MF-1

# THE DETERRENT EFFECT OF CAPITAL PUNISHMENT
# I. A TEST OF SOME RECENT STUDIES*



ABSTRACT

We develop a reliability statistic $\bar{K}$ to help estimate the level of systematic error in a mathematical model of homicide patterns. We then calculate $\bar{K}$-values for three recent studies on the deterrent effect of capital punishment; the results lead us to suggest that none of the models used is sensitive enough really to help resolve the question.

ARNOLD BARNETT
Sloan School of Management
Operations Research Center
Massachusetts Institute of Technology
Cambridge, Massachusetts

84552

*This paper was supported under Grant No. 78-NI-AX-0034, from the U.S. Department of Justice, Law Enforcement Assistance Administration, National Institute of Law Enforcement and Criminal Justice. Points of view or opinions stated in this document do not necessarily represent the official position or policies of the U.S. Department of Justice.

Acknowledgements

CHRIS McMAHON's extensive work in data gathering and computer programming was enormously helpful in this project. The perceptive comments of RICHARD LARSON, MICHAEL MALTZ, and EDWIN ZEDLEWSKI substantially improved the organization of the paper. The referees were unusually conscientious, insightful, and patient with a rather unwieldy first version.



## Introduction

In the United States, probably no crime-related issue has been as fiercely debated in recent years as the propriety of capital punishment. Since 1972, the Supreme Court has ruled several times on the issue, and pertinent statutes have been changed in over 35 states. While, outside of Utah, no U.S. executions have occurred this decade, it would be premature to conclude that opponents of the death penalty have won the day. Judicial appeals continue in some states; in others, capital punishment is still a major political issue. In 1978 alone, Californians voted 72-28 to extend the list of capital crimes, Oregon voters endorsed the death penalty, and, in Massachusetts and New York, opposition to capital punishment was viewed as a major political liability for certain gubernatorial candidates.

A major issue in the death-penalty debate is whether executions deter enough murders that their net effect is to save lives. A large number of statistical studies have appeared on this question; so many, in fact, that Professor Ernest Van den Haag spoke recently of a "cottage industry." While major exceptions exist, visitors to most of these cottages are informed that the death penalty is ineffective as a deterrent to homicide.

Virtually all recent researchers have recognized that, to use homicide data to assess the effect of executions, one must first try to "weed out" the effects of extraneous factors. They therefore attempted, usually through multivariate regression, to create a comprehensive theory of the determination of homicide levels. An important question about any such theory is whether its predictions are consistent enough with actual patterns of killing that its implications on capital punishment warrant serious attention. The issue is particularly crucial because, in the United States, executions have always been rare relative to homicides; thus even if each execution had considerable deterrent power, the overall impact of capital punishment would be small and hard to detect.

The "reliability" debate over capital punishment research has yielded an unusually wide range of positions. Some criminologists believe that the difficulty of identifying, let alone correcting for, the myriad influences on homicide rates precludes the assessment of the impact of executions. Others, while skeptical of the present generation of studies, nonetheless consider their flaws correctible and believe successful studies possible. Still others, though highly critical of some recent studies, are sufficiently impressed with others that they fully accept their conclusions. This division of opinion was reflected in the recent report of the National Research Council's Panel on Deterrence Research [1]. The Panel as a whole concluded that death penalty studies had a questionable past and a bleak future; those it commissioned to review such studies in detail, however, were far more optimistic and, indeed, one of them (Forst) shortly thereafter undertook a statistical study of his own.

How can intelligent scholars maintain such disparate views on what is ultimately a question of fact? The central problem is that statistical studies on capital punishment lack meaningful measures of their accuracy. Most statistical tests on regression models, for example, concern not whether the models are accurate depictions of reality but the far weaker question of whether they are not self-contradictory. In the absence of clear "standards of accountability," those examining a given death penalty study have been unsure whether its inevitable imperfections were crippling deficiencies or second-order effects. In such circumstances, accepting the conclusions of such a study has been more an act of faith than a logical necessity.

This situation differs sharply from that in such other areas of statistical endeavor as opinion polling. There a precise theory identifies the "margin of error" in a pollster's results given that his sample was genuinely random. This theory provides accuracy standards by which, for example, the pollster's pre-election predictions can be compared to the actual voting results and, over a period of



time, as assessment of his reliability made. It is the author's belief that despite current practices, one can develop similar standards of accountability for mathematical models of homicide patterns. To attempt to do so and then to apply these standards to some recent death penalty studies are the objectives of this paper.

