

New Claims about Executions and General Deterrence: Déjà Vu All Over Again?

Richard Berk *

March 23, 2005

Abstract

A number of papers have recently appeared claiming to show that in the United States executions deter serious crime. There are many statistical problems with the data analyses reported. This paper addresses the problem of “influence,” which occurs when a very small and atypical fraction of the data dominate the statistical results. The number of executions by state and year is the key explanatory variable, and most states in most years execute no one. A very few states in particular years execute more than 5 individuals. Such values represent about 1% of the available observations. Re-analyses of the existing data are presented showing that claims of deterrence are a statistical artifact of this anomalous 1%.

I. Introduction

Research on the possible deterrent value of capital punishment has a long history (Sutherland, 1925; Zimring and Hawkins, 1973; Gibbs, 1975, Ehrlich, 1975; 1977; Sellin 1980). Of late, a cluster of papers by economists has

*Department of Statistics, 8125 Mathematical Sciences Building, UCLA, Los Angeles, 90095-1554 (berk@stat.ucla.edu). Thanks go to Sam Gross, Jeff Fagan, Michael Radelet, and Rick Lempert for a number of very helpful comments on earlier drafts of this paper. Work on this paper was supported by a grant from the National Science Foundation: SES-0437169 — Ensemble Methods for Data Analysis in the Behavioral, Social and Economic Sciences

appeared attempting to rebut claims that earlier econometric analyses reporting deterrence effects were subject to fatal model specification errors (Ehrlich and Liu, 1999; Cloninger and Marchesini, 2001; Mocan and Gitting, 2003; Shepherd, 2004; Zimmerman, 2004).¹ In this work, the death penalty is treated as an intervention, either as a binary indicator variable or, more commonly, as the number of executions. Various measures of violent crime are the outcomes. A wide variety of control variables are used. Mocan and Gitting (2003), for example, claim to find strong deterrent effects.

Such research touches on many vexing issues about which there continues to be widespread controversy (Zimring, 2003, Forst, 2004; Blume et al., 2004; Gelman et al., 2004). But as an empirical matter, the research is necessarily based on observational data. It follows that there are a host of problems in trying to make credible casual inferences (Rosenbaum, 2002). These range from the conceptual leap of treating observational data as an experiment to a large number of nuts-and-bolts statistical difficulties. (See, for example, Box, 1976; Leamer, 1978; Rubin, 1986, Freedman, 1987; Manski, 1990; Heckman, 1999; Breiman, 2001, Berk, 2003.)

I have recently obtained the data used in papers authored by Mocan and others.² The full range of statistical concerns is surely relevant, but these data have special properties that make them especially problematic. Here I will focus on the nature of the “treatment.” What can be learned about the impact of the death penalty for the United States as a whole when a very few jurisdictions account for the vast majority of executions?

II. A Look at the Treatment and Response Variables

The data are a pooled cross-section time series of 50 states over 21 years (1977-1997), so that there are 1050 observations in all. Each observational unit is the year within a state. The canonical explanatory variable is the

¹For an evenhanded and thorough critique of the earlier deterrence claims see Klein et al. (1978).

²Jeffrey Fagan obtained the data from Mocan for replication and reanalysis and passed a copy of the data along to me. By and large, I have taken the data at face value as best I understand them, with important assistance from Fagan on how the variables are defined. I have also spot checked some of the observations. But I cannot guarantee that the data are fully what they purport to be or that there are no errors in the data themselves.

number of executions each year for each state. The number of homicides per state per year and the homicide rate per 1,000 are key response variables.

Mocan and Gittings construct their basic execution variable defining a year as October through September. Their rationale is that executions near the end of the usual calendar year cannot be expected to affect crime in that same calendar year, given that a year is the temporal unit. Thus, homicides in 1980, for example, respond to executions from January through September of 1980 plus executions from October through December 1979. In practice, however, they experiment with one and two year lags of their October-to-September execution variable. For the empirical analyses to follow, I will work with a one year lag. Mocan and Gittings try other lags and several transformations of their basic executions variable. But for the issues I want to raise, concentrating on the number of executions lagged by one year is for now sufficient. Alternative lags and transformations will be briefly addressed later. The one year lag for execution means that there are 1000 observations to analyze, not 1050.³

For the number of executions, the mean is .35, implying that on the average each state executes one prisoner about every three years. But this is very misleading. The standard deviation is 1.35, which given the mean and the lower boundary of 0, is a strong indication of skewing.

Figure 1 shows the empirical distribution for the number of executions. The “dose” ranges from 0 to 18, with 859 of the 1000 values (86%) equal to 0. As a result, the median is also 0. There are 78 values (8%) equal to 1. There are but 11 values (1%) larger than 5, ranging from 7 to 18 executions. Clearly, the distribution is highly skewed, and the mean is dominated by a few extreme values. Most states in most years execute no one.⁴

The risks of working with highly skewed explanatory variables are well known and thoroughly discussed in any number of accessible statistics textbooks (e.g., Cook and Weisberg, 1999). The farther values for the explanatory variables fall from the centers of their distributions, the more “leverage” they have. When such values tend to be paired with values for the response variable that fall some distance for the center of the response distribution, leverage becomes “influence.” The usual objective functions employed in the

³One year’s worth of data is necessarily lost when executions is lagged by one year. All of the observations for the response variables in 1977 are eliminated.

⁴Figure 1 is for the number of executions lagged by one year. For the unlagged number of executions, there are three more values larger than 5. The largest, for the most recent year in Texas, is equal to 29. If anything, the skewing is increased.

Number of Executions Per Year Offset by a Quarter

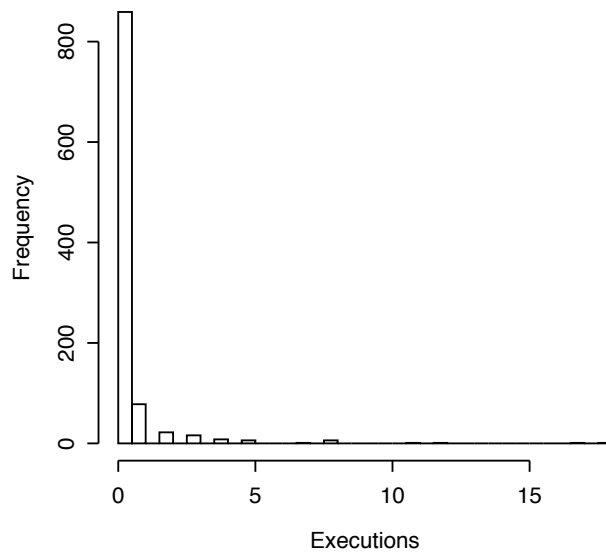


Figure 1: Empirical Distribution of the Number of Executions Lagged by One Year — *Note:* There is strong evidence of skewing

fitting, such as a quadratic, weight such observations very heavily. These observations can then dominate the results. Removing influential observations, or down-weighting them, can dramatically change the conclusions reached.

