On the Interpretation of Aggregate Crime Regressions

Steven N. Durlauf

Salvador Navarro

David A. Rivers

University of Wisconsin at Madison

September 30, 2007

We thank the National Science Foundation and University of Wisconsin Graduate School for financial support. Arthur Goldberger and Justin Wolfers provided immensely helpful comments on a previous draft.

1. Introduction

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 include the deterrent effect of shall-issue concealed weapons legislation (e.g. Lott and Mustard (1997), Lott (1998), Black and Nagin (1998), Ayers and Donohue (2003), Plassmann and Whitley (2003), and the deterrent effect of capital punishment (e.g. Dezhbakhsh, Rubin, and Shepherd, (2003), Donohue and Wolfers (2005)).

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 comprehensive, 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 between the aggregate data and individual behavior.[1]

Our analysis is closest in spirit to Horowitz (2004) which, although focusing on the context of concealed weapons laws, describes some general difficulties in drawing causal inferences about aggregate crime using observational data. Horowitz' discussion emphasizes the sensitivity of regression findings to the choice of control variables, functional forms, and other assumptions. He argues that lack of prior knowledge as to the validity of such assumptions essentially eliminates the ability of crime regressions to uncover policy effects. In contrast, he argues that inferences about policy effects may, in principle, be drawn in contexts where the policy has been implemented randomly; i.e. that the data may be conceptualized as the outcome of an experiment. We defend a different perspective in that we start with a microeconomic choice problem for individual decisions and discuss sufficient conditions from which the individual decisions aggregate

[1] The 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.

to linear regressions. These conditions are subject to the sorts of criticisms that Horowitz makes with respect to sensitivity of empirical findings to assumptions. However, we argue that this sensitivity may be interpreted as a consequence of model uncertainty and as such may be constructively addressed. The differences between Horowitz's perspective and ours reduce to what sorts of prior beliefs and reasoning one is willing to bring to a statistical exercise. We agree with him that the data cannot speak for themselves. We disagree with the degree of his pessimism about inferences with observational data in that we see a role for economic reasoning and decision-theoretic considerations in the determination of what information is provided by a given regression or set of regressions.

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 considers issues of model uncertainty in crime regressions. Section 6 discusses these general arguments in the context of two prominent papers in the capital punishment and deterrence literature. Section 7 concludes. Our discussion is conceptual: Durlauf, Navarro, and Rivers (2007) provides a more systematic treatment of many of the issues we raise as well as an empirical application.

2. Crime as a choice

From the vantage point of economics, 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). This logic underlies the renaissance of crime research in economics, exemplified in work such as Steven Levitt's e.g. Levitt (1996) and Donohue and Levitt (2001).

In constructing a formal model, the idea that crime is purposeful means that an observed criminal act is understood as the outcome of a decision problem in which a criminal maximizes an expected utility function subject to whatever constraints he faces. The utility function is not a primitive assumption about behavior (i.e. no economist thinks that agents carry explicit representations of utility functions in their heads), rather it is a mathematical representation of an individual's preferences, one which constitutes a rank ordering across the potential actions the individual may take.

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 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. That said, 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 individual's available information. While the relaxation of this notion of rationality has been a major theme in recent economic research (behavioral economics is now an established field of the discipline), we will use relatively standard notions 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.[2]

To see how crime choice may be formally described, we follow the standard binary choice model of economics; our only modifications are to place the model in a crime context. 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

---

[2]Becker (1993) provides a lovely statement of the economic perspective on behavior:

> "The analysis assumes that individuals maximize welfare *as they conceive it*, whether they be selfish, altruistic, loyal spiteful, or masochistic. Their behavior is forward looking, and it also assumed to be consistent over time. In particular they try as best they can to anticipate the consequences of their actions." (pg. 386)

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. A standard form for the expected utility associated with the choice $u_{i,t}\left(\omega_{i,t}\right)$ is

$$u_{i,t}\left(\omega_{i,t}\right) = Z_{l,t}\beta\omega_{i,t} + X_{i,t}\gamma\omega_{i,t} + \xi_{l,t}\left(\omega_{i,t}\right) + \varepsilon_{i,t}\left(\omega_{i,t}\right). \tag{1}$$

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, $\xi_{l,t}\left(\omega_{i,t}\right)$ and $\varepsilon_{i,t}\left(\omega_{i,t}\right)$ denote unobservable location-specific and individual-specific utility terms. 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

$$v_{i,t} = Z_{l,t}\beta + X_{i,t}\gamma + \xi_{l,t}(1) - \xi_{l,t}(0) + \varepsilon_{i,t}(1) - \varepsilon_{i,t}(0) \tag{2}$$

the choice-based perspective amounts to saying that a person chooses to commit a crime, i.e. $\omega_{\mathrm{i,t}} = 1$ if and only if $v_{\mathrm{i,t}} > 0$.

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.

We restrict the nature of the unobservable heterogeneity with three assumptions.

A.1 $$E\left(\varepsilon_{i,t}(1)-\varepsilon_{i,t}(0)\right)=0 \tag{3}$$

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

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

Assumption A.1, that the differences in individual utility errors have 0 mean is usually justified by assuming that $Z_{l,t}$ includes a constant term. Assumption A.2 allows us to consider the two types of unobservables separately. Assumption A.3 corresponds to the orthogonality of regressors and errors in the linear model. Assumptions A.2 and A.3 are sufficient rather than necessary.

Under our utility specification, it is immediate that a positive net utility from commission of a crime requires that $X_{i,t}\gamma + Z_{l,t}\beta + \xi_{l,t}(1)-\xi_{l,t}(0) > \varepsilon_{i,t}(0)-\varepsilon_{i,t}(1)$. Conditional on $X_{i,t}, Z_{l,t}$, and $\xi_{l,t}(1)-\xi_{l,t}(0)$, 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 a crime is committed. Formally,

$$\Pr\left(\omega_{i,t}=1 \middle| Z_{l,t}, X_{i,t}, \xi_{l,t}(1)-\xi_{l,t}(0)\right) = G_{i,t}\left(Z_{l,t}\beta + X_{i,t}\gamma + \xi_{l,t}(1)-\xi_{l,t}(0)\right). \tag{6}$$

This conditional probability structure fully captures the microfoundations of the economic model.[3]

---

[3]This formulation is a quite 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 sequences 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

3. Aggregation

How do the conditional crime probabilities for individuals described by (6) 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 (6), 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 $F_{X_{l,t}}$ denote the empirical distribution function of $X_{i,t}$ within *l*. The expected crime rate in a location at a given time is

$$E\left(\rho_{l,t} \middle| Z_{l,t}, F_{X_{l,t}}, \xi_{l,t}(1) - \xi_{l,t}(0)\right) = \int G_{i,t}\left(Z_{l,t}\beta + X\gamma + \xi_{l,t}(1) - \xi_{l,t}(0)\right) dF_{X_{l,t}}. \quad (7)$$

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

A.4 $dG_{i,t}$ is a uniform density. (8)

Applying A.4 to (7), the crime rate in locality *l* obeys

$$\rho_{l,t} = Z_{l,t}\beta + \bar{X}_{l,t}\gamma + \xi_{l,t}(1) - \xi_{l,t}(0) + \theta_{l,t}, \quad (9)$$

---

selection, these can be readily dealt with using the analysis of Heckman and Navarro (2007), albeit at the expense of considerable complication of the analysis.

where $\bar{X}_{l,t}$ is the empirical mean of $X_{i,t}$ within l and $\theta_{l,t} = \rho_{l,t} - E\left(\rho_{l,t} \middle| Z_{l,t}, F_{X_{l,t}}, \xi_{l,t}(1) - \xi_{l,t}(0)\right)$ captures the difference between the realized and expected crime rate within a locality. This is the model typically employed in aggregate crime regressions.