The keystone of out efforts is an attempt to estimate, through probabilistic reasoning, the size of the natural random fluctuations in homicide levels (i.e., those variations over time or space unrelated to systematic factors). Such fluctuations, like sampling error to the pollster, create an unavoidable level of inaccuracy in the individual predictions of any given model. In the long run, fluctuations are unlikely in themselves greatly to affect calculated results; the main reason to estimate their size is to allow the estimation of the systematic errors associated with the failings of particular models. If such systematic errors are far larger than any plausible estimate of the cumulative effect of past executions, then the model at hand is probably incapable of discerning the deterrent effect of capital punishment. These ideas will be elaborated in forthcoming sections of the paper.

The standards developed are applied to three recent studies on whether executions deter murder -- those of Passell [5], Ehrlich [2], and Forst [3]. These three studies have received widespread public attention and are among the most sophisticated of their breed. We find, however, that the underlying mathematical models of all three papers are subject to substantial systematic error; this finding raises serious doubt about their capacities to perform their intended function. The implications of the results are discussed at the end of the paper.

The paper is organized as follows: in Section I, we estimate the magnitude of the "random component" of recorded annual homicide levels. This estimate is utilized in Section II, where we develop a reliability statistic $\bar{K}$ to help measure the level of systematic error in a particular model of homicide patterns. In

Section III, we calculate $\bar{K}$ values for the Passell, Forst and Ehrlich models; our final remarks appear in Section IV.

I. Random Fluctuations in Homicide Levels

By a mathematical model of homicide levels we mean a formula for estimating a locality's homicide level on the basis of its prevailing demographic, social, and economic conditions, and its patterns of punishment for homicide. Researchers who have obtained such formulas have typically specified a functional form in advance, and then estimated particular parameters in the model so as to achieve greatest consistency with a given set of data. Regression theory, which has often been invoked in such efforts, provides confidence intervals as well as point estimates for the individual parameters; these intervals, however, are calculated from a series of assumptions, one of which is that the functional form assumed for the relationship is exactly correct. If this assumption is false, then accuracy estimates for the parameters -- like the parameters themselves -- may have no serious meaning.

No mathematical model of homicide levels, however obtained, can be expected to have perfect predictive power for, quite apart from systematic factors, sheer chance exerts an influence on recorded annual homicide levels. Even if a person has decided to commit murder, the time she attempts to do so may be related to random elements; whether she kills rather than just maims her victim is itself often subject to chance. By contrast, certain assults not at all intended to be lethal nonetheless wind up being so. Furthermore, it seems likely that some homicides are mistakenly identified as suicides, accidents, or natural deaths, while some deaths that appear to have been inflicted deliberately actually were not. But the existence of a random component in homicide levels should not become an "elastic clause" that can be used to explain away the prediction errors of any model. As we attempt to show below, it is possible reasonably to estimate the



the overall magnitude of such random effects.

Suppose that at the beginning of the calendar year, there is associated with each of the N individuals who live in or visit a given region, a probability distribution for the number of reported killings (s)he will commit there that year. More precisely, let $p_{ij}$ be the probability that individual will commit j reported killings that year in that region. (Clearly no mortal actually knows all the $p_{ij}$'s, though one can imagine ways to make estimates.) $\mu_i$, the mean number of recorded murders by person i, follows $\mu_i = \sum_{j=0}^{\infty} j\, p_{ij}$ ; the associated variance $\sigma_i^2$ follows $\sigma_i^2 = \sum_{j=0}^{\infty} j^2 p_{ij} - \mu_i^2$. A typical value of $\mu_i$ in the United States is about $10^{-4}$. We hypothesize that only very rarely does $\mu_i$ exceed $10^{-2}$, for there is no identifiable subgroup of the population -- even inner-city teenage gang members -- in which more than one percent commit murder per year. Under these circumstances, it seems reasonable to assume that $\mu_i^2$ is negligibly small compared to $\mu_i$; since $\sum_{j=1}^{\infty} j^2 p_{ij} > \mu_i$, we can further assume that the $\mu_i^2$ term can be ignored in the expression for $\sigma_i^2$.

