

Understanding Aggregate Crime Regressions

Steven N. Durlauf

Salvador Navarro

David A. Rivers

University of Wisconsin at Madison

revised

February 6, 2009

Abstract

This paper provides a general description of the relationship between individual decision problems and aggregate crime regressions. The analysis is designed to elucidate the behavioral and statistical assumptions that are implicit in the use of aggregate crime regressions for both the analysis of crime determinants as well in counterfactual policy evaluation. We apply our general arguments to the question of the deterrent effect of capital punishment and show how alternative assumptions affect estimates of the deterrent effect.

This paper was prepared for a festschrift in honor of Phoebus Dhrymes. We thank the National Science Foundation and University of Wisconsin Graduate School for financial support. We thank two referees for extremely helpful suggestions and Nonarit Bisonyabut and Xiangrong Yu for outstanding research assistance.

1. Introduction

While Phoebus Dhrymes is best known as an econometric theorist of the highest creativity and rigor, he has also been an important contributor to applied economics, in particular the areas of productivity growth.¹ This paper, while focusing on one of the few areas outside of Dhrymes' direct research interests - crime - nevertheless follows in his tradition of exploring the meaning of econometric exercises when one carefully examines the assumptions underlying a given analysis. Our analysis is partially inspired by Dhrymes (1985) remarkable critique of claims made by supply-side economists; like him, we hope to clarify what can and cannot be claimed when moving from data analysis to policy.

Despite recent efforts to employ microeconomic data and natural experiments, aggregate crime regressions continue to play a significant role in criminological analyses. One aspect is predictive, as illustrated by the literature that attempts to link crime to unemployment. A second aspect focuses on policy evaluation: prominent current controversies involving the use of aggregate regressions include the deterrent effect of shall-issue concealed weapons legislation (Lott and Mustard (1997), Lott (1998), Black and Nagin (1998), Ayers and Donohue (2003), Plassmann and Whitley (2003)), the deterrent effect of capital punishment (Dezhbakhsh and Rubin (2007), Dezhbakhsh, Rubin, and Shepherd, (2003), Donohue and Wolfers (2005), Katz, Levitt and Shustorovich (2001), Mocan and Gittings (2001,2006)) and perhaps most controversially, the effects of liberalized abortion laws on crime (Donohue and Levitt (2001,2004,2008), Foote and Goetz (2008), Joyce (2004) and Lott and Whitley (2007)).

The goal of this paper is to examine the construction and interpretation of these regressions. Specifically, we wish to employ aspects of contemporary economic and econometric reasoning to understand how aggregate crime regressions may be appropriately used to inform positive and normative questions. While by no means exhaustive of the relevant issues, we hope our discussion will prove useful in highlighting some of the limitations of the use of these regressions and in particular indicate how empirical findings may be overinterpreted when careful attention is not given to the link

¹On productivity, Dhrymes (1963) is one of his earliest papers while Dhrymes and Bartlesman (1998) is an important more recent contribution.

between the aggregate data and individual behavior.² Much of our discussion will involve arguments that are well understood from the vantage point of contemporary microeconometrics. While our objective is to provide a connection between current microeconomic thinking and the empirical crime literature in economics as well as other social sciences, the points we make apply to other applied areas in which aggregate regressions are used to test individual level hypotheses. In spirit, we follow Heckman (2000,2005) in trying to clarify how assumptions determine the nature of substantive empirical claims.

The paper is organized as follows. In Section 2, we describe a standard choice-based model of crime. Section 3 discusses how this individual-level model can be aggregated to produce crime regressions of the type found in the literature. Section 4 discusses the analysis of policy counterfactuals. Section 5 discusses these general arguments in the context of a prominent paper in the capital punishment and deterrence literature, Dezhbakhsh, Rubin, and Shepherd (2003). The empirical importance of some of the assumptions is demonstrated in Section 6. Section 7 concludes.

2. Crime as a choice

From the vantage point of economic reasoning, the fundamental idea underlying the analysis of crime is that each criminal act constitutes a purposeful choice on the part of the criminal. In turn, this means that the development of a theory of the aggregate crime rate should be explicitly understood as deriving from the aggregation of individual decisions. The basic logic of the economic approach to crime was originally developed by Gary Becker (1968) and extended by Isaac Ehrlich (1972,1973,1975,1977); for economists, Ehrlich's work represents the classic use of aggregate regressions to understand crime, in his context, capital punishment.

The choice-theoretic conception does not, by itself, have any implications for the process by which agents make these decisions, although certain behavioral restrictions are

²The interpretation of aggregate data continues to be one of the most difficult questions in social science; Stoker (1993) and Blundell and Stoker (2005) provide valuable overviews.

standard for economists. For example, to say that the choice of a crime is purposeful says nothing about how an individual assesses the various probabilities that are relevant to the choice, such as the conditional probability of being caught given that the crime is committed. At the same time, the economic analyses typically assume that an individual's subjective beliefs, i.e. the probabilities that inform his decision, are rational in the sense that they correspond to the probabilities generated by the optimal use of an individual's available information. While the relaxation of this notion of rationality has been a major theme in recent economic research, we will use a relatively standard notion of rationality in our discussion. This modeling choice is made both because we regard it as an appropriate baseline for describing individual behavior and for simplicity of exposition. But we emphasize that the choice-based approach does not require rationality as we model it.

To see how crime choice may be formally described, we follow the canonical binary choice model of economics. Consider the decision problem of a population of individuals indexed by i each of whom decides at each period t whether or not to commit a crime. Individuals live in locations l and it is assumed that a person only commits crimes within the location in which he lives. Suppose the choice is coded as $\omega_{i,t} = 1$ if a crime is committed, 0 otherwise. The standard form for the expected utility associated with the choice $u_{i,t}(\omega_{i,t})$ is

$$u_{i,t}(\omega_{i,t}) = Z_{l,t}\beta_l\omega_{i,t} + X_{i,t}\gamma_l\omega_{i,t} + \xi_{l,t}(\omega_{i,t}) + \varepsilon_{i,t}(\omega_{i,t}). \quad (1)$$

In this expression, $Z_{l,t}$ denotes a set of observable location-specific characteristics and $X_{i,t}$ denotes a vector of observable individual-specific characteristics. In contrast, $\xi_{l,t}(\omega_{i,t})$ and $\varepsilon_{i,t}(\omega_{i,t})$ denote unobservable location-specific and individual-specific utility terms; each will naturally be a function of unobserved location-specific and individual-specific characteristics. The distinction between observable and unobservable variables is made with respect to the modeler; the individual observes all the variables we have described. These unobservable terms represent sources of heterogeneity that are unexplained by the model. Defining the net expected utility of crime commission as

$$v_{i,t} = Z_{i,t}\beta_i + X_{i,t}\gamma_i + \xi_{i,t}(1) - \xi_{i,t}(0) + \varepsilon_{i,t}(1) - \varepsilon_{i,t}(0) \quad (2)$$

the choice-based perspective amounts to saying that a person chooses to commit a crime, i.e. $\omega_{i,t} = 1$ if and only if $v_{i,t} > 0$. The assumption of linearity of the utility function is standard in binary choice analysis. It is possible to consider nonparametric forms of the utility function, see Matzkin (1992). We focus on the linear case both because it is the empirical standard in much of social science and because it is not clear that more general forms will be particularly informative for the issues we wish to address. Some forms of nonlinearity may be trivially introduced, such as including the products of elements of any initial choice of $X_{i,t}$ as additional observables.

Eq. (1) and its equivalent (2) represent the substantive assumptions of the economic theory of crime. In order to make the theory operational, it is of course necessary to make statistical assumptions. First, we restrict the nature of the unobservable heterogeneity with four assumptions.

$$\text{A.1} \quad E_i(\varepsilon_{i,t}(1) - \varepsilon_{i,t}(0)) = 0 \quad (3)$$

$$\text{A.2} \quad \varepsilon_{i,t}(1) - \varepsilon_{i,t}(0) \text{ independent of } \xi_{i,t}(1) - \xi_{i,t}(0) \quad (4)$$

$$\text{A.3} \quad \varepsilon_{i,t}(1) - \varepsilon_{i,t}(0) \text{ independent of } X_{i,t}, Z_{i,t} \quad (5)$$

$$\text{A.4} \quad \beta_i = \beta_j; \gamma_i = \gamma_j \quad \forall i, j \quad (6)$$

The first three assumptions relate to the model errors. Assumption A.1, that the differences in individual utility errors have 0 mean is without loss of generality so long as either $X_{i,t}$ or $Z_{i,t}$ includes a constant term. Assumption A.2 allows us to consider the two types of unobservables separately. Assumption A.3 corresponds to the assumption of the orthogonality of regressors and errors in the linear model. Assumptions A.2 and A.3 are

sufficient rather than necessary; relaxation is not of interest for the issues we address. Assumption A.4, which imposes parameter homogeneity, is very standard in empirical work, even though it is not an implication of the choice-based approach. When the parameters relate to policy variables, this helps rule out treatment effect heterogeneity, something that may be quite important; see Abbring and Heckman (2007) for a comprehensive survey.

