Donohue and Wolfers estimates of net lives saved.

Section 7 Conclusions

We conclude from this study that there is little evidence of a deterrent effect from capital punishment laws. Our conclusion is based on a model averaging approach in which we integrate a number of varying approaches into a single analysis. The study has produced estimates based on a wide variety of models and a number of different data constructions.

This allows one to avoid the problem of needing to rely on specific assumptions about the appropriate data, control variables, model specification, etc.

Table 3

Dependent Variable: *Annual Homicides per 100,000 Residents*						
	Panel A: Replication of Dezhbakhsh, Rubin and Shepherd Estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
Probability of Arrest	-4.04*** (0.58)	-10.10*** (0.57)	-3.33*** (0.52)	-2.27*** (0.50)	-4.42*** (0.45)	-2.18*** (0.48)
Probability of Death Sentence Given Arrest	-21.80 (18.6)	-42.41*** (13.71)	-32.12** (16.22)	-3.62 (14.53)	-47.66*** (10.45)	-10.76 (13.13)
Probability of Execution Given Death Sentence	-5.17*** (0.81)	-2.89*** (0.46)	-7.40*** (0.72)	-2.71*** (0.62)	-5.20*** (0.27)	-4.78*** (0.56)
	Panel B: Replication of Dezhbakhsh, Rubin and Shepherd Estimates, Implied Life-Life Tradeoff					
Net Lives Saved	36.1*** (5.8)	19.7*** (3.3)	52.0*** (5.1)	18.5*** (4.4)	36.3*** (1.9)	33.3*** (4.0)

Notes: Columns 1-6 show slight differences in how to proxy for expectations of criminals vis-à-vis the deterrence variables. See Dezhbakhsh, Rubin and Shepherd note 11, pages 362-363. Panel A replicates the estimates of the impact of deterrence variables on murder rates, using the specification and county-level data from Dezhbakhsh, Rubin & Shepherd, *supra* note 11, at 362-63 tbls.3-4. Panel B converts these estimates into net lives saved per execution, showing a net savings of from eighteen to fifty-two lives per execution. Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership; and county and year fixed effects. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively. *(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996.

Table 4

Dependent Variable: *Annual Homicides per 100,000 Residents*						
	Panel C: Allowing Only One Partisanship Variable					
	(1)	(2)	(3)	(4)	(5)	(6)
Net Lives Saved	-24.5*** (8.0)	-53.8*** (6.0)	-43.3*** (8.2)	-17.7*** (6.0)	-0.9 (3.0)	-26.1*** (6.2)
	Panel D: Dropping Texas					
Net Lives Saved	-21.5*** (7.6)	33.7*** (4.4)	6.5 (7.9)	-41.6*** (5.6)	32.5*** (2.1)	-11.3* (5.9)
	Panel E: Dropping California					
Net Lives Saved	-26.1*** (7.0)	30.1*** (3.9)	33.3*** (6.5)	-28.7*** (4.9)	17.8*** (2.0)	9.6*** (4.8)

Notes (from Donohue and Wolfers): Panel C runs the regression as described by Dezhbakhsh, Rubin, and Shepherd, collapsing the partisanship variables into a single instrumental variable indicating the percentage of the Republican vote in the last presidential election (instead of six variables—one for each election); this specification then predicts that each execution will cost between one and fifty-four lives. Panels D and E show highly variable estimates when Texas and California are dropped. Population-weighted instrumental variables regressions are used. Endogenous independent variables are shown in panel A. Instruments include state-level police payroll, judicial expenditures, Republican vote shares, and prison admissions. Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; and state NRA membership. County and year fixed effects are used. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively. *(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996.

Table 5 Model Averaged Coefficients (first stage and data given)

Dependent Variable: *Annual Homicides per 100,000 Residents*						
	Panel A: Replication of Dezhbakhsh, Rubin and Shepherd Estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
Probability of Arrest	4.23 (8.81)	-1.08 (4.71)	5.55 (8.88)	5.58 6.86	2.17 (4.98)	6.76 (7.17)
Probability of Death Sentence Given Arrest	120.37 (131.58)	65.12 (47.96)	170.79 (137.76)	119.04 108.09	107.6 (38.75)	172.77 (123.2)
Probability of Execution Given Death Sentence	12.88 (18.78)	2.20 (7.61)	-17.47 (17.04)	-12.55 12.07	-4.05 (3.75)	-18.13 (13.30)
	Panel B: Implied Life-Life Tradeoff					
Net Lives Saved	-93.29 (131.87)	-14.79 (53.47)	126.18 (119.65)	90.96 (84.78)	30.05 (26.31)	130.92 (93.37)

Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership; and Ordinary Least Squares estimation.. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively. *(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996. Instrumental variables regressions are used. Endogenous independent variables are shown in panel A. Instruments include state-level police payroll, judicial expenditures, Republican vote shares, and prison admissions. Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership. The coefficients in this table are estimated by iterating over 12 key variables: assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; and state NRA membership.

Table 6 Model Averaged Coefficients (first and second stage given)

Dependent Variable: *Annual Homicides per 100,000 Residents*						
	Panel A: Replication of Dezhbakhsh, Rubin and Shepherd Estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
Probability of Arrest	1.31 (3.09)	-0.14 (2.91)	3.59 (4.15)	0.86 (4.07)	-0.050 (3.65)	1.73 (5.70)
Probability of Death Sentence Given Arrest	28.33 (16.88)	38.08 (14.33)	83.88 (31.17)	14.41 (26.06)	55.53 (28.4)	53.32 (59.08)
Probability of Execution Given Death Sentence	-4.31 (9.72)	-1.00 (5.00)	-11.81 (7.22)	-0.51 (13.31)	-2.85 (5.45)	-4.99 (16.66)
	Panel B: Implied Life-Life Tradeoff					
Net Lives Saved	31.89 (68.31)	8.19 (35.12)	85.62 (50.69)	4.63 (93.53)	21.39 (38.27)	36.78 (116.98)

Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership; and Ordinary Least Squares estimation. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively. *(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996. Instrumental variables regressions are used. Endogenous independent variables are shown in panel A. Instruments include state-level police payroll, judicial expenditures, Republican vote shares, and prison admissions. Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; and state NRA membership. The coefficients in this table are estimated by iterating over four combinations including the baseline in DRS, as well as the three variations in Table 4, above.

Table 7 Model Averaged Coefficients (first stage given, MA over model and data in second stage)

Dependent Variable: *Annual Homicides per 100,000 Residents*						
	Panel A: Replication of Dezhbakhsh, Rubin and Shepherd Estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
Probability of Arrest	1.707 10.099	-2.18 6.08	2.85 10.89	2.81 8.78	0.21 6.624	3.63 10.40
Probability of Death Sentence Given Arrest	112.78 126.82	68.42 50.28	160.43 163.02	106.58 110.2	88.18 63.13	153.82 156.77
Probability of Execution Given Death Sentence	-8.69 21.50	1.38 10.12	-14.23 23.62	-7.62 17.54	-1.99 7.51	-13.53 24.95
	Panel B: Implied Life-Life Tradeoff					
Net Lives Saved	63.28 (150.97)	-8.86 (71.11)	102.99 (165.88)	55.63 (123.20)	15.24 (52.72)	97.94 (175.19)

Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership; and Ordinary Least Squares estimation. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively. *(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996. Instrumental variables regressions are used. Endogenous independent variables are shown in panel A. Instruments include state-level police payroll, judicial expenditures, Republican vote shares, and prison admissions. Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership; and Ordinary Least Squares estimation. The coefficients in this table are estimated by iterating over 12 key variables and over four combinations of data. Data include the baseline in DRS, as well as the three variations in Table 4, above. Variables include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; and state NRA membership.

Table 8 Model Averaged Coefficients (full model)

Dependent Variable: *Annual Homicides per 100,000 Residents*						
	Panel A: Replication of Dezhbakhsh, Rubin and Shepherd Estimates					
	(1)	(2)	(3)	(4)	(5)	(6)
Probability of Arrest	1.30 (0.25)	1.30 (0.23)	1.30 (0.26)	6.15 (7.99)	2.90 (6.06)	6.37 (8.08)
Probability of Death Sentence Given Arrest	5.77 (37.35)	10.34 (36.16)	-1.90 (40.39)	136.98 (98.73)	101.40 (60.84)	173.44 (121.89)
Probability of Execution Given Death Sentence	-0.92 (7.48)	-2.57 (5.85)	0.20 (6.49)	-12.52 (17.78)	-2.29 (6.28)	-16.10 (20.19)
	Panel B: Implied Life-Life Tradeoff					
Net Lives Saved	7.60 (52.53)	19.41 (41.09)	-0.46 (45.57)	90.70 (124.85)	17.44 (44.08)	116.37 (141.75)