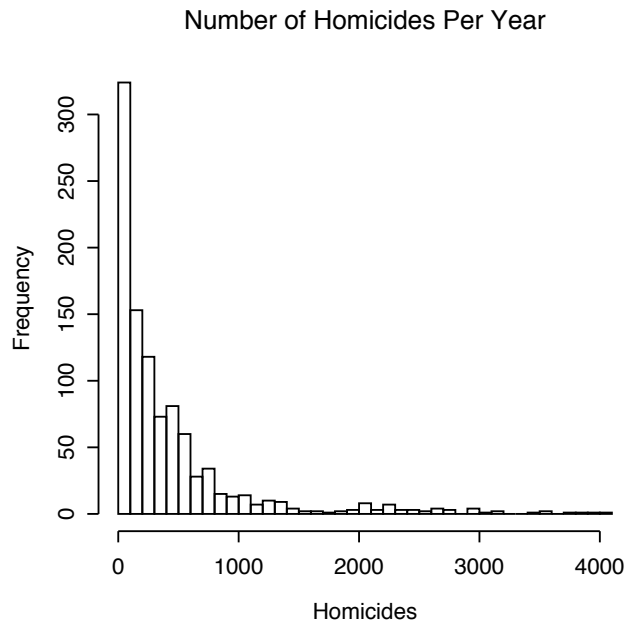


Figure 2: Empirical Distribution of the Number of Homicides — *Note:* There is strong evidence of skewing

However, influential observations are not necessarily “outliers.” An outlier is not just atypical, but is located some distance from the mass of the data. Such observations are often treated as qualitatively different. When influential observations are also outliers, there is even more grounds for concern.

It is clear from Figure 1 that there are several observations with large leverage that are also quite properly labeled outliers. We will soon consider whether they are also influential.

Figure 2 shows the empirical distribution for one of the two response variables: the number of homicides per year. The mean is 420, and the standard deviation is 607. The pattern in Figure 2 is much the same as for the number of executions, although a bit less extreme. One implication is

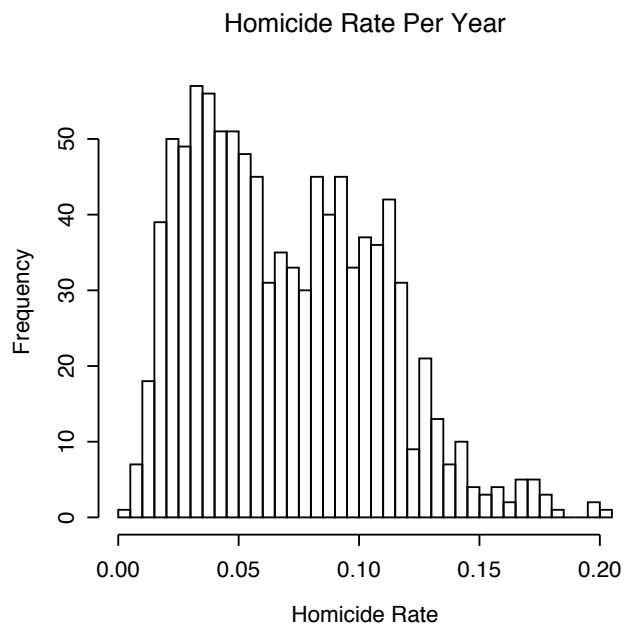


Figure 3: Empirical Distribution of the Homicide Rate
— *Note:* There is moderate evidence of skewing

that there is real potential for a few observations to be very influential and to be outliers as well.

Figure 3 shows the empirical distribution for a second response variable: the homicide rate per 1,000 people. The mean is .06, and the standard deviation is .04. Skewing is reduced substantially, and there is far less evidence of “daylight” between a few extreme values and the mass of the data. Although influence remains a genuine issue, the influential observations are less likely to include outliers for the homicide rate.

In summary, even before one moves beyond these histograms, there is good reason to worry about impact of highly skewed distributions for variables that are central to the analysis. A very few observations may dominate the results, and some of these observations may also be outliers.

III. Some Bivariate Relationships

Consider now Figure 4. The vertical axis is in units of the number of homicides per year. The horizontal axis is in units of the number of executions lagged by one year. Along the horizontal axis is a rug plot showing how the data cluster by the number of executions. The solid line represents the fitted values produced by a B-spline smoother within the generalized additive model (GAM). The dotted lines are the approximate 95% confidence interval.⁵

The generalized additive model (GAM) is a form of regression analysis in which each predictor (here, only one) is allowed to have its own functional relationship with the response variable. The functional form is determined empirically (Hastie and Tibshirani, 1990).⁶ Here, GAM is used solely as a

⁵The data are either a population or a convenience sample. Any statistical inference, therefore, must rely on model-based sampling (Thompson, 2002: section 7.2). No justification for model-based sampling is provided by Mocan and Gittings or by the authors of any of the papers cited earlier. Consequently, it is not clear that statistical inference is appropriate (Berk, 2003). Nevertheless, tests and confidence intervals will be reported, consistent with the practices in this literature.

⁶A bit more formally:

$$E(Y|\mathbf{X}) = \sum_{j=1}^p f_j(X_j), \tag{1}$$

where \mathbf{X} is a set of p predictors. Each function of \mathbf{X} , f_j , is determined by a smoother. Locally weighted regression (LOWESS) and spline functions are popular options. B-splines can be thought of as a computationally effective way to smooth a scatter plot with cubic

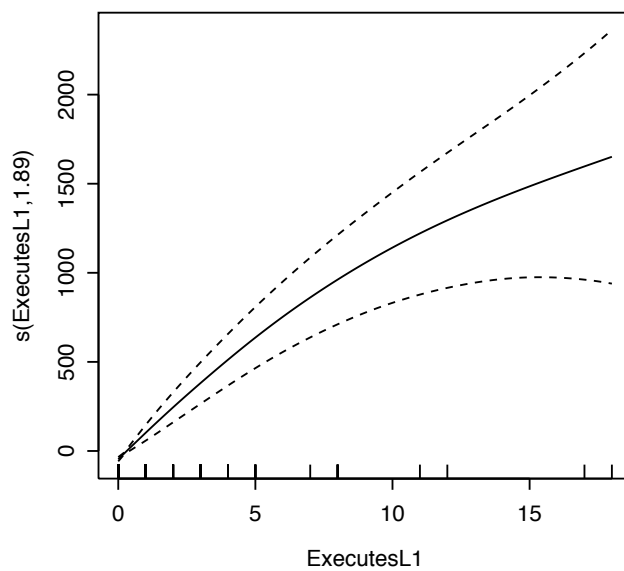


Figure 4: Number of Homicides as a Function of the Number of Executions Lagged by One Year (Explained Deviance = 7.2%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 1.89. The dotted lines contain the approximate 95% confidence interval. The overall relationship between the number of homicides and the lagged number of executions is positive.

descriptive tool.

Because the outcome is a count, it may be reasonable as first approximation assume that the conditional distribution is Poisson. But then, the output is somewhat more difficult to interpret. Proceeding with a gaussian conditional distribution leads to essentially the same descriptive results and is more easily understood.⁷

Associated with the fitted values are the effective degrees of freedom used up by the smoother, in this case 1.89 (See Berk, 2005 for an accessible discussion). Thus, the fitted values are a smooth function of the number of executions lagged by one year, using up almost 2 degrees of freedom. A key implication is that the fitted values are a non-linear function of the number of executions.

The relationship is positive overall. With more executions, there are more homicides about a year later. If the smoother is capturing anything causal, the effects can lead to increases of over 1000 homicides. However, the confidence interval is very wide beyond about 5 executions (where there are almost no data) and allows for the possibility that the relationship becomes flat or even slightly negative.⁸

Figure 5 repeats the analysis, but using the homicide rate as the response. Standardizing for population size is an effort to control for the obvious fact that with more people, there are more potential perpetrators and more potential victims. A gaussian conditional distribution is assumed, although once again, a LOWESS smoother leads to about the same story.