Our construction of eq. (9) from choice-based foundations illustrates how standard aggregate crime regressions require a number of statistical assumptions if they are to be interpreted as aggregations of individual behavior. The assumption of the linear probability model is of concern since it implicitly restricts the random utility terms in an odd way. Unfortunately, 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 $\log\left(\frac{\Pr_{i,t}\left(\omega_{i,t} = 1 \middle| Z_{l,t}, X_{i,t}, \xi_{l,t}(1) - \xi_{l,t}(0)\right)}{1 - \Pr_{i,t}\left(\omega_{i,t} = 1 \middle| Z_{l,t}, X_{i,t}, \xi_{l,t}(1) - \xi_{l,t}(0)\right)}\right)$ will be linear in the various payoff components, but will not produce a closed form solution for the aggregate crime rate. Methods are available to allow for analysis of aggregate data under logit type assumptions, see Berry, Levinsohn, and Pakes (1995), but have not been applied, as far as we know, to the crime context.

On its own terms, the linear aggregate crime model indicates how aggregation affects the consistency of particular estimators. 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 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 (9). By implication, this means that all the variables in (9) should be instrumented. Hence in our judgment the focus on instrumenting 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.

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

$$v_{i,t}^{A} = Z_{l,t}^{A}\beta^{A} + X_{i,t}^{A}\gamma^{A} + \xi_{l,t}^{A}(1) - \xi_{l,t}^{A}(0) + \varepsilon_{i,t}^{A}(1) - \varepsilon_{i,t}^{A}(0) \tag{10}$$

and

$$v_{i,t}^{B} = Z_{l,t}^{B}\beta^{B} + X_{i.t}^{B}\gamma^{B} + \xi_{l,t}^{B}(1) - \xi_{l,t}^{B}(0) + \varepsilon_{i,t}^{B}(1) - \varepsilon_{i,t}^{B}(0) \tag{11}$$

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

$$\begin{gathered} v_{i,t} = \\ Z_{l,t}^{A}\beta^{A} + X_{i,t}^{A}\gamma^{A} + \xi_{l,t}^{A}(1) - \xi_{l,t}^{A}(0) + \varepsilon_{i,t}^{A}(1) - \varepsilon_{i,t}^{A}(0) + \\ D_{l,t}\left(Z_{l,t}^{B}\beta^{B} - Z_{l,t}^{A}\beta^{A}\right) + D_{l,t}\left(X_{i,t}^{B}\gamma^{B} - X_{i,t}^{A}\gamma^{A}\right) + \\ D_{l,t}\left(\xi_{l,t}^{B}(1) - \xi_{l,t}^{B}(0) - \left(\xi_{l,t}^{A}(1) - \xi_{l,t}^{A}(0)\right)\right) \\ D_{l,t}\left(\varepsilon_{i,t}^{B}(1) - \varepsilon_{i,t}^{B}(0) - \left(\varepsilon_{i,t}^{A}(1) - \varepsilon_{i,t}^{A}(0)\right)\right) \end{gathered} \tag{12}$$

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

$$\rho_{l,t} =$$
$$Z_{l,t}^{A}\beta^{A} + \bar{X}_{l,t}^{A}\gamma^{A} + D_{l,t}\left(Z_{l,t}^{B}\beta^{B} - Z_{l,t}^{A}\beta^{A}\right) + D_{l,t}\left(\bar{X}_{l,t}^{B}\gamma^{B} - \bar{X}_{l,t}^{A}\gamma^{A}\right) + \tag{13}$$
$$\xi_{l,t}^{A}(1) - \xi_{l,t}^{A}(0) + \theta_{l,t}^{A} + D_{l,t}\left(\xi_{l,t}^{B}(1) - \xi_{l,t}^{B}(0) - \left(\xi_{l,t}^{A}(1) - \xi_{l,t}^{A}(0)\right) + \theta_{l,t}^{B} - \theta_{l,t}^{A}\right)$$

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. This allows the policy effect to be measured by $\left(Z_{l,t}^{B} - Z_{l,t}^{A}\right)\hat{\beta}$. Eq. (13) indicates how a number of assumptions are embedded in the standard approach, in particular the requirement $\xi_{l,t}^{B}(1) - \xi_{l,t}^{B}(0) - \left(\xi_{l,t}^{A}(1) - \xi_{l,t}^{A}(0)\right) = 0$, 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 we will argue below that there are cases, specifically capital punishment, where this does not receive adequate attention in the relevant empirical formulations.

5. Model uncertainty

In this section, we address the issue of how to analyze crime rates when the researcher is uncertain about the appropriate statistical model of the crime process. In raising this issue, we emphasize that the problem of inadequate attention to model uncertainty is in no way unique to criminology. Nor do we mean to suggest that criminological studies are unique in the extent to which authors fail to investigate how modifications in baseline models affect inferences.

i. characterizing model uncertainty

Our reading of the criminology literature suggests several general sources of model uncertainty. The categories we will describe have previously been proposed by Brock, Durlauf and West (2003) for economic growth models and Brock, Durlauf, and West (2007) for business cycle models. These categories thus capture uncertainties that are common in social science analyses. At the same time, our decomposition of model uncertainty should not be interpreted as based on natural kinds; one can well imagine alternative divisions.

theory uncertainty

Social science theories for a given phenomenon are often open-ended (Brock and Durlauf (2001)) which means that one theory does not logically exclude another as having additional explanatory power. Hence there is often no justification for focusing on a subset of plausible explanations in empirical work. Some evidence of why this matters is suggested by Levitt's (2004) evaluation of sources of the crime decline of the 1990's. Levitt identifies 10 alternative theories of the crime decline, all of which are mutually consistent. Without questioning any of his substantive conclusions, we do note that Levitt is to a large extent forced to evaluate the roles of the different theories based on studies each of which does not account for the full range of the other theories when measuring the empirical salience of a particular one.

statistical instantiation

Models may differ with respect to details of statistical specification which have nothing to do with the underlying social science theories which motivate them, but rather are necessary to translate these theories into representations that are amenable to data analysis. This is typically so even when the social science theories are themselves expressed mathematically. Differences in these assumptions can lead to different findings.