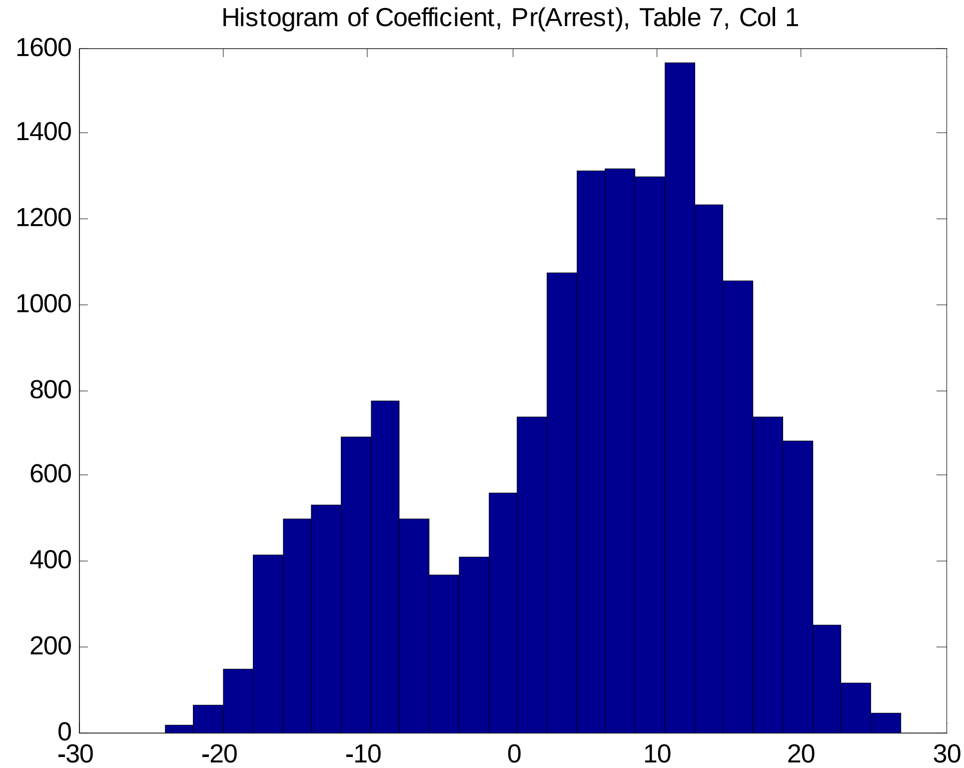
Controls include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; state NRA membership; and Ordinary Least Squares estimation. Standard errors are in parentheses, and ***, **, and * denote statistically significant at 1%, 5%, and 10%, respectively. *(a)* Implied life-life tradeoff reflects net lives saved evaluated for a state with the characteristics of the average death penalty state in 1996. The coefficients in this table are estimated by iterating over 12 key variables in the second stage, either 7 or twelve in the first stage and over four combinations of data. Data include the baseline in DRS, as well as the three variations in Table 4, above. Variables in the second stage include the assault rate; the robbery rate; real per capita personal income; real per capita unemployment insurance payments; real per capita income maintenance payments; population density; the proportion of the population aged 10-19, 20-29; black, white, or other; male or female; and state NRA membership. In the first stage, the averaging is over police expenditure, judicial expenditure, assault rate, robbery rate, state NRA membership, prison admissions, and either one or six voting variables.

Table 9: Most Likely Models

Rank	Variables												Data	Coefficient on Execution Variable
	AGA	ROB	1019	2029	Black	nonB	Male	NRA	PCI	IMP	UI	den		
1	0	X	0	0	X	0	X	X	X	X	X	X	VOTE	0.0377
2	X	0	X	0	0	0	X	0	X	X	X	X	VOTE	0.0283
3	X	0	X	X	X	X	X	X	X	0	X	0	VOTE	0.0361
4	X	X	X	X	X	X	X	X	X	0	X	0	VOTE	0.0286
5	0	X	X	X	X	X	0	X	X	0	X	0	VOTE	0.0368
6	X	X	X	X	X	X	0	X	X	0	0	X	VOTE	0.0312
7	X	X	X	0	X	0	X	X	X	X	0	X	VOTE	0.0311
8	0	X	X	X	X	X	X	X	X	0	X	X	VOTE	0.0397
9	X	X	0	X	0	X	X	X	X	X	X	0	VOTE	0.0278
10	X	0	X	X	X	X	X	X	X	0	0	X	VOTE	0.0288