Under our utility specification, it is immediate that a positive net utility from commission of a crime requires that $X_{i,t}\gamma + Z_{i,t}\beta + \xi_{i,t}(1) - \xi_{i,t}(0) > \varepsilon_{i,t}(0) - \varepsilon_{i,t}(1)$. Conditional on $X_{i,t}, Z_{i,t}$, and $\xi_{i,t}(1) - \xi_{i,t}(0)$, the individual choices are stochastic, with the distribution function of $\varepsilon_{i,t}(0) - \varepsilon_{i,t}(1)$, which we denote by $G_{i,t}$, determining the probability that a crime is committed.³ Formally,

$$\Pr(\omega_{i,t} = 1 | Z_{i,t}, X_{i,t}, \xi_{i,t}(1) - \xi_{i,t}(0)) = G_{i,t}(Z_{i,t}\beta + X_{i,t}\gamma + \xi_{i,t}(1) - \xi_{i,t}(0)). \quad (7)$$

This conditional probability structure captures the basic microfoundations of the economic model of crime. It is worth noting that this formulation represents a relatively elementary behavioral model in that we ignore issues such as 1) selection into and out of the population generated by the dynamics of incarceration and 2) those aspects of a crime decision at t in which a choice is a single component in a sequence of decisions which collectively determine an individual's utility, i.e. a more general preference specification is one in which agents make decisions to maximize a weighted average of current and future utility, accounting for the intertemporal effects of their decisions each period. While the introduction of dynamic considerations into the choice problem raises numerous issues, e.g. state-dependence, heterogeneity and dynamic selection, these can be readily dealt with

³We allow the distribution function $G_{i,t}$ to vary across individuals (and time) in order to accommodate the linear probability model described below. Of course, there is no behavioral reason why $G_{i,t}$ needs to be constant across individuals or across time; outside the linear probability model, the assumption that it is constant is nevertheless standard.

using the analysis of Heckman and Navarro (2007), albeit at the expense of considerable complication of the analysis.

3. Aggregation

How do the conditional crime probabilities for individuals described by (7) aggregate within a location? Let $\rho_{l,t}$ denote the realized crime rate in locality l at time t . Notice that we define the crime rate as the percentage of individuals committing crimes, not the number of crimes per se, so we are ignoring multiple acts by a single criminal. Given our assumptions, for the location-specific choice model (7), if individuals are constrained to commit crimes in the location of residence, then the aggregate crime rate in a locality is determined by integrating over the observable individual-specific heterogeneity in the location's population. Letting $F_{X_{i,t}}$ denote the empirical distribution function of $X_{i,t}$ within l , the expected crime rate in a location at a given time is

$$E\left(\rho_{l,t} \mid Z_{l,t}, F_{X_{i,t}}, \xi_{l,t}(1) - \xi_{l,t}(0)\right) = \int G_{i,t} \left(Z_{l,t} \beta + X \gamma + \xi_{l,t}(1) - \xi_{l,t}(0) \right) dF_{X_{i,t}}. \quad (8)$$

In order to convert this aggregate relationship to a linear regression form, it is necessary to further restrict the distribution function $G_{i,t}$:

$$\text{A.5} \quad dG_{i,t} \text{ is a uniform density.} \quad (9)$$

Notice that this assumption does not imply that $G_{i,t}$ is constant. Applying A.5 to (8), the crime rate in locality l obeys

$$\rho_{l,t} = Z_{l,t} \beta + \bar{X}_{l,t} \gamma + \xi_{l,t}(1) - \xi_{l,t}(0) + \theta_{l,t}, \quad (10)$$

where $\bar{X}_{l,t}$ is the empirical mean of $X_{i,t}$ within l and $\theta_{l,t} = \rho_{l,t} - E\left(\rho_{l,t} \mid Z_{l,t}, F_{X_{l,t}}, \xi_{l,t}(1) - \xi_{l,t}(0)\right)$ captures the difference between the realized and expected crime rate within a locality. This is the basic statistical model employed in aggregate crime regressions. As is standard in any exercise of this type (cf. Heckman (2000,2005)) the transition from eq. (1) to eq. (10) is done *with* loss of generality with respect to individual behavior, i.e. assumptions A.1-A.7 all restrict individual behavior, the extent to which this renders analyses based on (10) noncredible is a distinct issue.

In our subsequent discussion, we will assume that all location-specific variables are exactly measured; in other words, we assume that the entire set of individuals in each locality is observed. This assumption ignores sampling issues. The crime literature has in fact spent considerable time exploring sampling questions with respect to major data sets such as the Uniform Crime Reports (UCR) and National Crime Victims Survey (NCVS); see Lynch and Addington (2007). These arise in the UCR, for example, because of differences in reporting practices across police departments. While these issues are important, they are not germane to the questions we wish to explore.

Our construction of eq. (10) from choice-based foundations provides a basis for understanding when standard aggregate crime regressions may be interpreted as aggregations of individual behavior. We argue that there are at least three aspects of this construction that indicate limitations in the interpretability of standard aggregate crime regressions.

i. linear probability model

The assumption of the linear probability model is of concern since it implicitly restricts the random utility terms in ways that are well known to be odd; specifically, in order for the individual choice probabilities to always lie in the interval $[0,1]$, it is necessary for the support of the random utility terms to depend on the deterministic utility components. This sensitivity of the support to the deterministic utility component is in fact inconsistent with Assumptions A.2 and A.3. It might be possible to construct random utility terms which preserve weaker versions of the assumptions, e.g. median

independence, but such cases are very special. Unfortunately, as is well known, other random utility specifications do not aggregate in a straightforward manner. To illustrate the problem, note that if one assumes that $\varepsilon_{i,t}(\omega_{i,t})$ has a type-I extreme value distribution, which is the implicit assumption in the logit binary choice model, then $\log\left(\frac{\Pr_{i,t}(\omega_{i,t} = 1|Z_{l,t}, X_{i,t}, \xi_{l,t}(1) - \xi_{l,t}(0))}{1 - \Pr_{i,t}(\omega_{i,t} = 1|Z_{l,t}, X_{i,t}, \xi_{l,t}(1) - \xi_{l,t}(0))}\right)$ will be linear in the various payoff components, but will not produce a closed form solution for the aggregate crime rate.

ii. instrumental variables

On its own terms, the derivation of the linear aggregate crime model from individual choices indicates how aggregation affects the consistency of particular estimators and by implication affects how instrumental variables ought to be employed. The assumption we impose on the relationship between observables and unobservables, A.3, requires that individual unobservables are independent of the observables. The assumption does not require that the location-specific unobservables $\xi_{l,t}(\omega_{i,t})$ are independent of the aggregate observables that appear in the utility function $Z_{l,t}$ or those variables that appear as a consequence of aggregation $\bar{X}_{l,t}$. There is no a priori reason why the regression residual $\xi_{l,t}(1) - \xi_{l,t}(0) + \theta_{l,t}$ should be orthogonal to any of the regressors in (10). This means that all the variables in (10) should potentially be instrumented. Hence in our judgment the focus on instrumenting the endogenous regressors that one finds in empirical crime analyses is often insufficient in that, while this strategy addresses endogeneity, it does not address unobserved location-specific heterogeneity. Notice that if individual-level data were available, this problem would not arise since one would normally allow for location-specific, time-specific and location-time-specific fixed effects for a panel.