The region's total number of recorded murders in the forthcoming year, denoted by M, is treated as a random variable with expected value $\mu_M$ that follows:*

$$\mu_M = \sum_{i=1}^{N} \mu_i = q_1 + 2q_2 + 3q_3 + \ldots kq_k + \ldots \tag{1}$$

where

$$q_j = \sum_{i=1}^{N} p_{ij} .$$

The assumption that different would-be killers act independently is not literally true. But we argue in Appendix A that the deviations that arise from independence can reasonably be viewed as second-order effects. We therefore approximate $\sigma_M^2$, the variance of M, by the expression

$$\sigma_M^2 = \sum_{i=1}^{N} \sigma_i^2 = q_1 + 4q_2 + 9q_3 + \ldots k^2 q_k + \ldots \tag{2}$$

where $\mu_i^2$ terms are ignored in (2).

*When more than one person is implicated in a particular killing, we "credit" the homicide to the "ringleader."

Comparing (2) to (1) makes clear that unless multiple murders are neglected, $\sigma_M^2$ will exceed $\mu_M$; an important question is how much. Here some recent research becomes useful. Goehrke et. al. [4] studied homicide patterns from the mid 1950's to the mid 1960;s in several American states and cities that, while all in the East, were otherwise radically different (e.g., New York and Vermont). They found strong consistency across the localities in the fraction of killings that were double murders, had three or more victims, or were part of a series of murders by one person (e.g., the Boston Strangler). Multiple murders were actually quite rare, and accounted for only about 7/10 of one percent of all incidents of homicide. For every 1000 people who were the only known victims of their killers that year, it was estimated that 9.1 were slaim by people who also killed one other person, 4.8 were victims of triple murders, 2 of quadruple killers, 3 of killers with between 5 and 7 victims. If such patterns prevail in a given region, we would expect that $2q_2/q_1 \approx 9.1/1000$ or $q_2 \approx .0046q_1$; similarly, we would expect $q_3 \approx .0016q_1$; $q_4 \approx .0005q_1$; etc. Substituting these approximations into (1) and (2) yields $\mu_M \approx 1.019q_1$ and $\sigma_M^2 \approx 1.059q_1$, which in turn provides the estimate:

$$\sigma_M^2 \approx 1.04\mu_M \,. \tag{3}$$

This factor 1.04 should hardly be viewed as a universal constant, but it is a useful first approximation of the effect of mass killers.

Given the assumption that different potential killers act independently, the Central Limit Theorem implies that, except when $\mu_M$ is small (below 20 or so), M is roughly normally distributed.

We might summarize the discussion above as follows: <u>associated with a given locality and given time period, there is a parameter $\lambda$ such that the number of recorded killings in the period can be treated as one sample from the normal distribution with mean $\lambda$ and variance $1.04\lambda$.</u> $\lambda$ is in some sense the "true" homicide level for the period, devoid of random fluctuations. A mathematical model of



homicide levels can be viewed as an attempt to estimate $\lambda$ from the values of other variables that supposedly determine it. Such a model might be considered perfect if its estimates of $\lambda$ were always correct; even such a "perfect" model, we have argued, would be subject to normally distributed prediction errors, with means of zero and variances roughly equal in size to the predictions themselves.

Given the immense number of factors, many of which elude quantification, that might influence homicide rates, it would be naive to believe that any tractable model could predict $\lambda$ correctly over a wide range of circumstances. Thus the demonstration through statistical testing that the errors of a particular model cannot be explained by chance alone is neither surprising nor especially useful. A more constructive approach might involve trying to assess whether the systematic errors of a model are so large that they cast doubt on its ability to perform its stated purpose. Such an approach is discussed in the next section.

## II. <u>The Systematic Error of a Homicide Model</u>

We have suggested that a prediction of a homicide model should be considered an estimate of the mean of a certain normal distribution, while the actual number of killings is one sample pick from that distribution. Thus a general result about normal distributions that we will obtain now will soon be useful. Let $\mu$ and $\sigma^2$ be the mean and variance, respectively, of a normal distribution, let $w = \mu + k\sigma$ be an erroneous estimate of $\mu$ (we assume $k > 0$ for convenience), and let x be a sample pick from the distributed. We wish to compare the mean values of $|x - w|$ and $|x - \mu|$; this will indicate the effect of the estimation error for $\mu$ on the observed discrepancy between the estimate and the sample value (i.e., the "residual").