The relationship is positive over the range of executions where the mass of data are located, and only turns slightly negative for execution numbers greater than 5. But now the confidence interval starts to have real bite. It is so wide beyond about 5 executions that it is not clear what the overall trend may be. However, even if the downturn is real, the net effect over the entire curve is positive. Going from 0 executions to 15, increases the homicide rate substantially. The mean homicide rate is .06 (per 1000), so that an increase

polynomials. The generalized linear model (GLM) is a special case in which all of the functions are linear. GAM allows for the same selection of disturbance distributions as GLM. It also assumes the same canonical link functions.

⁷The results in Figure 4 are basically the same using a LOWESS smoother, which makes no assumptions about the conditional distribution of the response.

⁸By construction, the fitted values for a single predictor have a mean of zero. This may be hard to see in Figure 4 unless one keeps in mind that the vast majority of observations have no executions.

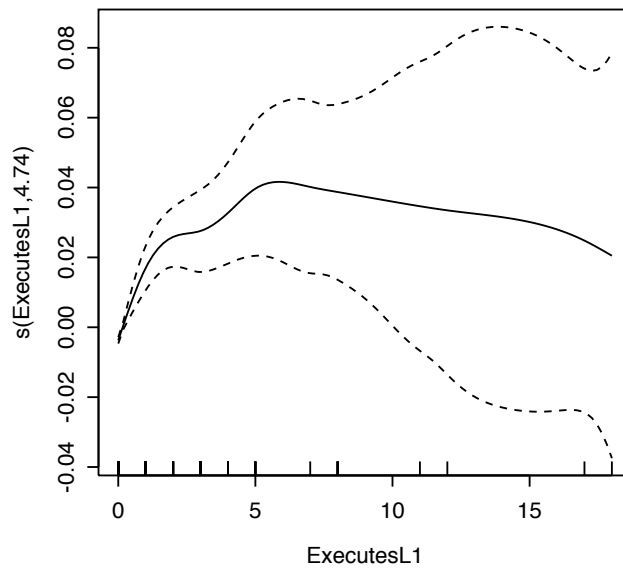


Figure 5: Homicide Rate as a Function of the Number of Executions Lagged by One Year (Explained Deviance = 7.4%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 4.74. The dotted lines contain the approximate 95% confidence interval. The relationship is between the homicide rate and the lagged number of executions is generally positive for up to 5 executions. Above 5 executions, it is difficult to tell.

of about .02 is important.

IV. Some Multivariate Relationships

What happens to Figures 4 and 5 when a more concerted effort is made to control for potential confounders? In particular, suppose one wanted covariance adjustments for the factors that on the average over the 20 years led to more homicides in some states than others (e.g, California versus Maine). In effect, each state would need its own intercept within the GAM formulation. One can do this in a fixed effects manner by simply adding an indicator variable for each of the state (less one to avoid linear dependence).⁹

Figure 6 shows the adjusted relationship between the number of homicides and the number of executions lagged by one year. The fit is excellent; 97% of the deviance can be accounted for. The results cannot be criticized for lack of fit, and it is clear that most of the variation in homicides is simply a function of the average number of homicides in each state from 1977 to 1997. In other words, once one accounts on the average for the number homicides state by state over that period, one has extracted from the data about all of the information there is. Knowing the number of executions adds virtually nothing. Indeed, dropping executions from the model reduces the deviance accounted for from 97.0% to 96.3%.

If, nevertheless, one is curious about the role of executions, there would seem at first to be evidence of a substantial negative relationship consistent with deterrence. However, for 5 executions or less, the relationship is flat or slightly positive overall. Only for more than 5 executions is the net effect negative.¹⁰ So any evidence for a deterrent effect on the number of homicides depends on 11 observations out of 1000.

Figure 7 shows the parallel analysis for the homicide rate. The fit is not quite as good, but with 87.9% of the deviance accounted for, still excellent. If executions is dropped from the model, the deviance accounted for drops to 87.1%. When allowance is made for average differences across states in the

⁹A random effects approach for the states would save some degrees of freedom, but would be require several heroic assumptions. The fixed effects approach is far more robust, and with 1000 observations, giving up 49 degrees of freedom for the states is not a major concern.

¹⁰The “wiggles” beyond 5 executions should not be taken at face value. The data are far too sparse. Only one or two observations are responsible for each bend.

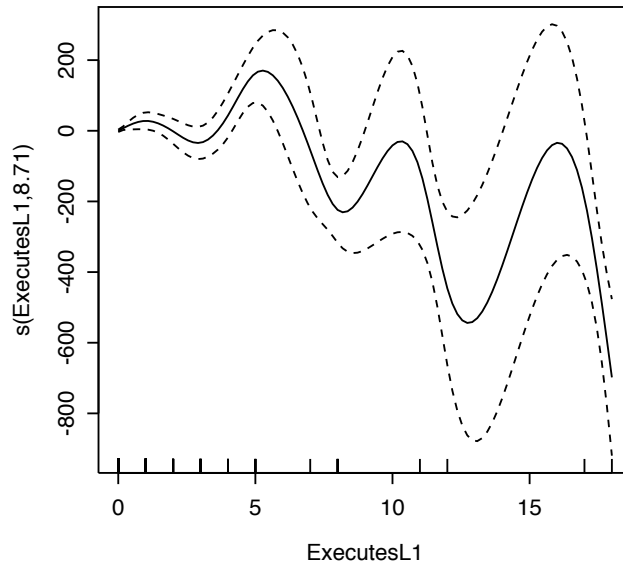


Figure 6: Number of Homicides as a Function of the Number of Executions Lagged by One Year, Controlling for State (Explained Deviance = 97.0%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 8.71. The dotted lines contain the approximate 95% confidence interval. There is no consistent relationship between the number of homicides and the lagged number of executions when there are five execution or less. The apparent negative relationship when there are six executions or more is based on only eleven observations out of one thousand and cannot be taken at face value.

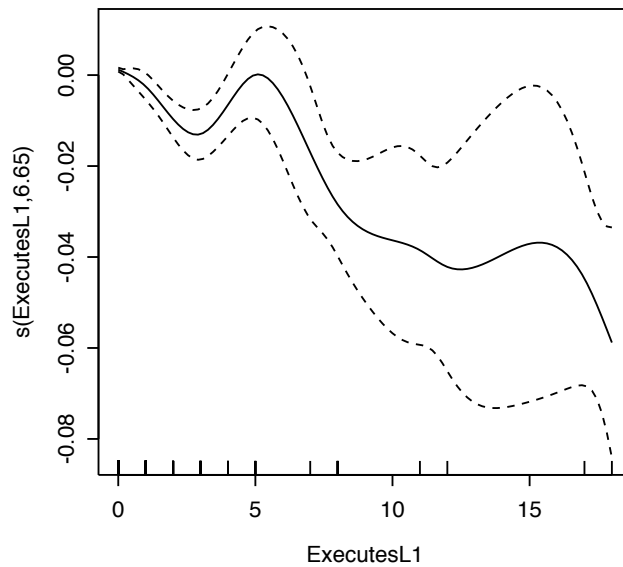


Figure 7: Homicide Rate as a Function of the Number of Executions Lagged by One Year, Controlling for State (Explained Deviance = 87.9%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal 6.65. The dotted lines contain the approximate 95% confidence interval. There is no consistent relationship between the homicide rate and the lagged number of executions when there are five execution or less. The apparent negative relationship when there are six executions or more is based on only eleven observations out of one thousand and cannot be taken at face value.

homicide rate, one has most of the story.