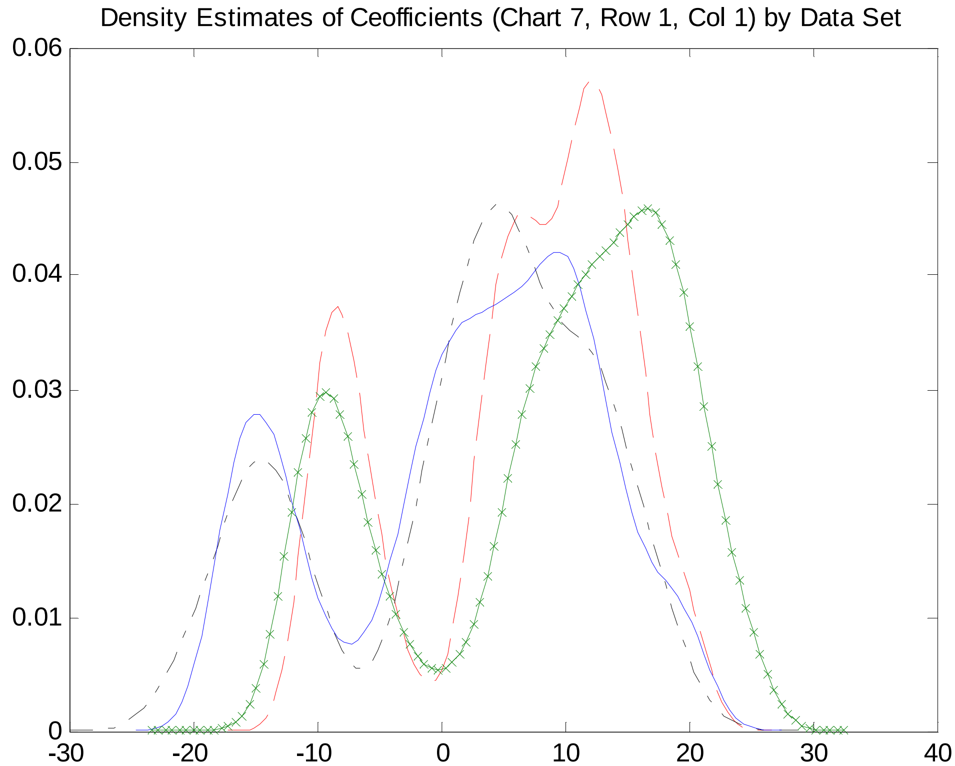
Notes: An X indicates that a variable was included in the given model, and a “0” that it was excluded. All regression run on with execution defined as in Column 4. Variables are: aggravated assault rate (AGA), the robbery rate(ROB), the population proportion of 10-19 year olds (1019), 20-29 year olds (2029), demographic percentages of blacks (black), percentage of non-black minorities (nonB), percentage of males (male), the percentage of NRA members (NRA) , real per capita income (PCI) , real per capita income maintenance payments (IMP), real per capita unemployment insurance payments (UI) , and the population density (den). There were four sets of data tested along with variations of the 12 variables. The data sets were the original Dezh Dezhbakhsh, Rubin and Shepherd data (DRS), a modification to allow for a single “voting” variable instead of 6 (VOTE), exclusion of Texas from the sample (exTX) and exclusion of California from the sample (exCA).