The need for instruments does not, of course, imply that valid instruments are available. In general, there do not exist good reasons to believe that lagged values of $Z_{l,t}$ and $\bar{X}_{l,t}$ are orthogonal to $\xi_{l,t}(1) - \xi_{l,t}(0) + \theta_{l,t}$ since a researcher will usually not have a

theory of the determinants of $\xi_{l,t}(1) - \xi_{l,t}(0) + \theta_{l,t}$. The reason for this is that the variables which appear in $Z_{l,t}$ and $\bar{X}_{l,t}$ are generally based on reasonable guesses about the sources of individual-specific and location-specific heterogeneity; as such, their relationship to unobserved heterogeneity is indeterminate. Put differently, the presence of one of these variables in a crime model does not usually rule out any other variable and so does not rule out those determinants that have been omitted from the specification. Brock and Durlauf (2001) describe this problem as theory openendness in arguing that many candidate instrumental variables in the economic growth literature are invalid in that they correlate with the associated regression errors. The same problems arise in crime contexts.

iii. parameter heterogeneity

The linear probability assumption matters for a third reason, which relates to the significance of Assumption A.4, i.e. parameter homogeneity. If this assumption is relaxed, the crime rate within locality l will equal

$$E\left(\rho_{l,t} \mid Z_{l,t}, F_{X_{l,t}}, \xi_{l,t}(1) - \xi_{l,t}(0)\right) = \iiint G_{i,t}\left(Z_{l,t}\beta + X\gamma + \xi_{l,t}(1) - \xi_{l,t}(0)\right) dF_{X_{l,t}} dF_{\beta} dF_{\gamma} \quad (11)$$

where F_{β} and F_{γ} are the distribution functions for these parameters when calculated across i . Under Assumption A.5, this heterogeneity has no qualitative effect, as (11) holds when evaluated at the mean values of β and γ . For other probability structures, this will not be the case. As suggested above, once one introduces a role for moments other than the mean of the policy-specific parameters to affect the aggregate, this introduces new complications in evaluating the aggregate effects of policy changes.

Methods are available to allow for analysis of aggregate data which account for parameter heterogeneity; Berry, Levinsohn, and Pakes (1995) is a seminal contribution, but has not been previously applied in crime contexts.

4. Policy effect evaluation

How can an aggregate crime regression be used to evaluate a change in policy? Given our choice-theoretic framework, a policy evaluation may be understood as a comparison of choices under alternative policy regimes A and B . The net utility to the commission of a crime will depend on the regime, so that

$$v_{i,t}^A = Z_{l,t}^A \beta^A + X_{i,t}^A \gamma^A + \xi_{l,t}^A(1) - \xi_{l,t}^A(0) + \varepsilon_{i,t}^A(1) - \varepsilon_{i,t}^A(0) \quad (12)$$

and

$$v_{i,t}^B = Z_{l,t}^B \beta^B + X_{i,t}^B \gamma^B + \xi_{l,t}^B(1) - \xi_{l,t}^B(0) + \varepsilon_{i,t}^B(1) - \varepsilon_{i,t}^B(0) \quad (13)$$

respectively. The net utility to individual i of committing a crime equals

$$\begin{aligned} v_{i,t} = & \\ & Z_{l,t}^A \beta^A + X_{i,t}^A \gamma^A + \xi_{l,t}^A(1) - \xi_{l,t}^A(0) + \varepsilon_{i,t}^A(1) - \varepsilon_{i,t}^A(0) + \\ & D_{l,t} \left(Z_{l,t}^B \beta^B - Z_{l,t}^A \beta^A \right) + D_{l,t} \left(X_{i,t}^B \gamma^B - X_{i,t}^A \gamma^A \right) + \\ & D_{l,t} \left(\xi_{l,t}^B(1) - \xi_{l,t}^B(0) - \left(\xi_{l,t}^A(1) - \xi_{l,t}^A(0) \right) \right) \\ & D_{l,t} \left(\varepsilon_{i,t}^B(1) - \varepsilon_{i,t}^B(0) - \left(\varepsilon_{i,t}^A(1) - \varepsilon_{i,t}^A(0) \right) \right) \end{aligned} \quad (14)$$

where $D_{l,t} = 1$ if regime B applies to locality l at t ; 0 otherwise. The analogous linear aggregate crime rate model is

$$\begin{aligned} \rho_{l,t} = & \\ & Z_{l,t}^A \beta^A + \bar{X}_{l,t}^A \gamma^A + D_{l,t} \left(Z_{l,t}^B \beta^B - Z_{l,t}^A \beta^A \right) + D_{l,t} \left(\bar{X}_{l,t}^B \gamma^B - \bar{X}_{l,t}^A \gamma^A \right) + \\ & \xi_{l,t}^A(1) - \xi_{l,t}^A(0) + \theta_{l,t}^A + D_{l,t} \left(\xi_{l,t}^B(1) - \xi_{l,t}^B(0) - \left(\xi_{l,t}^A(1) - \xi_{l,t}^A(0) \right) \right) + \theta_{l,t}^B - \theta_{l,t}^A \end{aligned} \quad (15)$$

The standard approach measuring how different policies affect the crime rate, in this case regimes A versus B, is to embody the policy change in $Z_{l,t}^A$ versus $Z_{l,t}^B$ and to assume that all model parameters are constant across regimes, i.e.

$$\text{A.6} \quad \beta^A = \beta^B = \beta; \gamma^A = \gamma^B = \gamma \quad (16)$$

This allows the policy effect to be measured by $(Z_{l,t}^B - Z_{l,t}^A)\beta$. In addition, it is necessary that

$$\text{A.7} \quad \xi_{l,t}^B(1) - \xi_{l,t}^B(0) - (\xi_{l,t}^A(1) - \xi_{l,t}^A(0)) = 0 \quad (17)$$

i.e. that the change of regime does not change the location-specific unobserved utility differential between committing a crime and not doing so. This requirement seems problematic as it means that the researcher must be willing to assume that the regime change is fully measured by the changes in $\bar{X}_{l,t}$ and $Z_{l,t}$. Changes in the detection probabilities and penalties for crimes typically come in bundles and not all aspects of these are measured (or measurable) and so A.7 seems unlikely to hold in practice even if A.6 holds. We will argue below that there are cases, specifically capital punishment, where this does not receive adequate attention in the relevant empirical formulations.

5. Application to capital punishment: theory

In this section, we describe how some of our arguments matter by considering their implications for the study of capital punishment. For this discussion we focus on the empirical study of deterrent effects by Dezhbakhsh, Rubin, and Shepherd (2003); we denote this paper as DRS. We choose this paper both because it has been quite influential in resurrecting the capital punishment/deterrence literature and because it has recently

come under criticism by Donohue and Wolfers (2005), denoted as DW. Our analysis is designed to see to what extent the specific issues we have raised affect deterrence inferences. In doing so, we do not mean to suggest that other issues are not important. For example DW argue that the DRS results are driven by California and Texas; an assessment of their argument would require an investigation of the appropriateness of exchangeability assumptions for state-level crime data, which is beyond the scope of this paper. Nor do we address potential criticisms of both DRS and DW, for example, both papers ignore possible spatial dependence in errors.⁴

The behavioral foundations of DRS recognize that the consequences for the commission of a murder involve three separate stages: apprehension, sentencing and carrying out of the sentence. Defining the variables C = caught, S = sentenced to be executed and E = executed, DRS estimate regressions of the form

$$\rho_{l,t} = \alpha + \bar{X}_{l,t}\gamma + Z_{l,t}\beta + P_{l,t}(C)\beta_C + P_{l,t}(S|C)\beta_S + P_{l,t}(E|S)\beta_E + \kappa_{l,t} \quad (18)$$

where

$P_{l,t}(C)$ = probability of being caught conditional on committing a murder,

$P_{l,t}(S|C)$ = probability of being sentenced to be executed conditional on being caught,

$P_{l,t}(E|S)$ = probability of being executed conditional on receiving a death sentence.

Here $Z_{l,t}$ denotes location-specific variables other than those associated with the death penalty and $\kappa_{l,t} = \xi_{l,t}(1) - \xi_{l,t}(0) + \theta_{l,t}$ is the composite regression error.

i. microfoundations

⁴We thank an anonymous referee for this observation.