At first glance there appears to be an overall downward trend. However, a close look at the graph again reveals effectively no relationship when there are 5 executions are less. It is the same 11 observations out of 1000 that are driving the negative relationship. And consistent with this tiny sub-sample, the 95% confidence interval is very large precisely where the substantial declines in the homicide rate are found. Note that the negative relationship could be essentially flat beyond 7 executions. But even giving the benefit of the doubt, it is apparent that for the vast majority of states in the vast majority of years, there is no evidence of a negative relationship between executions and homicides.

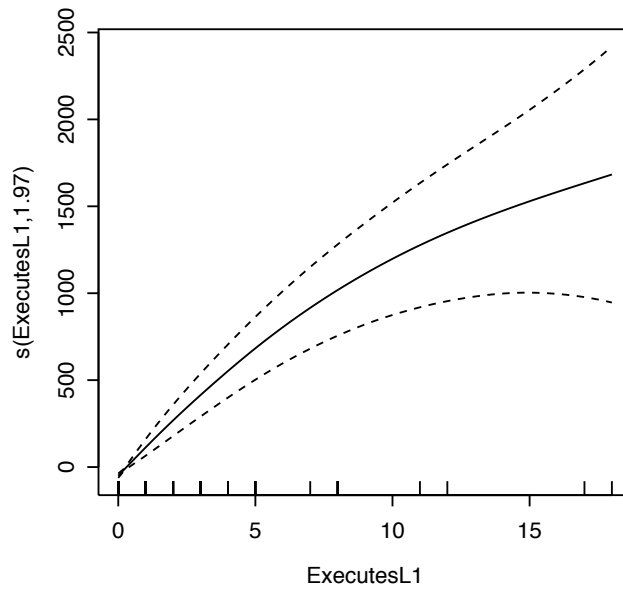


Figure 8: Number of Homicides as a Function of the Number of Executions Lagged by One Year, Controlling for Year (Explained Deviance = 7.8%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 1.97. The dotted lines contain the approximate 95% confidence interval. The relationship between the number of homicides and the lagged number of executions is positive.

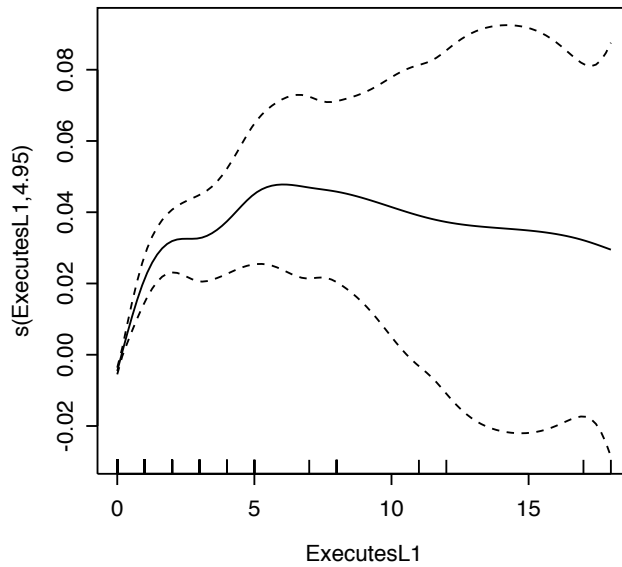


Figure 9: Homicide Rate as a Function of the Number of Executions Lagged by One Year, Controlling for Year (Explained Deviance = 11.8%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 4.95. The dotted lines contain the approximate 95% confidence interval. The relationship between the homicide rate and the lagged number of executions when there are five execution or less positive. The apparent negative relationship when there are six executions or more is based on only eleven observations out of one thousand and cannot be taken at face value. The confidence interval implies that the sign cannot be credibly determined.

So far, the temporal dimension has been ignored. Figures 8 and 9 show the results when indicator variables for years are used instead of indicator variables for state. These analyses control for average differences between years (over states) in the number of homicides and the homicide rate respectively. In effect, the year indicator variables control for national trends in the number of homicides and the homicide rate. But because year to year variation does not appear to be nearly as important as state to state variation, the results are very much like those reported earlier when no covariates were used.

Figures 10 and 11 show the results when indicator variables for states *and* years are used as indicator variables. These analyses control for average differences between states (over years) and between years (over states). Because variation between states dominates the story, these figures look much the same as Figures 6 and 7. Note that adding years contributes virtually nothing to the explained deviance.

A. Isolating the Role of Influential Observations

Another way to consider the role of the 11 anomalous observations is to drop them from the analysis. As an illustration, Figure 12 repeats the analysis shown in Figure 11, but deletes the 11 observations with more than 5 executions. That is, this figure uses 989 of the 1000 observations. The fit remains good; 90.4% of the deviance is accounted for. Visually, Figure 12 is basically a “blow-up” of the portion of Figure 11 located in the upper left corner, when the number of executions is 5 or less.

One can see that little of any importance is going on. There is a flat relationship in the average homicide rate when there are no executions compared to when there is 1. The relationship between 1 and 3 turns negative. The relationship between 3 and 5 turns positive. And any causal effects that might be inferred from these trends are tiny in any case.

B. Alternative Adjustments for Confounding

One might argue that controlling for state differences with a set of indicator variables is ham-fisted. After all, one of the ways that states may differ is in their inclination and ability to seek the death penalty. Such inclinations should not be removed from the treatment content of the number of executions.

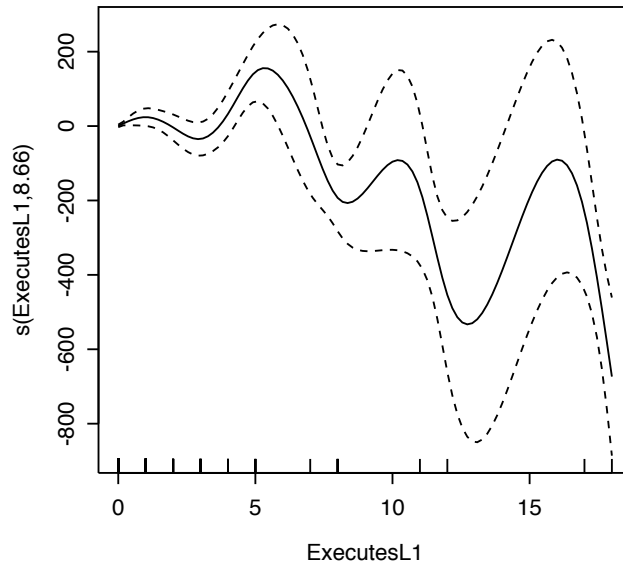


Figure 10: Number of Homicides as a Function of the Number of Executions Lagged by One Year, Controlling for State and Year (Explained Deviance = 97.4%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 8.66. The dotted lines contain the approximate 95% confidence interval. There is no consistent relationship between the number of homicides and the lagged number of executions when there are five execution or less. The apparent negative relationship when there are six executions or more is based on only eleven observations out of one thousand and cannot be taken at face value.