A good example of how differences in statistical assumptions can affect substantive conclusions is specification of time trends. In the context of the deterrence effects of shall-issue concealed carry laws, different time trend choices have proven to be important. Specifically, Black and Nagin (1998) find that the use of quadratic time trends in place of state-specific linear time trends eliminates the evidence of a link between liberalization of concealed weapons laws and crime rates found in Lott and Mustard (1997). Lott's rejoinder (1998) argues that it is hard to identify the effects of a policy change (in this case concealed weapons legality) because a quadratic trend will mask it; intuitively, if crime is rising before a law is passed and decreases thereafter, this will be approximated by the quadratic trend.[4] Lott's intuition may be correct, but his argument is question begging as it applies in both directions. If crime follows an exogenously determined quadratic trend over some crime interval, and rising crime levels lead to a change in legislation, then Lott's approach will spuriously identify a causal effect from the legislation. This is true even if state-specific trends are employed.

From the perspective of model uncertainty, Black and Nagin and Lott are working with different statistical instantiations of unexplained temporal heterogeneity. Under the Black and Nagin specification, there may be, as Lott argues, substantial collinearity between the variable used to measure temporal heterogeneity and the variables used to measure the effects of shall-issue concealed weapons legislation. This multicollinearity does not invalidate the Black and Nagin model on logical grounds. In our judgment, the differences between Black and Nagin and Lott on this issue reflect the absence of good explanations for much of the temporal evolution of crime rates. Neither a linear specification nor a quadratic specification (or for that matter, more complicated splines or alternative semiparametric methods) instantiate substantive ideas about the crime process. Rather, they constitute efforts to purge the data so that the residual components may be analyzed.

Trend specification also matters in the analysis of unemployment rates and crime. Greenberg (2001) criticizes Cantor and Land (1985) for modeling trends using deterministic rather than unit root methods. Again, social science theory does not dictate

---

[4]This argument is further developed in Plassmann and Whitley (2003).

a preference for one type of trend versus another. While both Greenberg and Cantor suggest microfounded arguments in favor of their trend specifications, neither of them demonstrates a one-to-one mapping from these arguments to their modeling assumptions.

Other examples of this type of model uncertainty include assumptions about additivity, linearity and the use of logarithms versus levels.

parameter heterogeneity

A third source of model uncertainty concerns parameter heterogeneity. Another source of differences between Black and Nagin and Mustard and Lott illustrates this in that Black and Nagin find sensitivity of at issue effects to the presence of Florida in the data set. Black and Nagin show that eliminating data from Florida eliminates the evidentiary support for a handgun/crime link from some of the Lott and Mustard specifications. A second example appears in the capital punishment literature, which we will discuss in detail below.

ii. model averaging

How can the sensitivity of empirical claims to model choice be constructively addressed? We describe a strategy based on model averaging; ideas associated with model averaging appear to originate in Leamer (1978). They have become prominent in the last decade within statistics; a valuable conceptual argument is made in Draper (1995) and the development of formal methods has been greatly advanced by Adrian Raftery e.g. Raftery, Madigan and Hoeting (1997). We proceed using Bayesian language for expositional convenience, but the analysis can be done using frequentist estimators.

For a given exercise, suppose that the objective of the researcher is to construct a density of crime rates $\rho_{l,t+1}$ based on data D_t. When model m is assumed to be correct, this density may be written as

$$\Pr\left(\rho_{l,t+1} \middle| D_t, m\right) \tag{14}$$

Many disagreements about substantive empirical questions such as forecasts or the effects of alternative policies, derive from disagreements about the choice of model, *m*. This is of course why model selection plays such a significant role in empirical work. From the perspective of some empirical questions, it is not obvious that this is the appropriate role for model choice. If the goal of an exercise is to compare policies, the model choice is a nuisance parameter. Similarly, if one wants to construct a forecast, then the model itself is not intrinsically interesting.

In order to avoid dependence on a particular model specification, an alternative strategy is to develop conclusions based upon a space of candidate models; denote this space as M. Probability statements about a future outcome such as $\rho_{l,t+1}$ can then be constructed conditioning on the entire model space rather than on one of its elements. In other words, one computes the probability density

$$\Pr\left(\rho_{l,t+1} \middle| D_t, M\right). \tag{15}$$

From this perspective, the true model is an unknown that needs to be integrated out of the probability density, in the sense that

$$\Pr\left(\rho_{l,t+1} \middle| D_t, M\right) = \sum_{m \in M} \Pr\left(\rho_{l,t+1} \middle| D_t, m\right) \Pr\left(m \middle| D_t\right). \tag{16}$$

Here $\Pr\left(m \middle| D_t\right)$ denotes the posterior probability that m is the correct model given the data.

The model averaging approach can be used to calculate the usual objects of empirical interest, the expected value and the variance of $\rho_{l,t+1}$. Following derivations due to Leamer (1978), one can show that

$$E\left(\rho_{l,t+1} \middle| D_t, M\right) = \sum_{m \in M} E\left(\rho_{l,t+1} \middle| D_t, m\right) \Pr\left(m \middle| D_t\right) \tag{17}$$

and

$$Var\left(\rho_{l,t+1}\middle|D_t, M\right) = \sum_{m\in M} Var\left(\rho_{l,t+1}\middle|D_t, m\right) P\left(m\middle|D_t\right) + \sum_{m\in M}\left(E\left(\rho_{l,t+1}\middle|D_t, M\right) - E\left(\rho_{t+1}\middle|D_t, m\right)\right)^2 P\left(m\middle|D_t\right) \tag{18}$$

The term $\sum_{m\in M}\left(E\left(\rho_{t+1}\middle|D_t, M\right) - E\left(\rho_{t+1}\middle|D_t, m\right)\right)^2 \Pr\left(m\middle|D_t\right)$ in the variance calculation is of particular interest as its analog does not arise in model-specific calculations. This term appears in the overall variance calculation because of cross-model variation in expected values. It is a key element in understanding the effects on the precision of an empirical analysis when model uncertainty is present.

While model averaging ideas in economics were initiated in Leamer (1978), the methodology has only recently become widespread; this represents a combination of the increases in computational capacity and theoretical advances. Model averaging is beginning to be employed in a range of economics contexts, most notably economic growth (Brock, Durlauf and West (2003), Doppelhofer, Miller and Sala-i-Martin (2004), Fernandez, Ley and Steel (2001)), finance (Avramov (2002)), forecasting (Garratt et al (2003)) and monetary policy (Brock, Durlauf and West (2003)). An application to a crime context, the deterrent effect of capital punishment, is Cohen-Cole, Durlauf, Fagan, and Nagin (2007). While model averaging methods are still being developed and a number of outstanding theoretical questions still exist,[5] the approach seems very promising.