We can write $|x - w| = |x - \mu| + D(x)$ where

$$D(x) = \begin{cases} k\sigma \text{ if } x \le \mu \\ -k\sigma \text{ if } x \ge w \\ k\sigma - 2(x - \mu) \text{ if } \mu < x < w \,. \end{cases}$$

Using the facts that $Pr(x < \mu) = .5$ and $\int_0^{w} x e^{-\frac{x^2}{2\sigma^2}} dx = \sigma^2\left[1 - e^{-\frac{k^2}{2}}\right]$, we easily deduce that:

$$E(D(x)) = 2k\sigma(.5 - R(k)) - \sigma\sqrt{2/\pi}\left[1 - e^{-\frac{k^2}{2}}\right] \qquad (4)$$

where $R(k) = Pr(x > w = \mu + k\sigma)$.

Since $E(|x - \mu|) = \sigma\sqrt{2/\pi}$, we obtain from (4) that:

$$E(|x - w|) = \sigma(2k(.5 - R(k) + \sqrt{2/\pi}\, e^{-\frac{k^2}{2}}) . \qquad (5)$$

One can establish through a simple symmetry argument that, for any $k > 0$, $E(|x - (\mu + k\sigma)|) = E(|x - (\mu - k\sigma)|)$.

When $k = 1$ and thus $\mu$ is overestimated by $\sigma$, (4) implies that $E(D(x)) = .367\sigma$; when $k = .5$, $E(D(x)) = .131\sigma$. These calculations remind us that, on the average, only a fraction of the error in estimating $\mu$ shows up as an increase in the residual. This happens because the fluctuation of $x$ around $\mu$ sometimes brings it closer to the incorrect estimate than to $\mu$ itself. Because of this, even a small excess in observed residuals over the random-noise level $\sigma\sqrt{2/\pi}$ can indicate considerable inaccuracy in the estimates of means.

Use of a homicide model entails a prediction problem of the kind just considered in which, in earlier terminology, $\mu = \lambda$, $\sigma^2 = 1.04\lambda$ and the $x$ is the observed number of killings. If $w$ is the model's predicted number of murders in a given situation, we know directly the total prediction error $|w - x|$. But in assessing the model's accuracy we are really interested in $|w - \lambda|$, the systematic (nonrandom) component of the error. And this quantity must be estimated while $\lambda$ itself is unknown.

We proceed in the discussion below on the assumption that predictions are made on data other than that on which the model was calibrated (i.e., from which the parameters in the model were estimated). If this circumstance does not obtain, modifications described later must be made.

Suppose that a homicide model is used to make $N$ independent predictions of the murder levels in given places and periods (e.g., Illinois in 1950). Let $q_i$



be the $i^{th}$ of these predictions, let $\lambda_i$ be the corresponding true mean of the *a priori* distribution of the recorded number of killings, and let $x_i$ be the number ultimately observed. We can write $q_i = \lambda_i + k_i\sigma_i$, where $\sigma_i = 1.02\sqrt{\lambda_i}$; $k_i$ represents the systematic error in the $i^{th}$ prediction, measured in standard deviations. While obtaining a reliable estimate of $k_i$ based solely on $x_i$ is not possible, one can make a reasonable estimate of the "typical" $|k_i|$ in a large number of predictions. Let $r_i = (q_i - x_i)/\sigma_i$, and let $S = \sum_1^N r_i / N$. (i.e., $S$ is the average normalized residual). Consider the nonnegative quantity $\bar{K}$ defined by the equation:

$$\bar{K} = 0 \quad \text{if } S \le \sqrt{2/\pi} \qquad (6A)$$

$$2\bar{K}(.5 - R(\bar{K}) + \sqrt{2/\pi}\, e^{-\frac{\bar{K}^2}{2}}) = S \qquad \text{if } S > \sqrt{2/\pi} \qquad (6B)$$

where

$$R(\bar{K}) = \frac{1}{\sqrt{2\pi}} \int_{\bar{K}}^{\infty} e^{-\frac{x^2}{2}} dx .$$

If all the $k_i$'s are equal and $N$ is large, it follows from comparing (6B) to (5) that $\bar{K}$ is a highly accurate estimate of the absolute value of each $k_i$. When the $k_i$'s vary, $\bar{K}$ tends slightly to exceed their average absolute value. If, for example, $N = 50$, half the $k_i$'s are 1 and half 2 (i.e., the average $k_i$ is 1.5), then given the normal distributions of the $x_i$'s, the data-based estimate of $\bar{K}$ from (6B) will have mean value 1.54 and standard deviaition .14. In either case, $\bar{K}$ is a useful indicator of how much systematic error enters the predictions of a homicide model (i.e., error based on imperfections of the model rather than chance fluctuations). While $\bar{K}$ is expressed in standard deviations, one can multiply it by typical values of $\sigma$ to approximate the actual size of the model's systematic error, a highly relevant quantity if one is considering the utility of the model for particular purposes. We will soon calculate $\bar{K}$'s for three different homicide models to help assess their power to discern the deterrent effect of capital punishment.