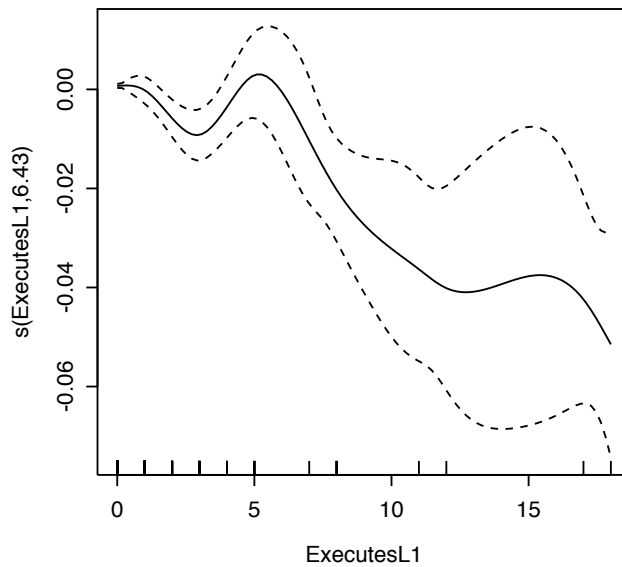


Figure 11: Homicide Rate as a Function of the Number of Executions Lagged by One Year, Controlling for State and Year (Explained Deviance = 90.2%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 6.43. The dotted lines contain the approximate 95% confidence interval. There is no consistent relationship between the homicide rate and the lagged number of executions when there are five execution or less. The apparent negative relationship when there are six executions or more is based on only eleven observations out of one thousand and cannot be taken at face value.

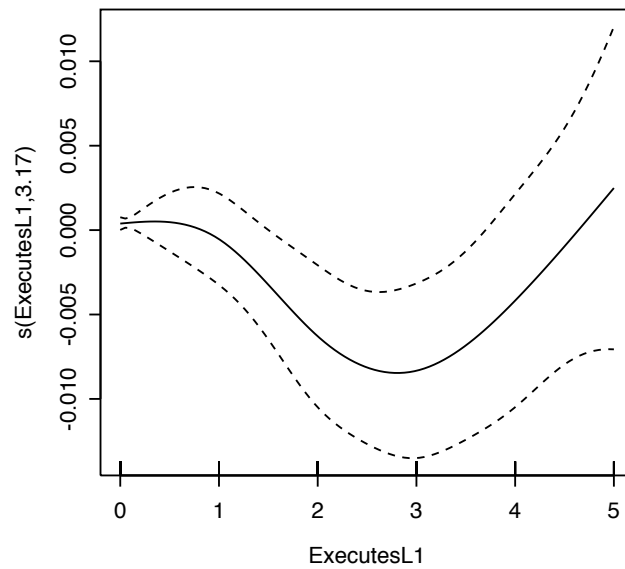


Figure 12: Homicide Rate as a Function of the Number of Executions Lagged by One Year Equal to 5 or Less, Controlling for State and Year (Explained Deviance = 90.4%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 3.17. The dotted lines contain the approximate 95% confidence interval. There is no consistent relationship between the homicide rate and the lagged number of executions when there are five execution or less.

In response, the model reported in Figure 12 was re-estimated. Instead of including the state indicator variables, the homicide rate in 1977 is used as a predictor, with its functional form estimated in a nonparametric manner. That is, a smoother is once again applied. Also, recall that because the number of executions is lagged by one year, the homicide rate as the response variable begins in 1978, one year after the homicide rate used as a control variable. This way, some of the same values are not on both sides of the equal sign.

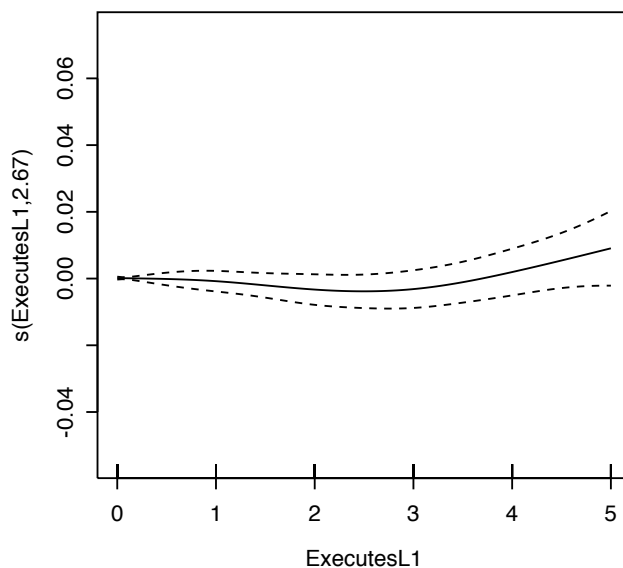


Figure 13: Homicide Rate as a Function of the Number of Executions Lagged by One Year Equal to 5 or Less, Controlling for the Homicide Rate in 1977 and Year (Explained Deviance = 82.8%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 2.67. The dotted lines contain the approximate 95% confidence interval. There is effectively no relationship between the homicide rate and the lagged number of executions when there are five execution or less.

If including all of the state indicators as predictors risks covariance adjustments that are too strong, using the homicide rate in 1977 risks covariance

adjustments that are not strong enough. It is not literally true that the factors affecting the homicide rate in each state were stationary from 1977 to 1997. For example, the “crack epidemic” was a phenomenon primarily of the 1980’s. But by weakening the covariance adjustments, more room is made for deterrence effects to surface.

Figure 13 shows the results. As one would expect, the explained deviance has dropped somewhat (to 83%), but the fit remains good. And there is still no evidence for any deterrence. The fitted values are essentially flat.

V. Some Variations on the Theme

What happens if instead of allowing the number of executions to take on a non-linear functional form, one proceeded in the same manner used to produce Figure 13, but with an assumed linear relationship? When the 11 extreme values are included, there is now a negative regression coefficient with a p-value of .09. This smacks of the deterrence findings popping up in the recent literature.

Similar conclusions are reached when the state-by-state controls used earlier are applied, and the number of homicides is the response. With the 11 extreme values included, a negative slope is found consistent with deterrence. Without the 11 extreme values, the slope is slightly positive, and there is no evidence for deterrence.

In short, by forcing a linear relationship and capitalizing on a very few extreme values, a negative relationship can be made to materialize. But the linear form is inconsistent with the data, and the 11 extreme values highly atypical.

What happens if instead of using the number of executions lagged by one year, one uses a binary indicator lagged by one year? The indicator is coded 0 for no executions and 1 for a single execution or more. This representation of the intervention eliminates the leverage of the long right tail of the execution distribution. The result is very small, negative regression coefficient, with a p-value of .51. There is no evidence for deterrence.

VI. Messin' with Texas

One can get a more grounded sense of the outliers in executions by considering what states produce them. In particular, it is well known that the state of Texas is highly unusual in the large number of prisoners executed. Figure 14 is a histogram for the number of executions in Texas during the years covered by the data. The differences between Figure 1, which is for all 50 States, and Figure 14 are dramatic. While for Texas the distribution is still somewhat skewed, the distribution is also highly atypical in the relatively large number of years in which a substantial number of prisoners were executed.

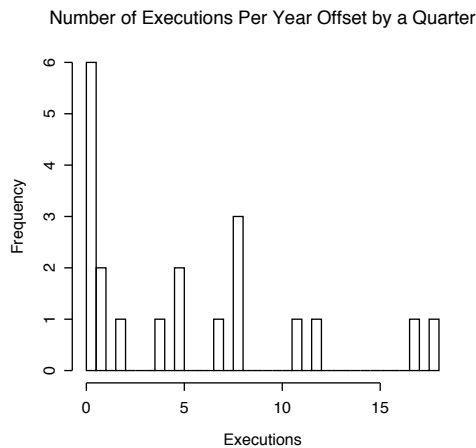


Figure 14: Empirical Distribution of the Number of Executions for Texas Alone — *Note*: There is still evidence for skewing, but the skewing is less extreme than for the data set as a whole.