From the perspective of our first argument, that aggregate models should flow from individual behavioral problems, the DRS specification can be shown to be flawed. Specifically, the way in which probabilities are used does not correspond to the probabilities that arise in the appropriate decision problem. For an individual who commits a murder, the potential outcomes are NC = not caught, CNS = caught and not sentenced to death, $CSNE$ = caught, sentenced to death, and not executed, and CSE = caught, sentenced to death and executed. The expected utility is therefore

$$\begin{aligned} & \Pr_{i,t}(NC)u_{i,t}(NC) + \Pr_{i,t}(CNS)u_{i,t}(CNS) + \\ & \Pr_{i,t}(CSNE)u_{i,t}(CSNE) + \Pr_{i,t}(CSE)u_{i,t}(CSE). \end{aligned} \quad (19)$$

The unconditional probabilities of the four possible outcomes are of course related to the conditional probabilities used in DRS. In terms of conditional probabilities, expected utility may be written as

$$\begin{aligned} & (1 - \Pr_{i,t}(C))u_{i,t}(NC) + \\ & (1 - \Pr_{i,t}(S|C))\Pr_{i,t}(C)u_{i,t}(CNS) + \\ & (1 - \Pr_{i,t}(E|S))\Pr_{i,t}(S|C)\Pr_{i,t}(C)u_{i,t}(CSNE) + \\ & \Pr_{i,t}(E|S)\Pr_{i,t}(S|C)\Pr_{i,t}(C)u_{i,t}(CSE) \end{aligned} \quad (20)$$

Assume that the utility difference between the potential outcomes is constant across individuals, i.e.

$$\begin{aligned} \beta_S &= u_i(CSNE) - u_i(CNS) \\ \beta_C &= u_i(CNS) - u_i(NC) \\ \beta_E &= u_i(CSE) - u_i(CSNE) \end{aligned} \quad (21)$$

Then, replacement of the relevant terms in (18) yields

$$\begin{aligned} \rho_{l,t} = & \alpha + \bar{X}_{l,t}\gamma + Z_{l,t}\beta + P_{l,t}(C)\beta_C + P_{l,t}(S|C)P_t(C)\beta_S + \\ & P_{l,t}(E|S)P_{l,t}(S|C)P_t(C)\beta_E + \kappa_{l,t}. \end{aligned} \tag{22}$$

Comparison of (22) with (18) reveals that the DRS specification does not derive naturally from individual choices. Their analysis considers how different conditional probabilities affect behavior, without accounting for how these probabilities interact from the perspective of expected utility analysis. Put differently, the effect of the conditional probability of execution given a death sentence on behavior cannot be understood separately from the effects of the conditional probability of being caught and being sentenced to death if caught.

An additional benefit of deriving the estimating equation from individual choices is that it allows the researcher to accurately analyze the effects of potential policies. Consider a policy which aims to reduce the murder rate by increasing the probability of arrest. By looking at eq. (22) it is easy to see that this will affect the murder rate through three different channels: the probability of being arrested; the joint probability of being arrested and sentenced to death; and the joint probability of being arrested, sentenced to death, and executed. These channels show up because a prerequisite for reaching the stages at which being sentenced to death and being executed are possible is to be arrested in the first place. In the misspecified model used by DRS the latter two channels are ruled out. When analyzing the effectiveness of a potential policy, having a model which is derived from the individual behavior problem allows for both an accurate evaluation of the policy and a clear understanding of the causal mechanisms through which the policy operates.

ii. linear probability model

The DRS model assumes an underlying linear probability model for individual choices, which we have argued is unappealing. This potential misspecification is of particular importance given the use of the net lives saved measure. The reason for this is that the net lives saved measure depends sensitively on how one formulates the probability structure for the individual model errors. Executions are low probability events and involve outcomes associated with draws from the tail of the distribution of the random

payoffs. While in practice the linear probability model and the logit or probit models are often very similar, this is only true when events in the tails are unimportant. Put differently, the linear probability model implies that the effect of a change in a given variable on an outcome probability is a constant that is independent of how probable the event is. Suppose that the effect of changing a variable (say the probability of being executed) by 1% is estimated to cause an increase in the probability of a murder by -0.05. Then, if the probability of committing a crime is 0.05 initially, the linear probability model would predict the crime probability is reduced to zero. A logit probability model would never make such a prediction.

iii. instrumental variables

Our second argument concerned the transition from an individual-level specification to an aggregate specification. Aggregation, as we argued, can induce correlations between all of the aggregated regressors and model errors because of unobserved location characteristics. This implies that all regressors should be instrumented in aggregate crime regressions. DRS only instrument the conditional crime probabilities in (18), doing so on the basis that these probabilities are collective choice variables by the localities. However, in the presence of unobserved location characteristics, it is necessary to instrument the regressors contained in $Z_{l,t}$ as well. Since instrumenting a subset of the variables in a regression that correlate with the regression errors does not ensure consistency of the associated subset of parameters, the estimates in DRS would appear to be inconsistent given the logic of their exercise.

DRS might respond to this objection by noting that they use location-specific fixed effects. However, these will not be sufficient to solve the problem, since the location-specific unobservables $\xi_{l,t}(\omega_{i,t})$ can (and very likely do) vary over time. There is no reason to believe that non-capital penalties vary any less than capital penalties. This involves the question of the range of penalties.

6. Application to capital punishment: reestimation

In this section we address each of these theoretical objections to assess their empirical salience. We report regression results based on the preferred specification employed by Dezhbakhsh, Rubin and Shepherd (2003).⁵ DRS assume that the aggregate murder rate can be explained by a linear and separable function of various controls. DRS estimate the regression using two-stage least squares to control for potential endogeneity with respect to the deterrence measures (the terms with the β coefficients in the equation below). The exact specification is

$$\begin{aligned}
 & \frac{Murders_{c,s,t}}{Population_{c,s,t} / 100000} = \\
 & \beta_1 \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} + \beta_2 \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} + \beta_3 \frac{Executions_{s,t}}{DeathSentences_{s,t-6}} + \\
 & \gamma_1 \frac{Assaults_{c,s,t}}{Population_{c,s,t}} + \gamma_2 \frac{Robberies_{c,s,t}}{Population_{c,s,t}} + \gamma_3 CountyDemographics_{c,s,t} + \quad (23) \\
 & \gamma_4 CountyEconomy_{c,s,t} + \gamma_5 \frac{NRAMembers_{s,t}}{Population_{s,t}} + \\
 & \sum_c CountyEffects_c + \sum_t TimeEffects_t + \eta_{c,s,t}
 \end{aligned}$$

We employ the same data set used by DRS: an annual county-level panel from 1977-1996 containing data on the number of homicides, various deterrence measures and controls. A complete description of the data can be found in DRS.

DRS estimate this regression using two-stage least squares to control for potential endogeneity with respect to the deterrence measures (the terms with the β coefficients). The instruments they use, which are all at the state-level, are police expenditures, judicial expenditures, prison admissions, and Republican vote shares in each of the past six presidential elections.

⁵This corresponds to Model 4 of Table 4 from DRS.

DW criticize DRS for failing to cluster errors for counties within the same state. This criticism seems fair since one would generally expect dependence in the county-specific regression residuals that is not captured by state-level fixed effects. However, other forms of spatial dependence may be present. For example, crime rates may be more correlated across state borders than within states. In our analysis confidence intervals are reported using both the DW (clustered) and DRS (unclustered) calculations, where the clustering is done at the state level. Despite our view that the DW calculation is appropriate, we report both since, unlike our criticisms, clustering does not involve point estimates per se.

In evaluating deterrence effects, DRS emphasize the number of net lives saved per execution. For any model for the murder rate $M(e)$, where e is the number of executions, net lives saved is calculated as

$$Net\ Lives\ Saved = \frac{\partial M(e)}{\partial e} * Population - 1 \quad (24)$$

where $\frac{\partial M(e)}{\partial e}$ is the derivative of the murder rate with respect to the number of executions implied by each model.

The number of net lives saved in eq. (24) varies by locality. There are many ways of summarizing this metric. The approach used by DRS evaluates each term ($\frac{\partial M(e)}{\partial e}$ and population) in (24) at the average value among states with the death penalty in 1996. An alternative is to calculate the number of net lives saved for each state in each year separately. Not only does this seem to more appropriately correspond to the counterfactual of an additional execution in a given state, but it also allows us to analyze not just the mean of these effects, but also the median, or any other part of the distribution in which we are

interested.⁶ While in the text we refer to the mean number of net lives saved across states, in the tables we also present the DRS measure and the median of our alternative measure.

i. replication

In Table 1 we report results based on DRS’s preferred specification. In column 1 we replicate the results of DRS using their methodology. Focusing on the point estimates of net lives saved, executing one more criminal will save 29.0 lives on average across states versus 18.5 lives in the “average” 1996 death penalty state (the DRS estimate). These numbers are associated with negative parameter estimates for the three deterrence variables, which DRS treat as implications of the choice model of crime (although we have argued this is not so). These are the results used by DRS as evidence for the deterrence effect of the death penalty.

For both net lives saved and the model parameters, we report both asymptotic and bootstrapped confidence intervals, using 5,000 replications for the bootstrap.⁷ We do this for two reasons, both because bootstrapped confidence intervals are useful for addressing possible deviations of the regression residuals from normality and in order to preserve comparability with some of our subsequent results. As the table indicates, the bootstrap confidence intervals are substantially wider than the asymptotic ones, so much so that the statistical significance of the number of net lives saved and deterrence parameters are both lost. With the clustered confidence intervals, statistical significance fails even when the asymptotic calculation is used; which was the finding made by DW.