$\bar{K}$, like the traditional $R^2$, is measure of the explanatory power of a model; the two indicators arise, however, from very different considerations. $R^2$ reflects the improvement associated with using the current model instead of a very crude one (specifically, one that assumes the dependent variable a constant, unaffected by any other factors). $\bar{K}$, by contrast, compares the model's accuracy with that one would expect of a model with no systematic error, and thus estimates the impact of its deficiencies on its predictions. More importantly, $R^2$ is primarily a relative measure of accuracy while $\bar{K}$ is an absolute measure. It is an absolute measure, we believe, that is central to assessing a model's usefulness.

We should stress that $\bar{K}$ is a measure of the observed systematic error of a model, which is not a perfect reflection of the accuracy of its assumptions. Consider, for example, a model with a low $\bar{K}$, one of whose variables, $y_1$, is highly correlated with a variable $y_2$ that it excludes. The model might erroneously attribute to $y_i$ the influence actually exerted by $y_2$ yet, because of the high correlation, the model's predictive power would be only slightly diminished. It seems fair to say that a high $\bar{K}$ is a surer indicator that a model is deficient than a low $\bar{K}$ is that it is successful.

## III. Evaluation of Three Death Penalty Studies

In this section we consider three recent studies on the deterrent effect of capital punishment. We describe their underlying mathematical models, calculate $r_i$'s and $\bar{K}$'s for their predictions, and attempt to assess the degree of confidence one might reasonably have in their conclusions. We should stress that we focus exclusively on how well the models work in practice, and make no comments about the reasoning that led to their construction. The first paper discussed is that of Passell [5].

### (i) Passell

Since death penalty statutes arise at the state level, Passell's study, like the others considered here, uses the state as the unit of observation. Passell performed a cross-sectional regressional analysis on data from 41 American states



for the year 1950, and then proceeded similarly for 44 states in 1960. He hypothesized that the differences in homicide races among the states were essentially reflections of cross-state differences in:

- P = the fraction of homicides that ultimately led to criminal convictions
- T = the median time spent in prison by convicted killers
- E = the fraction of those defendants convicted of homicide who were executed
- A = the percentage of the resident population between the ages of 15 and 24
- I = the percentage of families with incomes below the poverty level
- M = the ratio of net non-white migrants in the past ten years to the state's total population
- S = a regional indicator variable (1 for Southern states, 0 for others, meant to reflect the greater "tradition of violence" in the South than elsewhere)

Passell assumed that in a given year, the homicide rate H in any state could be estimated from an equation of the form:

$$H^B = C_0 + \sum_{i=1}^{N} C_i Z_i^B \qquad (7)$$

where $Z_i$ is the prevailing value of the $i^{th}$ explanatory variable and the $C_i$'s and B are constants that don't vary across states. He separately set B = 1 (in which case (7) is a linear equation) and B = .1, and in each case determined the ordinary least squares estimates of the $C_i$'s (i.e., those values under which the mean square deviation between predicted state homicide rates under (7) and the corresponding actual rates was minimized). Because he considered two separate years, two values of B, and alternative definitions of certain variables, Passell generated several regression equations of the form of (7). Under the $R^2$ measure he chose to describe goodness-of-fit, one of his most successful equations was linear and for 1960:

$$H = -15.1 - 4.68P - .013T + 4.87E + 1.45A + .122I + 1.95M + 2.04S$$
$$(R^2 = .860) \qquad (8)$$

(NOTE: P, E, A, I, and M are measured in units of .01 .)

Since the coefficient of E is positive, (8) suggests that increases in the execution rate E might actually stimulate homicides. While Passell argued against this interpretation (and, indeed, the coefficient of E in (8) is not statistically significant), he summarized his various results with the statement that he saw "no reasonable way of interpreting the cross-section data that would lend support to the deterrence hypothesis."