Figure 1



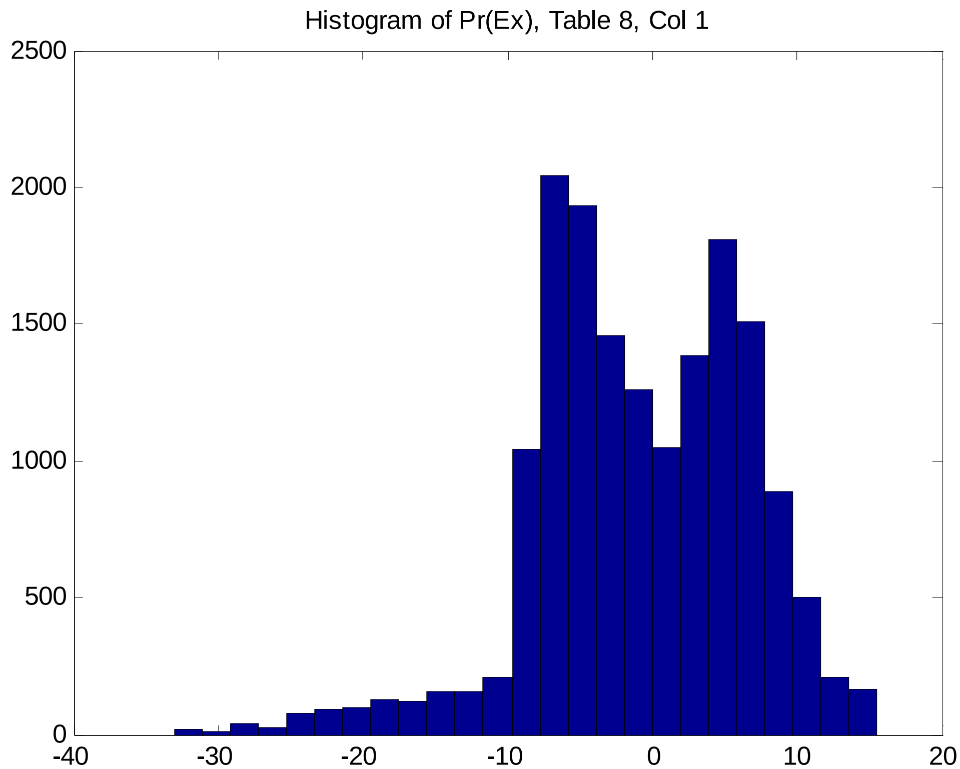
Note: Histogram of coefficients for probability of arrest variable in Table 7, row 1, column 1. Model averaging exercise produces 2^12*4 coefficients based on each possible model, data combination.

Figure 2



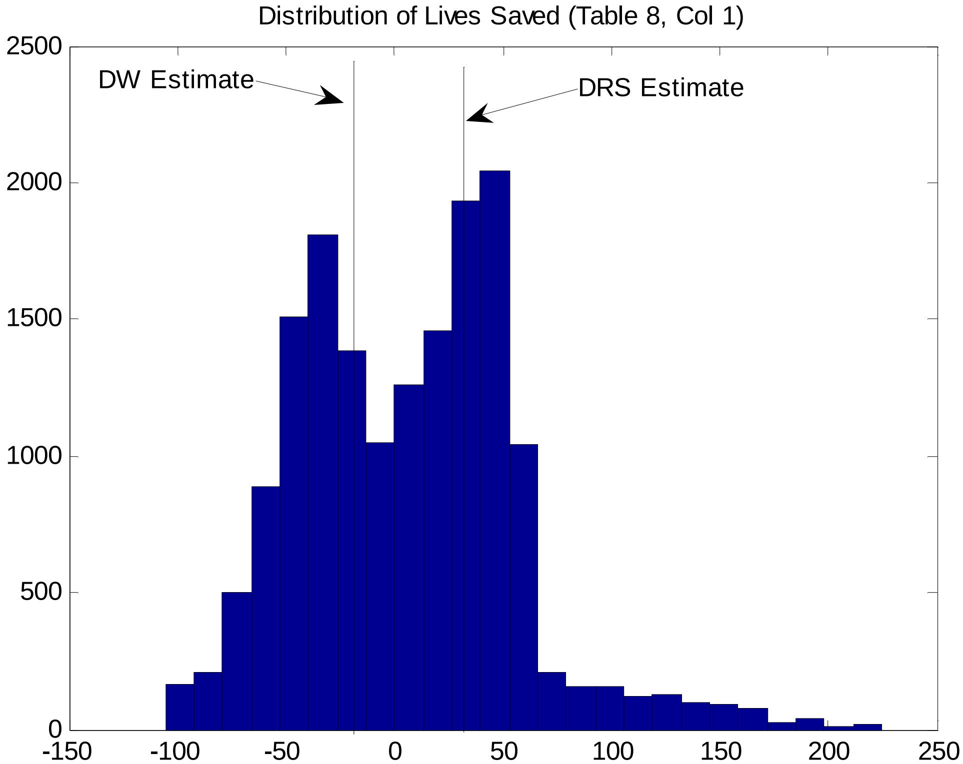
Note: Kernel density estimates for probability of arrest variable in Table 7, row 1, column 1. each of the four represent one of the "data sets" used in estimation and is based on 2^12 observations.