---

[5]One issue concerning model priors that is worth noting concerns the assignment of priors to similar models. Most of the model averaging literature has employed diffuse priors, i.e. all models are assigned equal prior weights. However, it can be the case that some models in a model space are quite similar, e.g. differ only with respect to a single included variable, whereas others are much more different from the perspective of theoretical or statistical assumptions. In this case, the diffuse prior can be very misleading. Brock, Durlauf, and West (2003) propose ways to construct model priors that mirror the nested structure of modern discrete choice theory, but much more needs to be done. The issue of model similarity is course often ignored in ad hoc analyses of the

iii. *Firearms and Violence* revisited

The potential value of model averaging in providing a constructive way of proceeding with model uncertainty may be seen in the controversy over the analysis of shall-issue concealed weapons laws in the National Academy of Science report *Firearms and Violence,* Wellford, Pepper, and Petrie (2004). This report concluded that

> "…with the current evidence it is not possible to determine that there is a causal link between the right to carry laws and crime rates…It is also the committee's view that additional analysis along the lines of the current literature is unlikely to yield results that will persuasively demonstrate a casual link between right-to-carry laws and crime rates (unless substantial numbers of states were to adopt or repeal right-to-carry laws), because of the sensitivity of the results to model specification." (pg. 150-151)

Committee member James Q. Wilson dissented from this part of the study, on the grounds that the sensitivity to specification found in the report did not account for the sensibility of different models; in particular, he questioned whether the failure of models that excluded socioeconomic control variables to find deterrent effects was of importance in assessing the deterrent effect. Wilson observes

> "Suppose Professor Jones wrote a paper saying that increasing the number of police in a city reduced the crime rate and Professor Smith wrote a rival paper saying that cities with few police officers have low crime rates. Suppose that neither Smith nor Jones used any control variables, such as income, unemployment, population density, or the frequency with which offenders are sent to prison in reaching their conclusions. *If* such papers were published, they would be rejected by the committee out of hand for the obvious reason they failed to produce a complete account of the factors that affect the crime rate." (pg. 270)

---

robustness of findings. Lott (1998) defends his findings on concealed weapons permits by stating "My article with David Mustard and my forthcoming book report nearly 1000 regressions that implied a very consistent effect…" (pg. 242). This claim is of little intrinsic interest without knowing what classes of models these regressions cover; put most simply, the different regression results are not independent, so the number 1000 is not informative.

The committee's rejoinder to Wilson argued that

> "…Everyone (including Wilson and the rest of the committee) agrees that control variables matter, but there is disagreement on the correct set. Thus the facts that there is no way to statistically test for the correct specification and that researchers using reasonable specifications find different answers are highly relevant. Given the existing data and methods, the rest of the committee sees little hope in resolving this fundamental statistical problem." (pg. 273-27)

The disagreement between Wilson and the rest of the NAS committee reflects the absence in the report of an explicit evaluation of how model uncertainty interacts with evidence of shall-issue laws. While the assertion that it is impossible to statistically identify the correct specification of a statistical model is true at some level of generality (though the report is frankly unclear on what is meant by this) this argument is an old one; it is known in the philosophy literature as the Duhem-Quine hypothesis (Quine (1951) is the classic statement) and refers to the idea that all theories are undetermined by available data. And at this level of generality the NAS committee majority's claim is an uninteresting observation with respect to social science research, since it begs the question of the relative plausibility of assumptions.[6] For the dispute at hand, we believe that Wilson is correct in his argument that a model whose specification includes controls suggested by social science theory should receive greater weight than one that does not. Further, these two models are statistically distinguishable. To conclude that one should only regard evidence of a deterrent effect as persuasive if both models produce the same findings makes little sense. The NAS report implicitly suggests that the models without control variables are intrinsically interesting, e.g.

---

[6]The NAS report's suggestion that randomized experiments represent the gold standard for research ignores the assumptions required for their conduct, e.g. integrity of the researcher, accuracy of data collection, etc. An advocate of randomized experiments would presumably dismiss concerns about such factors as implausible. But this is precisely our point.

> "No link between right-to-carry laws and changes in crime is apparent in the raw data…it is only once numerous covariates are included that the…effects…emerge." (pg. 150)

but this remark ignores the classic Simpson's paradox, in which a bivariate relationship has one direction whereas a multivariate relationship does not. The standard example of Simpson's paradox is the positive relationship between admission to hospital and the probability of death.

Model averaging provides a natural way of integrating the information across the alternative specifications considered in the NAS report. As we see it, the committee could have addressed the sensitivity of shall-issue deterrence effects by constructing a set of specifications that included those found in the literature as well as others that are formed by combining the assumptions underlying these models. Intuitively, one thinks of the assumptions that differentiate models as the axes of the model space, and fills the model space out with those combinations of assumptions that are coherent with one another. Averaging over this space would have integrated the information in the different models and indicated whether evidence of a shall-issue deterrent effect is present when one conditions on a model space rather than a particular model. Cohen-Cole, Durlauf, Fagan, and Nagin (2007) illustrate this approach in the context of capital punishment.

Our emphasis on model averaging as a method of addressing assumption discrepancies across studies differs from the perspective developed in Horowitz (2004). Horowitz makes a broad general argument against the utility of observational data analysis (and to be clear, specifically regression analysis) in the presence of model uncertainty. While we, as before, concur that there does not exist an algorithm to infallibly identify the "true" model when the analysis is conducted on a sufficiently broad universe of potential models, it is also the case that different models have different ex ante plausibility and ex post goodness of fit with respect to data. The accumulated body of knowledge that a researcher brings to a given question is a legitimate basis for restricting the class under study or for downweighting certain models. Hence the opposite findings of concealed weapons regressions with and without socioeconomic controls do not warrant equal prior consideration. And we do not know, given our priors, how the relative goodness of fit of the different models under consideration would translate into different posteriors, as the particular models compared in the NAS report are not observationally equivalent.

One answer to our advocacy of model averaging as a tool to address model uncertainty of the type facing the NAS panel is that a given body of empirical studies only captures a small fraction of the universe of potential models (and indeed might represent a measure 0 set). This is certainly a tenable position. But if this position is taken, then it would be irrelevant whether a given body of studies produced similar or conflicting results. If it is then claimed that the degree of consistency in results across models contained in a subspace is informative about the results that would be ascertained were the model space expanded, then it is difficult to see why the relative prior plausibility and relative evidentiary support within an initial model space are not informative as well.