An important criticism by DW concerns the fragility of the DRS results to the definition of one of the instruments: the Republican vote share in the most recent presidential election. They show that when these six variables (one for each election) are collapsed into one variable, which restricts the effects to be the same for each election,

⁶We report these different measures of net lives saved as a way of showing the potential heterogeneity of the effect across localities. For each measure the range of potential net lives saved that can be inferred is given by the confidence intervals we report.

⁷The bootstrapped confidence intervals are calculated using a clustered non-parametric bootstrap procedure, in which state-time pairs are sampled and all observations (counties) for each sampled state-time pair are included to create the bootstrap sample.

finding that the number of net lives saved changes from positive to negative. We replicate their results in column 3 of Table 1. The point estimates in this case imply that executing a criminal *induces* 25.7 additional murders.⁸ While DW dismiss this as an unreasonable possibility, we note that there is no a priori reason why an increased possibility of execution should not raise the murder rate; the reason is that a higher likelihood of execution for a given murder reduces the marginal deterrence effect for subsequent murders. The DW finding on net lives saved is mirrored by their finding that increases in the conditional probability of a death sentence given a conviction and/or the conditional probability of an execution given a death sentence each produce an increase in the murder rate. As before, the bootstrapped confidence intervals are wider than the asymptotic ones resulting in an insignificant parameter and net lives saved. The results are also statistically insignificant with clustering for both the asymptotic and the bootstrapped cases.

It is unclear why a researcher should prefer the restricted specification of DW to the DRS specification unless one believes that by separating the vote shares one generates a weak instruments problem. A Stock-Yogo test rejects the null hypothesis of weak instruments, so we see no reason to prefer the restricted DW specification (collapsing the six voting instruments).

ii. microfoundations

We next analyze what happens when we employ the theoretically appropriate joint probabilities instead of the conditional probabilities used in the DRS and the DW specifications. As discussed in Section 5, the use of the conditional probabilities of being arrested, sentenced to death, and executed as covariates in the regression is not consistent with an expected utility model. Instead one should use the joint probabilities. That is, we replace the first three terms in the RHS of (23) with

⁸Note that, because net lives saved already includes the death of the criminal, when we report the number of murders induced by the execution (when net lives saved is negative), we do not count the executed criminal. For example, in column 3 or table 1, net lives saved equals -26.7 so an execution induces 25.7 additional murders.

$$\begin{aligned}
& \beta_1 \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} + \\
& \beta_2 \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} * \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} + \\
& \beta_3 \frac{Executions_{s,t}}{DeathSentences_{s,t-6}} * \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} * \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}}
\end{aligned} \tag{25}$$

The results are presented in column 2 (for the DRS specification) and column 4 (for the DW specification). The use of joint probabilities changes the net lives saved estimate for DRS to 31.5 per execution. The changes for DW are quite dramatic as their instrumental variable choice no longer reverses the sign of the point estimate of executing a criminal; under their specification each execution saves 20.9 lives. These results demonstrate that, once a correctly specified individual choice model is used to generate the aggregate model, the dispute over instrumental variables between DRS and DW does not really matter in terms of whether the estimated deterrence effect is positive or negative. As with the previous results, bootstrapping and/or clustering leads to statistically insignificant parameters and net lives saved.

iii. logit versus linear probability model

As argued in Section 3, a linear specification for aggregate crime regressions places strong assumptions on the underlying individual-specific errors; assumptions that are generally regarded as unappealing. In addition, while the linear probability model can give very similar results to alternative nonlinear models, this is the case only when the probabilities in the data (i.e. the murder rates) are not very close to 0 or 1. When this is not the case – in our case the largest murder rate we observe is 0.008 – the differences in tail behavior between models lead to different results. While many alternative probability models exist, it is straightforward to analyze the model when the errors are logistically distributed. This leads to an aggregate model which is linear in the right-hand-side

variables and nonlinear in the left-hand-side variable.⁹ In Table 2 we present the results obtained from running the analysis assuming the errors follow a logit distribution. We use the joint outcome probabilities implied by the individual choice problem as opposed to the conditional probabilities, so these results employ the theoretically correct set of deterrence variables. We report marginal effects in addition to parameter estimates as the former are most comparable to the parameter estimates for the linear model.

Tables 2a and 2b show that the specification of the error term has very large effects on the estimated number of net lives saved. For the DRS specification, in Table 2a, moving from a linear to a logistic error distribution assumption reverses their claims: each execution is associated with an average increase of 1.8 murders. For the DW specification, in Table 2b, each execution is associated with 7.2 additional murders. For both DW and DRS, while the marginal effect of an increase in the arrest probability decreases the murder rate, both capital sentences and executions are associated with more murders. Together, these results suggest that the DRS estimates of substantial deterrent effects to capital punishment are an artifice of their use of the linear probability model.¹⁰ Interestingly, the logit estimates of net lives saved are much less sensitive to the way the calculation is made than for the linear probability model. The asymptotic and bootstrapped confidence intervals for both the clustered and standard cases imply statistical insignificance of almost all of the parameter and all of the net lives saved estimates.

iv. instruments

Section 3 also argued that the presence of unobserved location-specific effects implies that all variables should be instrumented when aggregate data are used. In columns

⁹In particular, the LHS of (23) becomes $\log\left(\frac{Y}{100000 - Y}\right)$ where

$$Y = \frac{\text{Murders}_{c,s,t}}{\text{Population}_{c,s,t} / 100000}.$$

¹⁰In fact, with the exception of one specification (collapsed partisan variables, instrumenting for all regressors), the point estimates no longer imply a deterrent effect once we switch to logit errors in all cases considered in the paper.

3 and 4 of Tables 2a and 2b we present results for the logit analysis when this is done. As there are not enough instruments in the original instrument set to do this, we use lagged values of the instruments we have available.¹¹ This is likely to generate collinearity and/or weak instruments but we present the results for completeness. When this is done, we find the surprising result that the DRS specification produces 8.5 lives lost per execution, whereas DW find 9.2 net lives saved. Full instrumenting thus restores the discrepancy in estimates in the original DRS and DW papers. That said, the confidence intervals are sufficiently wide that one cannot conclude that either of the estimates (or the associated model parameters) is statistically significant or that the estimated differences between the specifications are statistically significant.

v. preference heterogeneity

In all of the results reported so far, we have assumed that the parameters of the utility function are common across individuals. In Table 3a we relax this assumption by allowing the coefficients on each of the three deterrence variables from DRS to be individual-specific. We focus on a model specification that combines joint deterrence probabilities, logit errors and the use of all the partisanship variables. The net utility function for this specification is

¹¹Specifically, we add 1, 2, 3, and 4 period lags of police expenditures, judicial expenditures, and prison admissions.

$$\begin{aligned}
u_{i,c,s,t} = & \beta_{i,1} \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} + \beta_{i,2} \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} * \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} \\
& + \beta_{i,3} \frac{Executions_{s,t}}{DeathSentences_{s,t-6}} * \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} * \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} \\
& + \gamma_1 \frac{Assaults_{c,s,t}}{Population_{c,s,t}} + \gamma_2 \frac{Robberies_{c,s,t}}{Population_{c,s,t}} + \gamma_3 CountyDemographics_{c,s,t} \\
& + \gamma_4 CountyEconomy_{c,s,t} + \gamma_5 \frac{NRAMembers_{s,t}}{Population_{s,t}} \\
& + \sum_c CountyEffects_c + \sum_t TimeEffects_t + \eta_{c,s,t} + \varepsilon_{i,c,s,t}
\end{aligned} \tag{26}$$

where i denotes an individual, c denotes a county, s denotes a state, and t denotes a year.¹² This combination of assumptions is, in our judgment, the most theoretically appealing of those we have considered. As discussed above, when one ventures outside the linear probability model, the calculation of policy effects will depend on the distribution of parameters.

Following (11), one has to integrate the individual choice probabilities implied by (26) in order to produce aggregate crime rates with interpretable microfoundations. In order to operationalize the integration we make the standard assumption that the random coefficients are uncorrelated with the regressors and that they are normally distributed¹³; this is the approach followed by Berry, Levinsohn, and Pakes (1995). Under the assumption of logit errors, this gives us the following expression for the aggregate county-level murder rate, $\rho_{c,s,t}$.

$$\rho_{c,s,t} = \iiint \left\{ \frac{\exp(\delta_{i,c,s,t})}{1 + \exp(\delta_{i,c,s,t})} \right\} dF_{\beta,1} dF_{\beta,2} dF_{\beta,3} \tag{27}$$

¹²Here and elsewhere, when c does not appear in a subscript, this means that the variable is measured at the state rather than county level.