Passell's analysis rested on strong assumptions about homogeneity across states, about which factors are related to homicide levels, and about the functional form of the complete relationship. We attempt below using the concepts discussed earlier to estimate the systematic error in his models caused by inaccuracies in his assumptions. We focus on the predictive power of equation (8) above; results for his other equations are generally worse.

Once one has collected the data needed to use (8), one can make predictions for the homicide rates in 1960 in individual states; these rates, when multiplied by resident populations, yield corresponding estimates of total numbers of killings. (Checks on the accuracy of our database are described in Appendix B.) As described earlier, the predicted homicide level $q_i$ for state i can be compared to the actual level $x_i$ through the normalized residual $r_i = (q_i - x_i)/1.02\sqrt{x_i}$.* However, a complication exists; the eight parameters of equation (8) were chosen to achieve maximum agreement with the very data now being used to assess (8)'s predictive power. The regression theory invoked by Passell specifies that when 8 parameters are chosen for consistency with 43 data points (i.e., different states), a linear model's residuals are artificially reduced by an average of 9.5%. Thus, to correct for this bias, we estimate a "true" normalized residual $r_i$ from the equation $\tilde{r}_i = 1.105r_i$. Crudely

*In the original definition of $r_i$, the denominator was $1.02\sqrt{\lambda_i}$ but $\lambda_i$ cannot be known precisely. We suggest that in calculating $r_i$, one should estimate $\lambda_i$ by $x_i$ because (i) $x_i$ is an unbiased estimator of $\lambda_i$, equally likely to be above and below and (ii) for $\lambda_i$ near 200 (a typical value for an American state), $\sqrt{x_i}$ will, on the average, deviate from $\sqrt{\lambda_i}$ by less than 3 per cent.



speaking, $\tilde{r}_i$ is the ratio of the observed prediction error in state i to the size of a typical prediction error caused solely by random fluctuations.

Table 1: NORMALIZED ABSOLUTE RESIDUALS FOR PASSELL PREDICTIONS OF STATE HOMICIDE LEVELS IN 1960

| State | Normalized Residual ($\tilde{r}_i$) | State | Normalized Residual ($\tilde{r}_i$) |
|---|---|---|---|
| Alabama | 4.13 | Montana | 2.03 |
| Arizona | 2.32 | Nebraska | .48 |
| Arkansas | 1.87 | Nevada | 1.34 |
| California | 21.07 | New Hampshire | .95 |
| Colorado | 1.65 | New Mexico | 1.44 |
| Connecticut | 1.36 | New York | 5.49 |
| Delaware | 1.68 | North Carolina | .96 |
| Florida | .78 | Ohio | .52 |
| Georgia | 3.41 | Oklahoma | 4.64 |
| Illinois | 3.30 | Oregon | .59 |
| Indiana | 1.86 | Pennsylvania | 2.84 |
| Iowa | 5.34 | Rhode Island | 6.17 |
| Kansas | .61 | South Carolina | 1.05 |
| Kentucky | .57 | South Dakota | 2.61 |
| Louisiana | .87 | Tennessee | 3.26 |
| Maine | 1.18 | Texas | 5.25 |
| Maryland | 1.85 | Utah | 4.97 |
| Massachusetts | .36 | Virginia | .53 |
| Michigan | 2.29 | Washington | 7.34 |
| Minnesota | 6.06 | West Virginia | 2.10 |
| Mississippi | 2.34 | Wisconsin | 11.68 |
| Missouri | .59 | | |

$\bar{K} = 3.07$

The $\tilde{r}_i$'s associated with the use of (8) appear in Table 1. (Because of missing data, predictions can be made for only 43 states in 1960.) The corresponding summary statistic $\bar{K}$ calculated from (6B) appears at the bottom of the Table.

That $\bar{K} = 3.07$ for equation (8) implies that Passell's predictions typically include a systematic error of about three standard deviations. We can estimate an average $\sigma_i$ by averaging the $1.02\sqrt{x_i}$ 's, which approximate the $1.02\sqrt{\lambda_i}$'s. The result ($\sigma_i$ 12) implies that a typical Passell prediction errs by roughly 36* because of the inaccuracy of (8) quite apart from random fluctuations. The cumulative systematic error for the full set

*We are treating the normalized residuals as uncorrelated with actual murder levels; Table 1 suggests this assumption is somewhat charitable to Passell.

of predictions is nearly 1600.