Of course, our discussion of the assumptions that underlie the interpretation of aggregate crime regressions may all be interpreted as examples for Horowitz' claims about the limitations of regression analysis of crime. We do not claim to have thought through the question of how to integrate the different types of model uncertainty we have discussed into a single integrated framework, let alone introduce factors such as the extension of the basic crime model to intertemporal decisionmaking. Our disagreement with Horowitz amounts to our view that, since all empirical work is theory-laden (to use Quine's phrase), empirical models can inform beliefs about policy questions, even though the researcher is aware of untestable or even unappealing assumptions underlying them. The way in which models are used to inform beliefs necessarily requires judgments, which is different from rejecting them altogether. A researcher brings a body of social science and statistical knowledge to bear in the assessment of empirical results; this knowledge matters in assessing the sensitivity of a result to an assumption. In other words, concern about the dependence of an empirical finding on an assumption should depend on what the assumption is.

iv. from model estimation to policy evaluation

This discussion of model uncertainty contains an important limitation in that it does not account for the objectives of a given empirical exercise. Focusing on the use of a single model, it seems intuitive that this model must be correctly specified in order for the model to yield usable findings, so that no distinct considerations arise when one considers the reason why the model is employed. But even in this case, such intuition needs to be qualified.

For example, Horowitz argues that in order to use cross-county data to evaluate the average effect of shall-issue laws, if there are differences between the states[7], so that the crime rate in a county is determined by some set of factors *X*, then in order to identify the effect of the laws "one must use a set that consists of just the right variables and, in general, no extra ones." But as shown Heckman and Navarro (2004) this is true only for a particular set of empirical strategies known as matching,[8] of which linear regression is a special case. Heckman and Navarro demonstrate there are other strategies that are designed to deal with the problem of missing information, in particular the use of control functions, see Navarro (2007) for an overview. The control function approach is based on the idea that the presence of unobservable variables matters only to the extent that their relationship to the observables cannot be determined; for many cases this relationship can be determined. And if so, then other information contained in the omitted variables is irrelevant. The standard example is the Heckman selection correction method in which one adds a "Mills ratio" term to the regression under the assumption of normality but one can be much more general and use semi-parametric methods to estimate the control function term (see Navarro (2007)).

---

[7]Relative to eq. (13), if $\xi^B_{l,t}(1) - \xi^B_{l,t}(0) - \left(\xi^A_{l,t}(1) - \xi^A_{l,t}(0)\right) \neq 0$, then the observables $Z_{l,t}$ and $\bar{X}_{l,t}$ do not constitute the correct set to use when estimating the model since one needs to also control for the effect of the location-time unobservables.

[8]Under matching, endogeneity is solved by assuming that there exists a set of variables such that, conditional on these variables, endogeneity is eliminated. That is, the endogenous variables are not independent of the errors, rather it is assumed they are conditionally independent when the correct set of observable variables (to the econometrician) is conditioned on.

More generally, one cannot decouple the assessment of a model's specification from the objective for which the model is employed. Similarly, any assessment of fragility (or the lack thereof) of empirical claims can only be fully understood with reference to a decision problem.

6. Policy-relevant calculations

i. basic ideas

In this section, we explicitly consider the relationship between statistical models and policy evaluation from a decision-theoretic perspective. The fact that statistical significance levels do not equate to policy statements is well known (see Goldberger (1991) for a nice discussion), our goal here is to suggest some ways of reporting and interpreting results for policy contexts. In making this argument, we are drawing both on classic ideas in statistics, notably Savage (1951) and Wald (1950) as well as recent work in econometrics, e.g. Heckman (2005) and Manski (2005,2006); Brock, Durlauf and West (2003,2007) and Brock, Durlauf, Nason and Rondina (2007) implement some of these ideas. Again, our remarks apply with equal force to work in social sciences other than criminology.

Suppose that the policymaker has a payoff function

$$V\left(\rho_{l,t+1}, \middle| D_t, p, m\right). \tag{19}$$

where $p \in \{A, B\}$ denotes the policy regime and, as before, D_t represents the information available to the policymaker at time t. The conditioning of the utility function on D_t allows for the possibility that the policymaker's preferences depend on aspects of the particular locality since location-specific data $D_{l,t}$ are a subset of D_t. For an expected payoff maximizer, the optimal policy problem is

$$\max_{p \in \{A.B\}} \int V\left(\rho_{l,t+1}, \middle| D_t, p, m\right) \Pr\left(\rho_{l,t+1} \middle| D_t, p, m\right). \quad (20)$$

Eq. (20) implies that the relevant objects for policy analysis are

$$\Pr\left(\rho_{l,t+1} \middle| D_t, A, m\right) \quad (21)$$

and

$$\Pr\left(\rho_{l,t+1} \middle| D_t, B, m\right). \quad (22)$$

as these probabilities fully capture the aspects of the data that are relevant to an expected payoff calculation. Notice that these calculations may not require all aspects of a model to be correctly specified; this was seen in our discussion of the use of matching versus control functions; Heckman (2005) provides a deep analysis of the relationship between models and policy calculations, emphasizing what he denotes as "Marschak's maxim" given ideas found in Marschak (1953):

> "...for many policy questions it is unnecessary to identify full structural models…All that is needed are combinations of subsets of the structural parameters, corresponding to the parameters required to forecast particular policy modifications, which are much easier to identify (i.e. require fewer and weaker assumptions." (pg. 49)

.

One advantage of explicit calculations of posterior densities for policy effects is that they naturally allow one to assess the effects of portfolios of policies. Evidence on the effects of individual policies may be imprecise whereas evidence on the effects of combinations of policies may not be. We do not know whether there are cases of this type in criminology.

Another advantage is that such calculations avoid confusion between the lack of statistical significance of a coefficient for a policy variable and the claim that a policy has no effect; while this is a banal observation, the mistake is often seen. An example of this is found in Lott (1998) who, in evaluating Black and Nagin's (1998) critique of his work, asserts "On the basis of Black and Nagin's comment and our original article, the choice is between concealed handguns producing a deterrent effect or having no effect (one way or the other) on murders and violent crime generally." (pg. 242) Lott's exclusion of crime-enhancing effect of concealed weapons ignores the uncertainty associated with point estimates of the effects. That is, concluding that we cannot reject that the effect is equal to zero does not mean that the effect is indeed zero. One may not be able to reject that it is 0.1 (or -0.1) either. The point estimate is only the most likely (in a particular sense) value of the parameter given the data, not the only possible one. The policy relevant calculation requires assessing the probabilities for different magnitudes of positive and negative effects, which cannot be ascertained from the numbers he (and other participants in this literature) report.

ii. model averaging and policy evaluation

When model uncertainty is present, the optimal policy calculation (20) may be generalized in a straightforward fashion as the policymaker simply conditions on *M* rather than *m*. The relevant calculation in this case is

$$\max_{p\in\{A.B\}} \sum_{m\in M}\left(\int V\left(\rho_{l,t+1}, \middle| D_t , p, m\right)\Pr\left(\rho_{l,t+1}\middle|D_t, p, m\right)\right)\Pr\left(m\middle|D_t\right). \tag{23}$$

If the payoff does not depend on the model, this expression may be rewritten

$$\max_{p\in\{A.B\}} \int V\left(\rho_{l,t+1}, \middle| D_t , p\right)\left(\sum_{m\in M}\Pr\left(\rho_{l,t+1}\middle|D_t, p, m\right)\Pr\left(m\middle|D_t\right)\right) \tag{24}$$

which means the relevant policy calculations are

$$\Pr\left(\rho_{l,t+1}\middle|D_t, A, M\right) \tag{25}$$

and

$$\Pr\left(\rho_{l,t+1} \middle| D_t, B, M\right) \tag{26}$$

which correspond to eq. (15) above.