Figures 15 and 16 show the relationship between the number of executions lagged by one year and the number of homicides and the homicide rate respectively, with the observations from Texas removed. For the number of homicides, there is nothing but a bit of noise around an overall trend that is slightly positive. For the homicide rate, there is less noise captured by the smoother, but the conclusion is essentially the same.

The “no findings” pattern persists when all of the analyses reported earlier are repeated with observations from the state of Texas removed. For example, Figures 17 and 18 show the results when covariates include the 1977 homicide

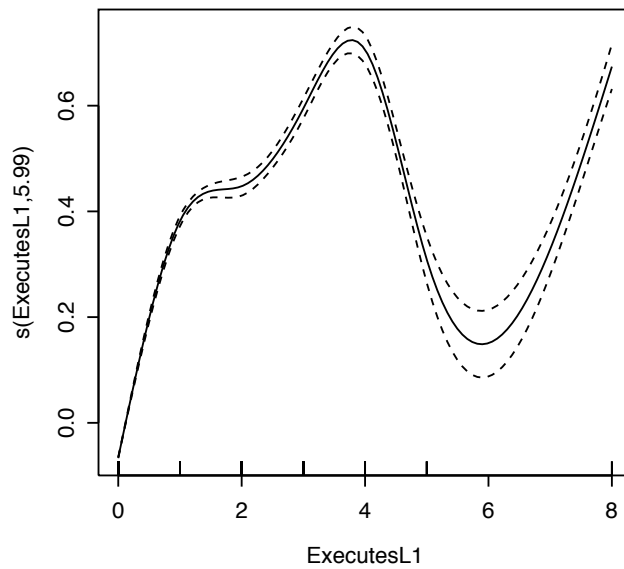


Figure 15: Number of Homicides as a Function of the Number of Executions Lagged by One Year with Texas Removed (Explained Deviance = 2.9%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 5.99. The dotted lines contain the approximate 95% confidence interval. There is no consistent relationship between the number of homicides and the lagged number of executions when Texas is removed from the data set.

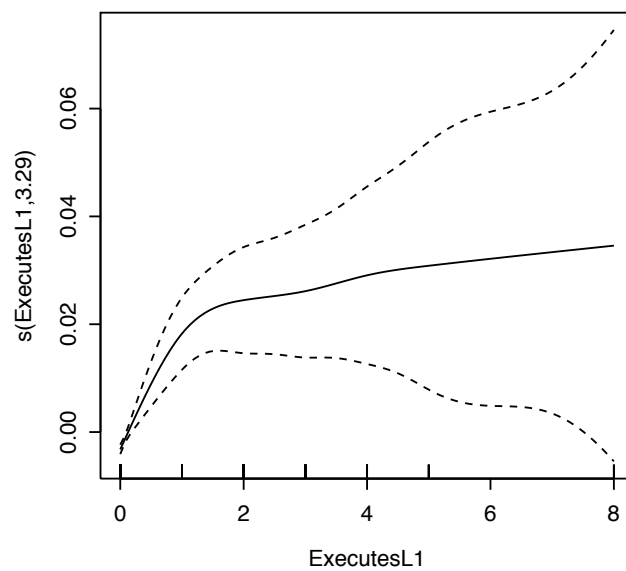


Figure 16: Homicide Rate as a Function of the Number of Executions Lagged by One Year With Texas Removed (Explained Deviance = 5.8% — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 3.29. The dotted lines contain the approximate 95% confidence interval. There is a positive relationship between the homicide rate and the lagged number of executions when Texas is removed from the data set.

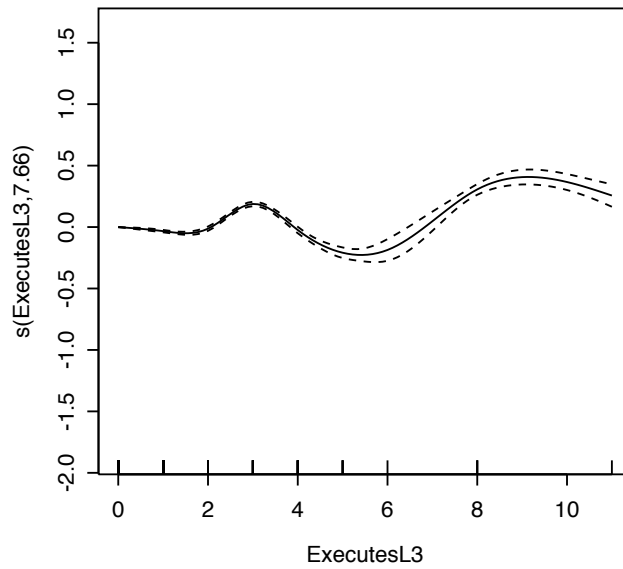


Figure 17: Number of Homicides as a Function of the Number of Executions Lagged by One Year, Controlling for the Homicide Rate in 1977 and Year with Texas Removed (Explained Deviance = 55.1%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 7.66. The dotted lines contain the approximate 95% confidence interval. There is effectively no relationship between the number of homicides and the lagged number of executions when Texas is removed from the data set.

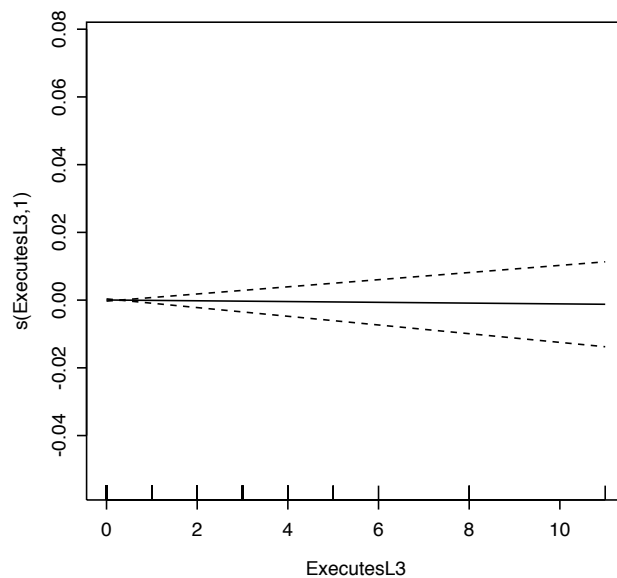


Figure 18: Homicide Rate as a Function of the Number of Executions Lagged by One Year, Controlling for the Homicide Rate in 1977 and Year with Texas Removed (Explained Deviance = 82.6%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 1.0. The dotted lines contain the approximate 95% confidence interval. There is no relationship between the homicide rate and the lagged number of executions when Texas is removed from the data set.

rate and year. These figures are selected because they are based on the more “gentle” of the covariance adjustments undertaken above. In both cases, it is hard to imagine a more compelling depiction of no effects.

In summary, one can either eliminate some outlier observations or what appears to be an outlier state. Any hint of a deterrent effect for executions then disappears. There is not much going on when all of the data are used, but clearly there is no evidence of a deterrent effect for 99% of the data.

A. A Little Simulation

Suppose it possible to alter the data so that for all states save Texas, there was no systematic relationship between the number of executions lagged by one year and the homicide rate. An empirical analysis of these altered data, with Texas excluded, should reveal no evidence of deterrence. But what would the analyses show if Texas were included? If there were then evidence of deterrence, one would have a powerful illustration of the role of influential observations.¹¹ Such a simulation is easy to do. For all states but Texas, one can just randomly shuffle the number of executions. The number of executions is then assigned just as in a randomized experiment and is then, on the average, unrelated to any other variables in the data set for 49 of the 50 states. It is also for these states unrelated on the average to any omitted variables.