Figure 3



Note: Histogram of coefficients for probability of execution variable in Table 8, row 1, column 1. Model averaging exercise produces 2^12*4 coefficients based on each possible model, data combination.

Figure 4



Note: Histogram of estimated number of "lives saved" for probability of execution variable in Table 8, row 1, column 1. Model averaging exercise produces 2^12*4 coefficients based on each possible model, data combination. The DW estimate used is from the 'single voting variable' model variation.

Bibliography

Baldus, David and James Cole. (1975). "A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment." *Yale Law Journal*, 85, 170-184.

Bowers, William J. and Pierce, Glenn. (1975) "The Illusion of Deterrence in Isaac Ehrlich's Research on Capital Punishment." *Yale Law Journal*, 85, 187-208.

Bowers, William J. and Pierce, Glenn. (1980) "Deterrence or Brutalization? What Is the Effect of Executions?" 26 *Crime and Delinquency*, 453-484.

Brock, W. and S. Durlauf, (2001), "Growth Economics and Reality," *World Bank Economic Review*, 15, 229-272.

Brock, William A.; Durlauf, Steven N. and West, Kenneth D. (2003) "Policy Evaluation in Uncertain Economic Environments." *Brookings Papers on Economic Activity*, 2003(1), 235-322.

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

Chressanthis, George A. (1989) "Capital Punishment and the Deterrent Effect Revisited: Recent Time Series Econometric Evidence." *Journal of Behavioral Economics,* 18(2), 81-97.

Cloninger, Dale and Marchesini, Roberto. (2001) "Execution and Deterrence: A Quasi-Controlled Group Experiment." *Applied Economics*, 35(5), 569-76.

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

Cook, Philip; Molliconi, Stephanie and Cole, Thomas. (1995) "Regulating Gun Markets." *Journal of Criminal Law and Criminology*, 86(1), 59-92.

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

Dezhbakhsh, Hashem; Rubin, Paul H. and Shepard, Joanna M. (2003) "Does Capital Punishment Have a Deterrent Effect? New Evidence from Post-Moratorium Panel Data." *American Law and Economics Review*, 5(2), 344-76.

Dezhbakhsh, Hashem and Shepard, Joanna M. (2006) "The Deterrent Effect of Capital Punishment: Evidence from a 'Judicial Experiment'" *Economic Inquiry*, 44, 512.

Donohue, John J and Justin Wolfers, (2005), "Uses and Abuses of Empirical Evidence in the Death Penalty Debate." *Stanford Law Review* 791.

Donohue, John J and Levitt, Steven D. (2001), "The Impact of Legalized Abortion on Crime." *The Quarterly Journal of Economics*, CXVI(2), 379-420.

Draper, David, (1995), "Assessment and Propagation of Model Uncertainty," *Journal of the Royal Statistical Society, series B*, 57, 45-70.

Ehrlich, Isaac and Gibbons, Joel. (1977), "On the Measurement of the Deterrent Effect of Capital Punishment and the Theory of Deterrence." *Journal of Legal Studies*, 6(1), 35-50.

Ehrlich, Isaac and Liu, Zhiqiang. (1999), "Sensitivity Analyses of the Deterrence Hypothesis: Let's Keep the Econ in Econometrics." *Journal of Law and Economics*, XLII, 455-87.

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

Ehrlich, Isaac. (1977), "Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence." *Journal of Political Economy*, 85, 741-88.

Fagan, Jeffrey. (2006), “Death and Deterrence Redux: Science, Law and Causal Reasoning on Capital Punishment.” *Ohio State Journal of Criminal Law, forthcoming*

Fernandez, C., E. Ley, and M. Steel, (2001), “Model Uncertainty in Cross-Country Growth Regressions,” *Journal of Applied Econometrics*, 16, 563-576.

Grogger, Jeffrey. (1990) "The Deterrent Effect of Capital Punishment: An Analysis of Daily Homicide Counts." *Journal of the American Statistical Association*, 85(410), 295-303.