Eq. (24) indicates an important feature of policy evaluation, namely that, unless the payoff function is model-specific, the identity of the true model does not directly affect policy evaluation. For the purposes of policy evaluation what matters is the distribution of outcomes under alternative policies. Unlike the case of the social scientist, the model has no intrinsic interest to a policymaker; it is simply an additional source of uncertainty in the effects of a policy.

iii. beyond model averaging

Once model uncertainty is involved in policy evaluation, new considerations can arise. One reason for this is that a policymaker may be unwilling to condition decisions on model priors. A form of this is involved in recent theoretical work on decisionmaking under ambiguity, which focuses on how agents should make decisions in environments where certain probabilities cannot be defined. For our purposes, what matters is that in such cases, there do exist alternative ways to engage in policy evaluation. The minimax approach, advocated by Wald (1950) and recently explored in macroeconomic contexts by Hansen and Sargent (2007) evaluates policies by the criterion

$$\max_{p \in \{A.B\}} \min_{m \in M} \int V\left(\rho_{l,t+1}, \middle| D_t, p\right) \Pr\left(\rho_{l,t+1} \middle| D_t, p, m\right) \tag{27}$$

whereas the minimax regret approach, due to Savage (1951) and recently explored in microeconomic contexts by Manski (2005,2006) evaluates policies by the criterion

$$\min_{p \in \{A,B\}} \max_{m \in M} R\left(p, D_t, m\right). \tag{28}$$

Where regret, $R(p,d,m)$ is defined by

$$R(p,d,m) = \max_{p \in P} \left(\int V\left(\rho_{l,t+1}, \middle| D_t, p\right) \Pr\left(\rho_{l,t+1} \middle| D_t, p, m\right) \right) - \int V\left(\rho_{l,t+1}, \middle| D_t, p\right) \Pr\left(\rho_{l,t+1} \middle| D_t, p, m\right) \quad (29)$$

Minimax selects the policy that does best for the least favorable model in the model space. Minimax regret selects the policy with the property that the gap between the model-specific optimal policy and its performance is smallest, when comparisons are made across the model space. The latter is generally regarded as less conservative. Brock, Durlauf, Nason, and Rondina (2007) employ minimax regret in monetary policy evaluation. Manski (2006) applies minimax regret in the context of treatment assignment. An important finding is that optimal treatment rules can be fractional as agents with identical observables receive different treatments. This may be of particular interest in crime policy contexts as it suggests a tradeoff between the fairness and deterrence objectives of punishment that policymakers ought to address.

This discussion allows us to complete the contrast in our perspective with the NAS discussion of the effects of concealed weapons. Suppose that a policymaker possesses minimax preferences with respect to model uncertainty. Then (see Brock, Durlauf and West (2003) for detailed discussion) the presence of different models with opposite policy implications can be used to argue against a change in policy, regardless of the relative goodness of fit of the models. This claim requires that the decision problem of the policymaker be structured so that the status quo law restricts concealed weapons permits, and that a change to shall-issue requires an increase in the policymaker's minimax-based utility relative to the status quo. This seems to be the way in which the NAS report is evaluating shall-issue concealed weapons laws. But minimax preferences do not, by themselves, generate the NAS conclusion as they need to be coupled with a presumption against shall-issue.

Therefore, a recommendation we make for policy evaluation studies such as *Firearms and Violence* is that claims about the robustness or fragility of various findings be evaluated with respect to different loss functions, with particular attention to minimax and minimax regret calculations as supplements to the standard Bayesian ones. We would also observe that it is possible to couple minimax and Bayesian expected payoff considerations, see Brock, Durlauf, and West (2003) for discussion.

7. Example: capital punishment

In this section, we describe how some of our arguments matter by considering their implications for the study of capital punishment. For most of the discussion we focus on the empirical study of deterrent effects by Dezhbakhsh, Rubin, and Shepard (2003). 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). DRS does not make general policy claims about the desirability of the death penalty, so we will also focus on Sunstein and Vermeule (2005), who have argued that evidence in favor of a capital punishment deterrence effect can make it morally obligatory. This latter paper, while philosophical in orientation, focuses on the role of deterrence evidence in justifying the death penalty and so is related to our discussion of policy evaluation.

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

$$\rho_{l,t} = \alpha + Z_{l,t}\beta + P_{l,t}\left(C\right)\beta_C + P_{l,t}\left(S\middle|C\right)\beta_S + P_{l,t}\left(E\middle|S\right)\beta_E + \kappa_{l,t} \tag{30}$$

where

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

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

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

i. microfoundations

From the perspective of our first argument, that aggregate models should flow from aggregation of individual behavioral equations, 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 DRS, the potential outcomes are

$$NC = \text{not caught,}$$
$$CNS = \text{caught and not sentenced to death,}$$
$$CSNE = \text{caught, sentenced to death, and not executed,}$$
$$CSE = \text{caught, sentenced to death and executed.}$$

The expected utility of a person who commits a murder is therefore

$$\begin{aligned}&\Pr_{l,t}\left(NC\right)u_{i,t}\left(NC\right)+\Pr_{l,t}\left(CNS\right)u_{i,t}\left(CNS\right)+\\&\Pr_{l,t}\left(CSNE\right)u_{i,t}\left(CSNE\right)+\Pr_{l.t}\left(CSE\right)u_{i,t}\left(CSE\right).\end{aligned} \tag{31}$$

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

$$
\begin{gathered}
\left(1-\Pr_{l,t}(C)\right)u_{i,t}(NC)+ \\
\left(1-\Pr_{l,t}(S|C)\right)\Pr_{l,t}(C)u_{i,t}(CNS)+ \\
\left(1-\Pr_{l,t}(E|S)\right)\Pr_{l,t}(S|C)\Pr_{l,t}(C)u_{i,t}(CSNE)+ \\
\Pr_{l,t}(E|S)\Pr_{l,t}(S|C)\Pr_{l,t}(C)u_{i,t}(CSE)
\end{gathered}
\tag{32}
$$

A comparison of (32) with (30) reveals that the DRS specification does *not* derive naturally from individual choices since the conditional probabilities in (30) interact with each other in the calculation of expected utility as in (32) . If one substitutes in a linear representation of the utility functions for the different outcomes, it is evident that (32) cannot produce an aggregate crime equation in which the conditional probabilities additively as in (30); a full analysis may be found in Durlauf, Navarro, and Rivers (2007). Put differently, the effect of the conditional probability of execution given a death sentence on behavior cannot be understand separately from the effects of the conditional probability of being caught and being sentenced to death if caught.