Figure 19 shows the results when Texas is excluded from the analysis. By construction, there should be no relationship between the number of executions lagged by one year and the homicide rate. And in fact, no such relationship is apparent. The slope is flat and the associated p-value is .35.¹² Interestingly, Figure 19 looks remarkably like Figure 18 and many earlier figures in which outlier observations are excluded.

One cannot argue that the null finding derives from important confounded variables that have been neglected. By construction, the number of executions is on the average unrelated to any and all potential confounders. Including them in the analysis would on the average not have changed the story.

Figure 20 show the result when the state of Texas is included. One now sees a substantial downward slope roughly comparable to the downward slope

¹¹Thanks for this idea go to Rick Lempert.

¹²In GAM, the null hypothesis, loosely speaking, is that the slope does not improve the fit.

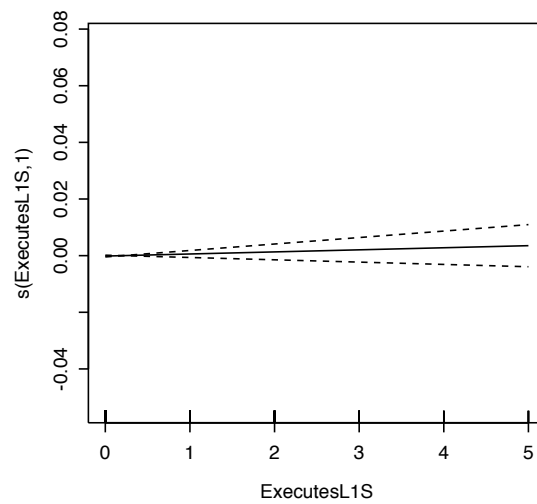


Figure 19: Homicide Rate as a Function of the Shuffled Number of Executions Lagged by One Year, Controlling for the Homicide Rate in 1977 and Year with Texas removed (Explained Deviance = 82.6%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 1.0. The dotted lines contain the approximate 95% confidence interval. There is no relationship between the homicide rate and the shuffled lagged number of executions when the data from Texas is removed.

found earlier in the real data. (See Figure 11.) The associated p-value is well below the .0001, implying that the negative relationship is “statistically significant” as well. Yet, these “findings” are clearly an artifact of the Texas observations.

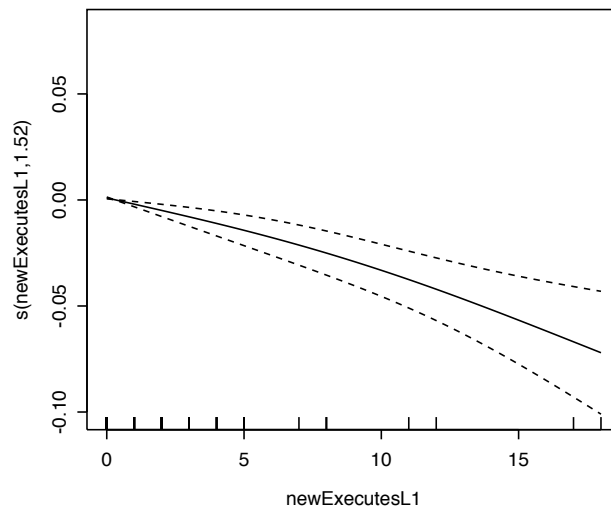


Figure 20: Homicide Rate as a Function of the Shuffled Number of Executions Lagged by One Year, Controlling for the Homicide Rate in 1977 and Year with Texas Included (Explained Deviance = 63.7%) — Note: The solid line is the smoothed fitted values, with the effective degrees of freedom equal to 1.52. The dotted lines contain the approximate 95% confidence interval. There an artifactual negative relationship between the homicide rate and the lagged number of executions when the data from Texas are put back into the analysis.

Three points need to be emphasized. First, the negative slope surfaces even though we *know* that in the altered data the homicide rate and the number of executions is unrelated for 49 of the 50 states. Second, by adding the true observations for Texas, one can produce the appearance of deterrence. Clearly this is an artifact of several very influential observations. Third, one must *not* take this as evidence for deterrence in Texas. There are not enough data to judge for any single state, even Texas.

VII. A few Comments about Transformations, Standardizations and lags

There is no strong theoretical justification for using a one year lag for executions, but even less theoretical justification for other lags. If someone has done serious research on how executions might be viewed by potential murderers, or as a cognitive matter, what the most appropriate time lags should be, I cannot find it in any reputable journals.

There also seems to be no empirical justification for a wide variety of standardizations, such as dividing the number of executions in a given year by the number of death sentences that year and/or in some earlier years. I can find no empirical evidence that potential murderers have any sense of the number of death sentences that, in turn, need to be taken into account when the import of executions is considered. A researcher might well want to estimate the probability of an execution conditional upon the a death sentence, but there is absolutely no serious research I know of showing that potential murderers think this way (even if they could get the data).¹³

It seems, therefore, that complicating matters by significantly adjusting or altering how the number of executions enters the analysis is not a good strategy. It just makes the analysis more obscure. Moreover, it further increases the risks of overfitting, which are already substantial. But I leave that for another time.

Finally, the reanalyses reported above should be relevant for any research that relies on the number of executions as the “treatment dose.” The number of executions is available from Bureau of Justice Statistics. Transformations, lags, and standardizations cannot alter the reality that these data are the building blocks for all that follows. If they are problematic, so are the variables derived from them.

¹³Researchers in this tradition cite Gary S. Becker (Becker, 1968) as the wellspring for their work. Without even getting into the empirical research that creates serious problems for rational choice theory (e.g., Kahneman and Tversky, 1974; 1996) or whether the deterrence formulations make unreasonable demands on the computational skills of prospective murderers (Parisi and Smith, 2004), all would surely agree that if the goal is to measure how executions and death sentences are taken into account before a murder is committed, the official statistics do a questionable job. In a multivariate analysis, therefore, inconsistent estimates result even if the measurement error is unsystematic. And calling flawed measures “proxies” does not solve the problem; it is actually an admission of faulty measurement (Greene, 2000: section 5.6).

Thus, a recent paper by Dezhbakhsh and his colleagues (2003) uses county-level data, where available. But most counties during most years do not have a single capital case that results in an execution. The highly skewed distribution for the number of executions remains.¹⁴ Shepherd (2004) disaggregates still further with months now the temporal unit. However, it should be no surprise that most counties in most months do not have a single capital case that results in a execution. Indeed most counties *never* have such a case. The distributional problems with the number of executions do not go away.¹⁵ Zimmerman (2004) overlays instrumental variable estimation, which addresses a totally different matter. The impact of the few influential observations remains.¹⁶

VIII. Conclusions

The conclusions that follow from the analyses are straightforward. From 1977 to 1997, most states during most years executed no one. A few states on rare occasion executed up to 5 individuals in a particular year. Years with 5 executions or less represent 99% of the data. Limiting the analysis to these 99% of the observations, there is no evidence of a negative relationship between the number of executions lagged by one year and either the homicide rate or the number of homicides. Including the 11 execution extreme values can for some analyses suggest possible negative effect, but *only* for these 11 observations and *only* if one ignores the very wide confidence intervals.