¹³Relaxing normality of the random coefficients allowing for a multinomial distribution instead (that is the nonparametric maximum likelihood estimator of Heckman and Singer (1984)) produces similar results.

where

$$\begin{aligned}
\delta_{i,c,s,t} = & \beta_{i,1} \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} + \beta_{i,2} \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} * \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} \\
& + \beta_{i,3} \frac{Executions_{s,t}}{DeathSentences_{s,t-6}} * \frac{DeathSentences_{s,t}}{Arrests_{s,t-2}} * \frac{HomicideArrests_{c,s,t}}{Murders_{c,s,t}} \\
& + \gamma_1 \frac{Assaults_{c,s,t}}{Population_{c,s,t}} + \gamma_2 \frac{Robberies_{c,s,t}}{Population_{c,s,t}} + \gamma_3 CountyDemographics_{c,s,t} \quad (28) \\
& + \gamma_4 CountyEconomy_{c,s,t} + \gamma_5 \frac{NRMembers_{s,t}}{Population_{s,t}} \\
& + \sum_c CountyEffects_c + \sum_t TimeEffects_t + \eta_{c,s,t}
\end{aligned}$$

and $dF(\beta_{i,k})$, $k=1,2,3$, is a normal density with a mean and variance that will be estimated. We estimate this model using the algorithm developed by Berry, Levinsohn, and Pakes (1995). The parameters of the distribution of the preference parameters are estimated using a GMM procedure. The individual-specific heterogeneity is integrated out numerically. By matching the observed murder rates to those predicted by the model, we concentrate out the location-time specific unobservables, $\eta_{c,s,t}$ since this implicitly defines them as a function of the parameters of the distribution of the preference heterogeneity. The moments are then constructed by interacting the location-time specific unobservables with the control variables and the instruments.

Table 3a contains results from the random coefficients version of the specification with logit errors, microfounded (i.e. joint) deterrence probabilities, and the DRS set of the partisan variables.^{14,15} Once one allows for the possibility of individual specific responses to

¹⁴If one assumes that the random utility shocks are distributed uniformly instead of according to the logit distribution, it can be shown that the random coefficients integrate out and do not appear in the aggregate equation used for estimation. Therefore a model of random coefficients with uniformly distributed errors will collapse to a model without random coefficients.

the deterrence variables, the point estimate for the number of net lives changes from an additional execution inducing 1.8 murders to inducing 11.4 murders, although the results are still statistically insignificant (compare Table 3a to column 2 from Table 2a).

Table 3b illustrates the magnitude of the estimated preference heterogeneity in the data. In order to see how this heterogeneity relates to outcomes, in Table 3b we calculate the number of net lives saved under the assumption that everyone is at a given percentile of the distribution of the random coefficients (for each of the three deterrence probabilities: arrest rate; joint arrest and sentencing rate; and joint arrest, sentencing and execution rate). For example, in the first column we assume that everyone's preference for each of the three deterrence variables is at the 10th percentile of the estimated distribution. This is done for five different points along the estimated distribution.

As expected, for a proportion of the probability mass of preferences, people are in fact deterred from committing crimes by the deterrence variables. However, the effect varies so widely across the population that, if all the population were located at the same point, the response to an additional execution can vary from saving 6 net lives to inducing 108 more murders.

This heterogeneity of preferences has important policy implications. Consider a policy in which a state increases the probability of being executed conditional on being sentenced to death. Individuals at the far left of the preference distribution for this deterrence variable are the ones that are most deterred by an increase in this probability. However, they are also the ones who, *ceteris paribus*, are less likely to commit a crime in the first place, given their disutility associated with the probability of being executed. Similarly, individuals at the far right of the distribution, the ones who are most likely to commit a murder, are less affected by the increase. Consequently, such a policy may have very different effects in the presence of heterogeneity of preferences as opposed to the homogeneous case.

¹⁵We do not estimate a random coefficients model for the specification which collapses the partisan variables into one variable. We need at least six instruments for identification since there are three endogenous variables for which we need to instrument and three additional parameters to estimate (the standard deviations of the random coefficients). Collapsing the partisan variables leaves us with only four instruments.

Together, these results indicate that the assumptions implicit in the DRS specification do matter for their claims. The variability we have found in their estimates as substantive and statistical assumptions are altered suggests that their evidence for deterrence is in fact quite fragile.

7. Conclusions

In this paper, we have attempted to illustrate the assumptions needed to justify aggregate crime regressions along the causal lines that are conventional in the empirical crime literature. We have tried to make clear that aggregate crime regressions involve a range of statistical and behavioral assumptions whose validity is easily seen to be problematic. We have illustrated these abstract claims in the context of an important study of the deterrence effects of capital punishment, which illustrated that these assumptions matter in terms of deterrence effect claims.

Do our arguments imply that aggregate crime regressions have no role to play in positive and normative analyses? Horowitz (2004) essentially takes this view. We disagree. For example, there exist methods such as model averaging, which can allow for the integration of inferences from different sets of assumptions; Durlauf, Navarro, and Rivers (2008) discuss how these methods may be applied to crime studies, Cohen-Cole, Durlauf, Fagan and Nagin (2008) show how model averaging can adjudicate some of the disparate findings in the capital punishment and deterrence literature. More generally, as emphasized in Heckman (2000,2005), the interpretation of empirical work always requires assumptions, particular exercises are more or less informative based upon the strength of those assumptions. Aggregate crime regressions are thus no different in kind from other empirical exercises; as in any empirical analysis, what is important is to understand how assumptions matter and for a researcher to assess how this sensitivity ought to affect substantive conclusions such as a policy evaluation. See Durlauf, Navarro, and Rivers (2008) for further discussion and a general defense of the utility of aggregate crime regressions.

Table 1: Linear Probability Models

	DRS (All Partisan Variables)				DW (Collapsed Partisan Variables)			
	Conditional Probabilities		Joint Probabilities		Conditional Probabilities		Joint Probabilities	
	1		2		3		4	
	Asymptotic	Bootstrap	Asymptotic	Bootstrap	Asymptotic	Bootstrap	Asymptotic	Bootstrap
COEFFICIENTS								
Arrest Probability	-2.3 (-6.8, 2.2) [-3.3, -1.3]	-2.3 (-15.0, 6.4) [-11.2, 3.1]	-1.5 (-6.5, 3.5) [-2.5, -0.5]	-1.5 (-14.2, 6.8) [-11.1, 4.0]	-1.3 (-8.5, 6.0) [-2.7, 0.2]	-1.3 (-17.0, 12.8) [-7.6, 5.5]	-3.0 (-11.3, 5.3) [-4.2, -1.7]	-3.0 (-22.5, 10.5) [-12.2, 7.7]
Death Sentence Probability	-3.6 (-87.4, 80.2) [-32.1, 24.9]	-3.6 (-287.3, 322.8) [-170.2, 176.6]	--	--	212.9 (-205.7, 631.4) [161.2, 264.5]	212.9 (-275.9, 740.1) [-21.9, 520.8]	--	--
Execution Probability	-2.7 (-11.0, 5.6) [-3.9, -1.5]	-2.7 (-12.2, 12.7) [-7.5, 2.9]	--	--	2.3 (-10.4, 15.0) [0.7, 4.0]	2.3 (-8.9, 20.5) [-3.5, 8.4]	--	--
Death Sentence and Arrest Probability	--	--	-136.2 (-329.1, 56.6) [-174.7, -97.8]	-136.2 (-464.0, 347.9) [-400.1, 175.9]	--	--	400.4 (-274.9, 1,075.7) [316.6, 484.1]	400.4 (-609.3, 975.4) [-54.5, 1,050.4]
Execution, Sentence, Arrest Probability	--	--	-124.4 (-508.0, 259.2) [-190.1, -58.6]	-124.4 (-617.7, 596.1) [-382.1, 234.7]	--	--	-83.6 (-407.7, 240.5) [-167.4, 0.1]	-83.6 (-727.9, 950.9) [-558.2, 295.1]
NET LIVES SAVED								
DRS method	18.5 (-41.1, 78.0) [9.8, 27.1]	18.5 (-91.8, 86.4) [-21.5, 52.8]	14.8 (-33.9, 63.4) [6.4, 23.1]	14.8 (-76.5, 77.3) [-30.7, 47.4]	-17.7 (-108.6, 73.2) [-29.4, -6.0]	-17.7 (-147.9, 62.5) [-61.0, 23.8]	9.6 (-31.5, 50.7) [-1.1, 20.3]	9.6 (-121.5, 91.3) [-38.4, 69.8]
DNR method--mean	29.0 (-62.8, 120.7) [15.6, 42.4]	29.0 (-142.0, 132.7) [-32.6, 82.1]	31.5 (-68.8, 131.9) [14.3, 48.8]	31.5 (-156.8, 159.8) [-62.8, 99.8]	-26.7 (-166.9, 113.4) [-44.8, -8.7]	-26.7 (-228.4, 98.5) [-93.6, 37.3]	20.9 (-63.9, 105.7) [-1.2, 43.0]	20.9 (-246.8, 189.3) [-78.5, 146.5]
DNR method--median	16.8 (-37.7, 71.4) [8.9, 24.8]	16.8 (-84.5, 79.1) [-19.8, 48.5]	19.8 (-44.5, 84.1) [8.8, 30.9]	19.8 (-100.3, 102.1) [-40.6, 63.6]	-16.3 (-99.7, 67.1) [-27.1, -5.6]	-16.3 (-137.2, 57.4) [-56.1, 21.8]	13.0 (-41.3, 67.3) [-1.2, 27.2]	13.0 (-158.2, 122.4) [-50.7, 93.5]