Therefore, we conclude that the DRS specification fails to properly model the implicit decision problem involved in homicides. Their analysis is based on a misspecification of the implications of their assumed behavioral model.

ii. aggregation

Our aggregation discussion suggests how correlations can arise between regressors and model errors because of unobserved location characteristics. DRS only instrument the conditional crime probabilities in (30), 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 (in the statistical sense).

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 $\xi_{l,t}\left(\omega_{i,t}\right)$ can vary over time.

iii. policy effect estimation

Our discussion of policy effect evaluation also calls into question the DRS analysis. DRS assume that the fluctuations in their arrest, sentencing and execution probabilities constitute the full set of changes in policies across time periods. This seems problematic. The decision to commit a homicide, under the economic model of crime, depends on entire range of penalties and their associated probabilities. As noted in Fagan (2006) and elsewhere, changes in the rates at which murderers are sentenced to life imprisonment without parole, for example, are not accounted for in DRS, or other capital punishment deterrence studies. Hence these studies suffer from an obvious omitted variables problem.

This argument can be pushed farther. As shown in Gelman, Liebman, West and Kiss (2004), the probability that a given death sentence will be overturned by a state or federal appeals court is at least 2/3. These authors also find that only 5% of the death sentences between 1975 and 1993 led to the eventual execution of those sentenced. Relative to our choice model, the Gelman, Liebman, West and Kiss findings mean that the reintroduction of capital punishment in a state, on average, substantially increases this means that an increases the probability that the commission of murder leads to the outcome *CSNE*, i.e. arrested, sentenced to death, and *not executed*. Since exonerations are rare, it is reasonable to conjecture that murderers with outcome *CSNE* experience longer prison sentences than they would have had they not been sentenced to death. This suggests that periods in which criminals face higher probabilities of capital sentencing and actual execution are also associated with longer prison sentences. Yet this increase is not reflected in the DRS regression. Put differently, if an increase in the conditional probability of a death sentence given arrest, $\Pr_{l,t}\left(S\middle|C\right)$, is associated with an increase in $\Pr_{l,t}\left(CSNE\right)$, then it is no longer clear what it means to say that a DRS-type regression provides evidence on the effects of capital punishment; does an increase in long prison sentences because of death sentences followed by reversals correspond to what is understood to be the deterrent effect of capital punishment?

iv. model uncertainty

Donohue and Wolfers (2005) (denoted as DW) have argued that the DRS findings of strong deterrence effects are fragile as small changes in their baseline specification can lead to an absence of a statistically significant effect or even evidence that a larger number of executions is associated with a larger number of murders. Specifically, DW show that the DRS findings change when one alters the lag structure for the instrumental variables used for the punishment probabilities as well as when one drops California and Texas from the sample. The latter may be interpreted as a change in the assumption that all states are exchangeable with respect to the model employed by DRS.

Cohen-Cole, Durlauf, Fagan and Nagin (2007) attempt to adjudicate the differences between DRS and DW by treating the problem as one of model uncertainty. To do this, a space of potential models was generated using different combinations of the assumptions found in the two papers. Cohen-Cole, Durlauf, Fagan, and Nagin conclude that the evidence for deterrence in the sample studied by DRS is quite weak.

v. policy-relevant calculations

Following our general discussion, the statistical significance of the capital punishment variables in a murder regression does not produce the appropriate information needed to make policy comparisons. This has implications for the way such evidence is employed in death penalty debates. Sunstein and Vermeule (2005) argue that evidence of a deterrent effect can produce a moral case for capital punishment, in that the decision of a government to fail to implement a life saving policy is equivalent to the decision to implement a policy that costs lives.

Relative to Sunstein and Vermeule (2005), we would note several points. First, while they do state that their argument is conditional on evidence of a deterrence effect, recent studies such as DRS do not provide that support.[9] These studies suffer from a failure to construct appropriate counterfactuals with respect to different punishment regimes and fail to address model uncertainty.

Second, the Sunstein and Vermeule analysis treats the expected number of lives saved as the variable of interest to the policymaker; in DRS this value is a function of the estimated parameter β_E in (30) . The expected number of lives saved is not necessarily sufficient in describing a policymaker's utility function, even if this function is a monotonically increasing function of the number of lives saved. As such, their attention to this figure is analogous to making a utilitarian as opposed to a welfarist calculation, see Sen (1979). While Sunstein and Vermeule would presumably respond that they are assuming that the precision associated with estimates of the expected number of lives saved is high, precision needs to be defined with respect to the policymaker's utility function.

Third, the fragility of deterrence claims to modeling assumptions, as demonstrated by DW and extended in Cohen-Cole, Durlauf, Fagan, and Nagin (2007), raises the issues we have discussed with respect to decisionmaking under ambiguity and the evaluation of policies when one does not wish to base them on a choice of model priors. The sensitivity of deterrence estimates to model specification renders their use problematic for expected value calculations (as well as more sophisticated policymaker utility calculations) without a justification of the choice of priors. Our impression of the philosophy literature is that the issue of policy evaluation under ambiguity has generally not been discussed, although Gaus (2006) makes an interesting argument in favor of following principles rather than expected effect calculations when assessing policies when the effects of policies are associated with substantial uncertainty.

To be clear, none of this means that Sunstein and Vermeule (2005) are incorrect in their conclusions about the ethical implications of a certain deterrent effect for a policymaker or that the death penalty is either moral or immoral per se. Rather, our claim

[9]At the same time they also state that

> "The foundation of our argument is a large and growing body of evidence that capital punishment may well have a deterrent effect, possibly a quite powerful one…The particular numbers do not much matter." (p. 706).

is that the policy implications of the uncertainty associated with deterrence effects cannot be assessed outside of the policymaker's preferences.

8. Conclusions

In this paper, we have described some issues we regard as important in the econometric study of crime: microfoundations, aggregation, counterfactual analysis and policy evaluation. We have tried to make clear the various assumptions that must be maintained to interpret aggregate crime regressions with respect to individual behavior and have emphasized how standard uses of these regressions to evaluate policy presuppose a number of assumptions. In light of disagreements about these assumptions, which ultimately underlie claims of fragility or robustness of an empirical result, we have outlined some ways of using model averaging methods and statistical decision theory to make progress. Throughout, we have emphasized the role of judgment in empirical work, for which no algorithm exists.

One response to the discussion in this paper would be to search for alternative ways of uncovering aggregate criminological facts. The critiques we have raised are part of the source of interest in so called natural experiments, in which an exogenous event of some type allows a comparison of aggregate crime outcomes; see Levitt (1996) for a nice example. In fact this is the approach that Horowitz (2004) advocates.