Imposing a linear relationship between executions and either the number of homicides or the homicide rate, and including the 11 extreme values, can for some analyses generate the appearance of deterrence effects much like those recently reported. But the linear relationship is contradicted by the data, and any hint of deterrence disappear if the 11 observations are excluded from the analysis. There is also no evidence in these data that replacing the number of executions with an indicator variable coded “0” for no executions and “1” for 1 or more executions changes the overall conclusions.

¹⁴And moving to the county level exacts a very high, and apparently unappreciated, price. One might be able to ignore spatial dependence between states, but certainly not between counties.

¹⁵And the analysis now requires that potential homicide perpetrators revise their calculations about every 30 days. Federal, state and county agencies do not do as well. The demographic proxy variables relied upon in the analysis are at best updated once a year.

¹⁶And new collection of fanciful assumptions is introduced.

All of the points made about the 11 extreme values apply to the data from Texas. Moreover, a simple simulation demonstrates that even when one knows for certain that in the 49 other states executions are on the average unrelated to the homicide rate, including the data from Texas can give the false impression that a deterrence relationship exists.

The distributional problems that characterize the number of executions remain when counties are the spatial units and/or months are the temporal units. Disaggregating the data does not make the skewing disappear, and can actually introduce a new set of difficulties. More elaborate estimation procedures only paper over the problems and introduce a new layer of dubious assumptions.

Whatever one makes of those 11 observations, it would be bad statistics and bad social policy to generalize from the 11 observations to the remaining 989. So, for the vast majority of states for the vast majority of years, there is no evidence for deterrence in these analyses. And even for the remaining 11 observations, the credible evidence for deterrence is lacking.

The analyses reported here are hardly exhaustive and are perhaps affected by misunderstandings about the data provided, or by errors in the data themselves. Nevertheless, the results raise serious questions about whether anything useful about the deterrent value of the death penalty can ever be learned from an observational study with the data that are likely to be available. With an intervention that is so highly skewed, a very small portion of the data will likely impart significant influence on the results. Generalizations to the mass of the data then become very risky. It is difficult to imagine how such problems can be overcome no matter how skilled or sophisticated the data analyst.

Finally, the focus in this paper has been in the role of influential observations because their impact is well understood in the statistical literature and because in this instance, the errors that can result are easily demonstrated. But as noted in the introduction, there are many other statistical problems could have been tackled.

1 References

- Becker, G.S. (1968) "Crime and Punishment: An Economic Approach," *Journal of Political Economy* 76: 169-217.
- Berk, R.A. (2003) *Regression Analysis: A Constructive Critique*. Sage Publications, Newbury Park, CA.
- Berk, R.A. (2005) "Data Mining within a Regression Framework," in *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*, Oded Maimon and Lior Rokach (eds.), Kluwer Academic Publishers, Forthcoming, 2005.
- Blume, J., Eisenberg, T., and M.T. Wells (2004) "Explaining Death Row's Population and Racial Composition," *Journal of Empirical Legal Studies* 1(1): 165-207.
- Box, G.E.P. (1976) "Science and Statistics." *Journal of the American Statistical Association* 71: 791-799.
- Breiman, L. (2001) "Statistical Modeling: Two Cultures," (with discussion) *Statistical Science* 16: 199-231.
- Cloninger, D.O., and R. Marchesini. (2001) "Executions and Deterrence: A Quasi-Controlled Group Experiment." *Applied Economics* 35, 5: 569-576
- Cook, D.R. and Sanford Weisberg (1999) *Applied Regression Including Computing and Graphics*. New York: John Wiley and Sons.
- Dezhbakhah, H., Rubin, P.H. and J.M. Shepherd (2003) "Does Capital Punishment Have a Deterrent Effect? New Evidence from Postmoratorium Panel Data," *American Law and Economics Review* 5(2): 344-376.
- Gelman, A., Liebman, J.S., West, V., and A. Kiss. (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.
- Ehrlich, I. (1975) "The Deterrent Effect of Capital Punishment: A Question of Life and Death." *American Economic Review* 65, 3: 397-417.

- Ehrlich, I. (1977) "Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence." *Journal of Political Economy* 85, August: 741-788.
- Ehrlich, I., and Z. Liu (1999) "Sensitivity Analysis of the Deterrence Hypothesis: Let's Keep the Econ in Econometrics." *Journal of Law and Economics*, 41,1: 455-488.
- Forst, B. (2004) *Errors of Justice*. Cambridge: Cambridge University Press.
- Freedman, D.A. (1987) "As others See Us: A Case Study in Path Analysis" (with discussion). *Journal of Educational Statistics* 12: 101-223.
- Gibbs, J.P. 1975) *Crime, Punishment, and Deterrence*. New York: Elsevier.
- Greene, W.H. (2000) *Econometric Analysis*, fourth edition. New York: Prentice Hall.
- Gross, S.R., and R. Mauro. (1989) *Death and Discrimination*. Boston: Northeastern University Press.
- Hastie, T.J. and R.J. Tibshirani. (1990) *Generalized Additive Models*. New York: Chapman & Hall.
- Hastie, T., Tibshirani, R., and J. Friedman, (2001) *The Elements of Statistical Learning*. Springer-Verlag.
- Heckman, J. (1999) "Causal Parameters and Policy Analysis in Economics: A Twentieth Century Retrospective," *The Quarterly Journal of Economics*, February: 45-97.
- Kahneman D., and A. Tversky (1974). "Judgment Under Uncertainty: Heuristics and Bias," *Science* 185: 112431.
- Kahneman D., and A. Tversky (1996) "On The Reality of Cognitive Illusions," *Psychological Review* 103: 58291.
- Klein, L.R., Forst, B., V. Filatov(1978) "The Deterrent effects of Capital Punishment An Assessment of Estimates," in A. Blumstein, D. Nagin, and J.Cohen (eds.), *Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates*. Washintgon, D.C., National Academy of Sciences.

- Leamer, E.E. (1978) *Specification Searches: Ad Hoc Inference with Non-Experimental Data*. New York, John Wiley.
- Manski, C.F. (1990) "Nonparametric Bounds on Treatment Effects." *American Economic Review Papers and Proceedings* 80: 319-323.
- Mocan, H.N. and K. Gittings (2003) "Getting off Death Row: Commuted Sentences and the Deterrent Effect of Capital Punishment." (Revised version of NBER Working Paper No. 8639) and forthcoming in the *Journal of Law and Economics*.
- Parisi, F., and V. L. Smith (eds.) (2004) *The Law and Economics of Irrational Behavior*. Palo Alto, CA: Stanford University Press, forthcoming.
- Rosenbaum, P.R. (2002) *Observational Studies*, second edition. Springer-Verlag.
- Rubin, D. B. (1986) "Which Ifs Have Causal Answers." *Journal of the American Statistical Association* 81: 961-962.
- Sellin, T. (1980) *The Penalty of Death*. Newbury Park, CA: Sage Publications.
- Shepherd, J.M. (2004) "Murder of Passion, Execution Delays, and the Deterrence of Capital Punishment," *Journal of Legal Studies*, 33: 283-321
- Sutherland, E. H. (1925) "Murder and the Death Penalty." *Journal of Criminal Law and Criminology* 15: 522-536.
- Thompson, S.K. (2002) *Sampling*, second edition. New York, John Wiley.
- Zimmerman, P.R. (2004) "State Executions, Deterrence, and the Incidence of Murder," *Journal of Applied Economics*, 7: 163-193.
- Zimring, F.E. (2003) *The Contradictions of American Capital Punishment*. Oxford: Oxford Press.
- Zimring, F.E., and G.J. Hawkins (1973) *Deterrence* Chicago, University of Chicago Press.