Notes:

1. For each specification, four sets of 95% confidence intervals are reported in parentheses below the parameter estimates. The left column reports confidence intervals based on the asymptotic distribution. The right column reports bootstrapped confidence intervals based on 5,000 replications. Clustered confidence intervals (at the state-level) are reported in parentheses. Unclustered confidence intervals are reported in brackets below.
2. The results are based on an available 43,535 observations.
3. In DRS, the net lives saved is calculated by plugging in the average of characteristics for states with the death penalty in 1996.
4. In the DNR method, we calculate the net lives saved for each state in each year separately, and then compute the mean and median of this distribution.

Table 2a: Logistic Probability Models*

	DRS (All Partisan Variables)					
	Marginal Effect		Coefficient		Instrumenting for all Regressors	
	1	2		3	4	
		Asymptotic	Bootstrap		Asymptotic	Bootstrap
COEFFICIENTS						
Arrest Probability	-0.3	-0.1 (-0.5, 0.3) [-0.2, 0.1]	-0.1 (-0.9, 0.6) [-0.6, 0.3]	0.6	0.1 (-0.4, 0.7) [-0.2, 0.4]	0.1 (-0.8, 0.5) [-0.7, 0.5]
Death Sentence and Arrest Probability	18.0	3.3 (-12.9, 19.4) [-2.1, 8.7]	3.3 (-24.5, 27.8) [-16.4, 20.1]	-11.5	-2.1 (-20.6, 16.4) [-11.3, 7.1]	-2.1 (-15.1, 17.1) [-15.6, 22.6]
Execution, Sentence, Arrest Probability	11.5	2.1 (-31.7, 35.9) [-7.1, 11.3]	2.1 (-34.4, 51.6) [-20.9, 25.2]	53.4	9.7 (-24.2, 43.6) [-2.6, 22.1]	9.7 (-54.5, 39.1) [-38.5, 41.0]
NET LIVES SAVED						
<i>DRS method</i>			-2.6 (-40.8, 25.8) [-20.9, 15.2]			-8.3 (-31.4, 39.2) [-32.5, 28.3]
<i>DNR method--mean</i>			-2.8 (-46.1, 28.9) [-23.0, 17.1]			-9.5 (-35.1, 46.6) [-36.6, 32.4]
<i>DNR method--median</i>			-2.7 (-42.1, 26.4) [-20.6, 15.1]			-8.8 (-32.5, 42.8) [-32.7, 28.7]

* The dependent variable for the logit model is $\log[P/(100,000-P)]$, where P is the murder rate per 100,000 people. The RHS is the same as the linear model. Columns 1 and 3 report the average marginal effects (per 100,000 people) so that they are comparable to the linear coefficients. That is, $1/n \sum_{i=1,n} \partial 100,000P(x_i)/\partial x_{ij}$.

Notes:

- For each specification, four sets of 95% confidence intervals are reported for the coefficient estimates in parentheses below the parameter estimates. The left column reports confidence intervals based on the asymptotic distribution. The right column reports bootstrapped confidence intervals based on 5,000 replications. Clustered confidence intervals (at the state-level) are reported in parentheses. Unclustered confidence intervals are reported in brackets below. For the net lives saved measures only bootstrapped confidence intervals are reported due to the computation involved in calculating the asymptotic standard errors given the large number of fixed effects.
- The results in columns 1 and 2 are based on an available 28,256 observations. The results in the remaining columns are based on 24,842 observations. The dependent variable for the logit is not defined when the murder rate is zero, so these observations are dropped for the logit specifications. For columns 3 and 4 we use lagged values of the instruments as our additional instruments, so this causes us to drop observations from earlier periods for which we do not have values for all of the instruments.
- In DRS, the net lives saved is calculated by plugging in the average of characteristics for states with the death penalty in 1996.
- In the DNR method, we calculate the net lives saved for each state in each year separately, and then compute the mean and median of this distribution.

Table 2b: Logistic Probability Models*

	DW (Collapsed Partisan Variables)					
			Instrumenting for all Regressors			
	Marginal Effect	Coefficient		Marginal Effect	Coefficient	
	1	2	3	4	6	
	Asymptotic	Bootstrap		Asymptotic	Bootstrap	
COEFFICIENTS						
Arrest Probability	-0.8	-0.1 (-0.8, 0.5) [-0.3, 0.0]	-0.1 (-1.7, 1.1) [-1.0, 0.9]	-2.5	-0.4 (-1.1, 0.2) [-0.9, 0.0]	-0.4 (-2.2, 1.7) [-1.9, 1.9]
Death Sentence and Arrest Probability	153.4	27.9 (-23.5, 79.3) [16.2, 39.6]	27.9 (-53.6, 82.6) [-18.3, 86.5]	87.4	15.9 (-10.8, 42.6) [3.4, 28.4]	15.9 (-44.8, 48.8) [-51.6, 52.3]
Execution, Sentence, Arrest Probability	45.5	8.3 (-28.4, 45.0) [-3.5, 20.0]	8.3 (-52.7, 94.5) [-39.0, 42.2]	-64.4	-11.7 (-80.5, 57.1) [-32.2, 8.8]	-11.7 (-125.9, 121.8) [-99.9, 102.1]
NET LIVES SAVED						
DRS method			-7.5 (-73.6, 39.5) [-34.3, 29.7]			8.0 (-95.6, 96.3) [-80.7, 75.8]
DNR method--mean			-8.2 (-83.6, 44.8) [-37.8, 33.2]			9.2 (-106.6, 108.4) [-90.3, 85.4]
DNR method--median			-7.6 (-76.7, 41.3) [-33.7, 29.5]			8.4 (-98.5, 99.9) [-80.3, 76.2]

* The dependent variable for the logit model is $\log[P/(100,000-P)]$, where P is the murder rate per 100,000 people. The RHS is the same as the linear model. Columns 1 and 3 report the average marginal effects (per 100,000 people) so that they are comparable to the linear coefficients. That is, $1/n \sum_{i=1,n} \partial 100,000P(x_i)/\partial x_{ij}$.

Notes:

- For each specification, four sets of 95% confidence intervals are reported for the coefficient estimates in parentheses below the parameter estimates. The left column reports confidence intervals based on the asymptotic distribution. The right column reports bootstrapped confidence intervals based on 5,000 replications. Clustered confidence intervals (at the state-level) are reported in parentheses. Unclustered confidence intervals are reported in brackets below. For the net lives saved measures only bootstrapped confidence intervals are reported due to the computation involved in calculating the asymptotic standard errors given the large number of fixed effects.
- The results in columns 1 and 2 are based on an available 28,256 observations. The results in the remaining columns are based on 24,842 observations. The dependent variable for the logit is not defined when the murder rate is zero, so these observations are dropped for the logit specifications. For columns 3 and 4 we use lagged values of the instruments as our additional instruments, so this causes us to drop observations from earlier periods for which we do not have values for all of the instruments.
- In DRS, the net lives saved is calculated by plugging in the average of characteristics for states with the death penalty in 1996.
- In the DNR method, we calculate the net lives saved for each state in each year separately, and then compute the mean and median of this distribution.