While we are sympathetic to such efforts, it is important to understand that these methods also require assumptions. An experiment of the type proposed by Horowitz with respect to shall-issue weapons permit laws – randomized legalization across states – would, if one is to use the findings to inform policymakers, require assumptions about 1) the degree to which criminals can alter the location in which crimes are committed, 2) the nature of migration by potential criminals across state boundaries both before the experiment and in response to it, 3) the effect on the current crime choices of potential criminals of knowledge that an experiment which may affect future laws in their state of residence is being conducted, etc.[10] Also, the translation of findings from such an experiment into a recommendation for those states that did not implement the policy requires exchangeability assumptions on the states. Does one assume that deterrence effects common across states? If state-level deterrent effects are heterogeneous, how is this heterogeneity to be modeled, via random effects, varying coefficients or some other method? Randomized experiments cannot avoid the need for judgments; as emphasized in Heckman (2005), judgment is intrinsic to scientific inquiry. Hence, we do not see good reasons to rank order regressions and natural experiments in terms of their relative utility as means of understanding crime. Each has a contribution to make in criminological research.

[10]Heckman (2005) and Manski (2007) provide general discussions of the limitations of experiments, with a particular focus on the assumptions implicit in treatment effect analysis; Heckman and Navarro (2004) compares the strengths and weaknesses of different empirical strategies for uncovering the determinants of individual choice.

Bibliography

Avramov, D. (2002). “Stock Return Predictability and Model Uncertainty,” *Journal of Finance*, 64, 423-458.

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

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

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

Becker, G., (1993), “Nobel Lecture: The Economic Way of Looking at Behavior,: *Journal of Political Economy,* 101, 3, 385-409.

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 Economics and Reality,” *World Bank Economic Review*, 15, 229-272.

Brock, W., S. Durlauf, J. Nason, and G. Rondina, (2007), “Simple versus Optimal Rules as Guides to Policy,” *Journal of Monetary Economics*, 54, 5, 1372-1396.

Brock, W., S. Durlauf, and K. West (2003), “Policy Analysis in Uncertain Economic Environments (with discussion).” *Brookings Papers on Economic Activity*, 1, 235-322.

Brock, W., S. Durlauf, and K. West, (2007), “Model Uncertainty and Policy Evaluation: Some Theory and Empirics,” *Journal of Econometrics*, 136, 2, 629-664.

Cantor, D and K. Land, (1985), “Unemployment and Crime Rates in the post-World War II United States: A Theoretical and Empirical Investigation,” *American Sociological Review*, 50, 317-332.

Cohen-Cole, E., S. Durlauf, J. Fagan, and D. Nagin, (2007), “Model Uncertainty and the Deterrent Effect of Capital Punishment,” mimeo, University of Wisconsin.

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

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

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

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

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

Durlauf, S., S. Navarro, and D. Rivers, (2007), "Notes on the Econometrics of Deterrence," mimeo, University of Wisconsin.

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.

Fagan, J., (2006), "Death and Deterrence Redux: Science, Law and Causal Reasoning on Capital Punishment," *Ohio State Journal of Criminal Law*, 4, 1, 255-320.

Fernandez, C., E. Ley and M. Steel (2001), "Model Uncertainty in Cross-Country Growth Regressions," *Journal of Applied Econometrics*, 16, 5, 563-76.

Garratt, A., K. Lee, M. H. Pesaran, and Y. Shin, (2003) "Forecasting Uncertainties in Macroeconometric Modelling: An Application to the UK Economy," *Journal of the American Statistical Association* 98, 464, 829-838.

Gaus, G., (2006), "Social Complexity and Evolved Moral Principles," mimeo, University of Arizona.

Gelman, A. J. Liebman, V. West, and A., (2004), “A Broken System: The Persistent Patterns of Reversals of Death Sentences in the United States,” *Journal of Empirical Legal Studies*, 1, 2, 209-261.

Goldberger, A., (1991), *A Course in Econometrics*, Cambridge: Harvard University Press.

Greenberg, D., (2001), “Time Series Analysis of Crime Rates,” *Journal of Quantitative Criminology*, 17, 4, 291-327.

Hansen, L. and T. Sargent, (2007), *Robustness*, Princeton: Princeton University Press.

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, (2004), “Using Matching, Instrumental Variables, and Control Functions to Estimate Economic Choice Models, *Review of Economics and Statistics*, 86, 1, 30-57.

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 Academy Press.

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

Levitt, S., (1996), “The Effect of Prison Population Size on Crime Rates: Evidence from Prison Overcrowding Litigation,” *Quarterly Journal of Economics*, 111, 319-51.

Levitt, S., (2004), “Understanding Why Crime Fell in the 1990’s: Four Factors that Explain the Decline and Six that Do Not,” *Journal of Economic Perspectives*, 18, 1, 163-190.

Lott, J. and D. Mustard, (1997), “Crime, Deterrence, and Right-to-Carry Concealed Handguns,” *Journal of Legal Studies*, 26, 1, 1-68.

Lott, J., (1998), “The Concealed-Handgun Debate,” *Journal of Legal Studies*, 27, 1, 221-291.

Manski. C., (1995), *Identification Problems in the Social Sciences*, Harvard: Harvard University Press.

Manski, C., (2005), “Statistical Treatment Rules for Heterogeneous Populations,” *Econometrica*, 72, 1221-1246.

Manski, C., (2006), “Minimax-Regret Treatment Choice with Missing Outcome Data,” *Journal of Econometrics*, forthcoming.

Manski, C., (2007), *Identification for Prediction and Decision*, Cambridge: Harvard University Press, in press.

Marschak, J., (1953), “Economic Measurements for Policy and Prediction,” in *Studies in Econometric Method*, W. Hood and T. Koopmans, ed. New York: John Wiley.

Matzkin, R., (1992), “Nonparametric and Distribution-Free Estimation of the Binary Threshold Crossing and the Binary Choice Models,” *Econometrica*, 60, 2, 239-270.

Navarro, S., (2007), “Control Functions,” in *New Palgrave Dictionary of Economics*, revised edition, L. Blume and S. Durlauf, eds., London: Palgrave-Macmillan, in press.

Plassmann, F. and J. Whitley, (2003), “Confirming “More Guns, Less Crime,” *Stanford Law Review*, 55, 1313-1369.

Quine, W. V. O., (1951), “Two Dogmas of Empiricism,” *Philosophical Review*, 60, 20-43.

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.

Savage, L. J., (1951), “The Theory of Statistical Decision,” *Journal of the American Statistical Association*, 46, 55-67.

Sen, A., (1979), “Utilitarianism and Welfarism,” *Journal of Philosophy*, 76, 9, 463-489.

Stoker, T. (1993), “Empirical Approaches to the Problem of Aggregation over Individuals,” *Journal of Economic Literature*, 31, 4, 1827-1874.

Sunstein, C. and A. Vermeule, (2005), “Is Capital Punishment Morally Required? Acts, Omissions and Life-Life Tradeoffs,” *Stanford Law Review*, 58, 3, 703-751.

Wald, A., (1950), *Statistical Decision Functions*, New York: John Wiley.

Wellford, C. J. Pepper, and C. Petrie, eds. (2004), *Firearms and Violence*, Washington DC: National Academy of Sciences.