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

Hoeting, J., M. Clyde, D. Madigan and A. Raftery, (1999), “Bayesian Model Averaging: A Tutorial,” *Statistical Science*, 14, 382-401.

Katz, Lawrence; Levitt, Steven D and Shustorovich, Ellen.(2001) "Prison Conditions, Capital Punishment, and Deterrence." *American Law and Economics Review*, 5, 318-43.

Layson, Stephen. (1983) "Another View of the Canadian Time-Series Evidence on Homicide and Deterrence." *Canadian Journal of Economics*, 16(1), 52-73.

____. (1985) "Homicide and Deterrence: A Reexamination of the United States Time-Series Evidence." *Southern Economic Journal*, 52(1), 68-89.

Leamer, E., (1978), *Specification Searches*. New York: John Wiley and Sons.

Leamer, Edward E. (1983), "Let's Take the Con out of Econometrics." American Economic Review, 73, 31-43.

Liu, Zhiqiang. (2004), "Capital Punishment and the Deterrence Hypothesis: Some New Insights and Empirical Evidence," *Eastern Economic Journal* 30: 237.

Masanjala, W. and C. Papageorgiou, (2004), "Rough and Lonely Road to Prosperity: A Reexamination of the Sources of Growth in Africa Using Bayesian Model Averaging," mimeo, Louisiana State University.

McAleer, Michael and Veall, Michael R. (1989), "How Fragile Are Fragile Inferences? A Re-Evaluation of the Deterrent of Capital Punishment." *Review of Economics and Statistics*, 71, 99-106.

McManus, Walter S. (1985), "Estimates of the Deterrent Effect of Capital Punishment: The Importance of the Researcher's Prior Beliefs." *Journal of Political Economy*, 93 (2), 417-25.

Mocan, Naci and Gittings, R Kaj. (2001), "Getting Off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment." *Journal of Law and Economics*, 46(2), 453-78.

Passell, Peter and Taylor, John B. (1977), "The Deterrent Effect of Capital Punishment: Another View." *American Economic Review*, 85, 445-58.

Raftery, A., (1995) "Bayesian Model Selection in Social Research (with discussion)," in *Sociological Methodology 1995*, P. Marsden ed., (Cambridge, MA: Blackwell)

Raftery, A. (1996). "Approximate Bayes factors and Accounting from Model Uncertainty in Generalised Linear Models." Biometrika 83: 251–266.

Raftery, A., D. Madigan, and J. Hoeting, (1997), "Bayesian Model Averaging for Linear Regression Models," *Journal of the American Statistical Association* 92, 437, 179-91.

Raftery, Adrian E. ; Madigan, David and Hoeting, Jennifer A. (1997), "Bayesian Model Averaging for Linear Regression Models." *Journal of the American Statistical Association*, 92, 179-91.

Sala-i-Martin, X., G. Doppelhofer, and R. Miller, (2004), "Determinants of Long-Term Growth: A Bayesian Averaging of Classical Estimates (BACE) Approach," *American Economic Review*, 94(4) 813-835.

Shepherd, Joanna M. 2005. "Deterrence versus Brutalization: Capital Punishment's Differing Impact Among States." *Michigan Law Review* 104: 203.

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

Sorenson, Jon; Wrinkle, Robert and Brewer, Victoria, (1999), "Capital Punishment and Deterrence: Examining the Effect of Executions on Murder in Texas." *Crime and Delinquency*, 45(4), 481-93.

Steiker, Carol S. (2005), "No, Capital Punishment Is Not Morally Required: Deterrence, Deontology, and the Death Penalty," *Stanford Law Review*. 58, 751.

Sunstein, Cass R. and Adrian Vermeule, (2005), "Is Capital Punishment Morally Required? Acts, Omissions, and Life-Life Tradeoffs" *Stanford Law Review* 58, 703-706.

Wasserman, L. (2000), "Bayesian Model Selection and Bayesian Model Averaging," *Journal of Mathematical Psychology* 44, 97-112.

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

Zimmerman, Paul, (2004), "State Executions, Deterrence, and the Incidence of Murder," *Journal of Applied Economics* 7(1), 163-193.

Zimmerman, Paul, (2006), "Estimates of the Deterrent Effect of Alternative Execution Methods in the United States: 1978-2000," *American Journal of Economics and Sociology* forthcoming.