Table 3a: Random Coefficients Specification

	All Partisan Variables Joint Probabilities			
	Mean		Standard Deviation	
	Asymptotic	Bootstrap	Asymptotic	Bootstrap
<u>COEFFICIENTS</u>				
Arrest Probability	-1.7 (-28.9, 25.5) [-13.4, 10.0]	-1.7 (-4.4, 0.3) [-3.9, 0.0]	1.2 (-10.4, 12.8) [-4.6, 7.0]	1.2 (0.0, 2.5) [0.0, 1.8]
Death Sentence and Arrest Probability	1.1 (-227.4, 229.5) [-88.8, 91.0]	1.1 (-52.7, 33.3) [-30.8, 27.3]	3.3 (-273.6, 280.3) [-162.6, 169.2]	3.3 (0.0, 33.1) [0.0, 27.0]
Execution, Sentence, Arrest Probability	8.9 (-202.5, 220.3) [-84.4, 102.3]	8.9 (-103.1, 60.5) [-49.8, 58.4]	19.5 (-152.3, 191.3) [-53.7, 92.6]	19.5 (0.0, 99.8) [0.0, 47.2]
<u>NET LIVES SAVED</u>				
<i>DNR method--mean</i>			-12.4 (-78.6, 121.7) [-76.6, 54.7]	
<i>DNR method--median</i>			-8.8 (-54.9, 82.7) [-53.4, 37.5]	

Notes:

1. This specification assumes that the random utility shocks are distributed according to the logit distribution, and that there are normally distributed random coefficients on each of the joint deterrence variables.
2. Four sets of 95% confidence intervals are reported for the coefficient estimates in parentheses below the parameter estimates. The left column reports confidence intervals based on the asymptotic distribution. The right column reports bootstrapped confidence intervals based on 5,000 replications. Clustered confidence intervals (at the state-level) are reported in parentheses. Unclustered confidence intervals are reported in brackets below. For the net lives saved measures only bootstrapped confidence intervals are reported due to the computation involved in calculating the asymptotic standard errors given the large number of fixed effects.
3. The results are based on an available 28,256 observations.

Table 3b: Random Coefficients Specification

	Percentile				
	10th	30th	50th	70th	90th
<u>NET LIVES SAVED</u>					
<i>DNR method--mean</i>	5.5	-0.2	-8.5	-26.4	-108.9

Notes:

1. The net lives saved are calculated based on the estimates from the random coefficient specification from Table 3a.
2. In each column, the number of net lives saved is calculated under the assumption that every individual's taste for the joint deterrence probabilities is at the given percentile of the estimated distribution. For example, the number in the third column is the number of net lives saved by one additional execution if everyone had the median estimated taste for the arrest, joint arrest and sentencing, and joint arrest, sentencing, and execution probabilities.

Bibliography

Abbring, J. and J. Heckman, (2007), "Econometric Evaluation of Social Programs, Part III: Distributional Treatment Effects, Dynamic Treatment Effects, Dynamic Discrete Choice, and General Equilibrium Policy Evaluation," in *Handbook of Econometrics*, vol 6b, J. Heckman and E. Leamer, eds., Amsterdam: North Holland.

Ayers, I. and J. Donohue, (2003), "Shooting Down the "More Guns, Less Crime" Hypothesis," *Stanford Law Review*, 55, 1193-1312.

Becker, G., (1968), "Crime and Punishment: An Economic Analysis," *Journal of Political Economy*, 78, 2, 169-217.

Berry, S., J. Levinsohn, and A. Pakes, (1995), "Automobile Prices in Market Equilibrium," *Econometrica*, 63, 4, 841-890.

Black, D. and D. Nagin, (1998), "Do Right-to-Carry Laws Deter Violent Crimes," *Journal of Legal Studies*, 27, 1, 209-219.

Blundell, R. and T. Stoker, (2005) "Heterogeneity and Aggregation," *Journal of Economic Literature*, 43, 2, 347-391.

Brock, W. and S. Durlauf, (2001), "Growth Empirics and Reality," *World Bank Economic Review*, 15, 2, 229-272.

Cohen-Cole, E., S. Durlauf, J. Fagan, and D. Nagin, (2008), "Model Uncertainty and the Deterrent Effect of Capital Punishment," *American Law and Economics Review*, forthcoming.

Dezhbakhsh, H. and P. Rubin, (2007), "From the "Econometrics of Capital Punishment" to the "Capital Punishment" of Econometrics: On the Use and Abuse of Sensitivity Analysis," mimeo, Emory University.

Dezhbakhsh, H., P. Rubin, and J. Shepherd, (2003), "Does Capital Punishment Have a Deterrent Effect? New Evidence from Post-Moratorium Panel Data," *American Law and Economics Review*, 5, 2, 344-76.

Dhrymes, P., (1963), "A Comparison of Productivity Behavior in the Manufacturing and Service Industries, U.S. 1947-1958," *Review of Economics and Statistics*, 45, 64-69.

Dhrymes, P., (1985), "Review of On the Foundations of Supply-Side Economics by Victor A. Canto, Douglas H. Jones, and Arthur Laffer," *Journal of the American Statistical Association*, 3, 2, 174-177.

Dhrymes, P. and E. Bartlesmann, (1998), "Productivity Dynamics: US Manufacturing Plants, 1972-1986," *Journal of Productivity*, 9, 5-34.

Donohue, J. and S. Levitt, (2001), "The Impact of Legalized Abortion on Crime," *Quarterly Journal of Economics*, 116, 2, 379-420.

Donohue, J. and S. Levitt, (2004), "Further Evidence that Legalized Abortion Lowered Crime: A Reply to Joyce," *Journal of Human Resources*, 39, 1, 29-49.

Donohue, J. and S. Levitt, (2008), "Measurement Error, Legalized Abortion, and the Decline in Crime: A Response to Foote and Goetz," *Quarterly Journal of Economics* . 123, 1, 425-440.

Donohue, J. and J. Wolfers, (2005), "Uses and Abuses of Empirical Evidence in the Death Penalty Debate," *Stanford Law Review*, 58, 3, 791-846.

Durlauf, S., S. Navarro, and D. Rivers, (2008), "On the Use of Aggregate Crime Regressions in Policy Evaluation," in *Understanding Crime Trends*, A. Goldberger and R. Rosenfeld eds., Washington DC: National Academies Press.

Ehrlich, I., (1972), "The Deterrent Effect of Criminal Law Enforcement," *Journal of Legal Studies*, L, 2, 259-276.

Ehrlich, I., (1973), "Participation in Illegal Activities: A Theoretical and Empirical Investigation," *Journal of Political Economy*, 81, 3, 521-565.

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

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

Foote, C. and C. Goetz, (2008), "The Impact of Legalized Abortion on Crime: Comment," *Quarterly Journal of Economics*, 123, 1 407-423.

Heckman, J., (2000), "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *Quarterly Journal of Economics*, 115, 1, 45-97.

Heckman, (2005), "The Scientific Model of Causality," *Sociological Methodology*, 35, 1-98.

Heckman, J. and S. Navarro, (2007), "Dynamic Discrete Choices and Dynamic Treatment Effects," *Journal of Econometrics*, 136, 2, 341-396.

Horowitz, J., (2004), "Statistical Issues in the Evaluation of Right-to-Carry Laws," in *Firearms and Violence*, C. Wellford, J. Pepper, and C. Petrie, eds., Washington DC: National Academies Press.

- Joyce, T., (2004), "Did Legalized Abortion Lower Crime?," *Journal of Human Resources*, 39, 1, 1-28.
- Katz, L., S. Levitt, and E. Shustorovich, (2001), "Prison Conditions, Capital Punishment, and Deterrence," *American Law and Economics Review*, 5, 318-43.
- Lott, J., (1998), "The Concealed-Handgun Debate," *Journal of Legal Studies*, 27, 1, 221-243.
- Lott, J. and D. Mustard, (1997), "Crime, Deterrence, and Right-to-Carry Concealed Handguns," *Journal of Legal Studies*, 26, 1, 1-68.
- Lott, J. and J. Whitley, (2007), "Abortion and Crime: Unwanted Children and Out-of-Wedlock Births," *Economic Inquiry*, 45, 2, 304-324.
- Lynch, J.P. and L. Addington, (2007). *Understanding Crime Incidence Statistics: Revisiting the Divergence of the NCVS and the UCR*. New York: Cambridge University Press.
- Matzkin, R., (1992), "Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models," *Econometrica*, 60, 2, 239-270.
- Mocan, N. and R. Gittings, (2001), "Getting Off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment," *Journal of Law and Economics*, 46, 2, 453-78.
- Mocan, N. and R. Gittings, (2006), "The Impact of Incentives on Human Behavior: Can We Make it Disappear? The Case of the Death Penalty," *National Bureau of Economic Research Working Paper no. 12631*, Cambridge MA.
- Plassmann, F. and J. Whitley, (2003), "Confirming "More Guns, Less Crime," *Stanford Law Review*, 55, 1313-1369.
- Stoker, T. (1993), "Empirical Approaches to the Problem of Aggregation over Individuals," *Journal of Economic Literature*, 31, 4, 1827-1874.