Consider two very different hypotheses on the deterrent effect of capital punishment, $H_0$: executions exert no deterrent effect, and $H_1$: each execution for murder deters, on the average, five killings in the state where it occurred. In 1960, 44 persons were executed for murder in the U.S.; the numbers for 1958 and 1959 were a bit lower. Thus, despite their large difference in content, $H_0$ and $H_1$ imply a difference of only about 200 in the aggregate U.S. homicide level for 1960. To discriminate meaningfully between the two hypotheses, a homicide model must be sensitive enough to pick up national effects of order 200.

Is Passell's equation (8), subject to national systematic error near 1600, equal to such a demanding task? We believe the best one-word answer is NO. We do not say this because of confidence that an erroneous preference for $H_0$ over $H_1$ contributed to Passell's systematic error. Rather we believe that, were his relatively large error somehow corrected, it is quite conceivable that the effect attributed to capital punishment would change in sign as well as magnitude. Though less extreme, the problem is somewhat akin to trying to count microscopic particles with a magnifying glass: the inadequacies of the device reduce greatly the relevance of the results of using it.

We now proceed to the studies of Ehrlich and Forst. For brevity we omit the analogs of Table 1 and are terser in interpreting calculated $\bar{K}$-values.

(ii) Forst

Unlike Passell and Ehrlich, Forst [3] tried to estimate not the homicide rates in different states, but rather the changes in these rates between 1960 and 1970. Using linear regression techniques on data from 32 states, he obtained a rule for approximating the difference $\Delta H$ between a state's 1960 and 1970 homicide rates:

$$H = 1422 + 17.64\Delta E - 5.97\Delta P - 24.91\Delta z + .0015\Delta CR + 39.60\Delta NW + .0047\Delta Y \quad (9)$$

where

P, E, and Y are defined as before

z = fraction of citizens living below officially defined poverty level



Cr = nonhomicide crime rate (i.e., commission rate for all crimes other than homicide)

NW = fraction of citizens who are nonwhite

and $\Delta$ refers to the changes in these variables between 1960 and 1970.

Forst performed other linear regressions after changing the definitions of certain variables and the weights given to data from different states; his results, as he pointed out, were largely unaffected. The coefficient of E in (9) is inconsistent with the hypothesis that executions deter homicides.

To evaluate Forst's model, we use (9) to estimate changes in the numbers of killings in the various states between 1960 and 1970. We generate an estimate of a given state's 1970 homicide level $M_{70}$ from the following equation:

$$M_{70} = P_{70}(H_{60} + \Delta H) \quad (10)$$

where $\Delta H$ is obtained from (9).

$P_{70}$ = the state's 1970 population (in 100,000's)

$H_{60}$ = the state's recorded homicide rate in 1960.

The change in the number of homicides over the 1960's is estimated as $M_{70}$ minus the recorded homicide total for 1960.

In our earlier notation, Forst attempted to estimate $\lambda_{70} - \lambda_{60}$ for each state (i.e., the mean of the probability distribution from which $X_{70} - X_{60}$, the observed change in the number of homicides, in one sample pick). Since $X_{70} - X_{60}$ has a variance of $1.04(\lambda_{60} + \lambda_{70})$, the normalized residual $\tilde{r}_i$ corresponding to our general framework follows:

$$\tilde{r}_i = \frac{1.131(d_i - a_i)}{1.02\sqrt{X_{60} + X_{70}}}$$

where $a_i = X_{70} - X_{60}$ in state i, $d_i$ = estimate change in state i's homicide level according to (10), and 1.131 = correction that arises because 7 parameters of Forst's model were estimated from 1960 and 1970 data.

Like Forst, we were unable to obtain the data needed to use (9) for 18 American states. For the remaining 32, the $\tilde{r}_i$'s were calculated and a $\bar{K}$-value of 1.58 was obtained. By averaging the $1.02\sqrt{X_{60} + X_{70}}$'s for the 32 states, we reach the estimate of 16 for a typical $\sigma_i$, and therefore the estimate of about 18 X 1.58 = 28 for a typical systematic error in a Forst prediction. For his entire set of 32 predictions, we estimate a systematic error near 900.

In 1960, there were 28 executions for murder in the states Forst considered; in 1970 there were none. Thus, once again, the systematic error of the model seems far too large to allow confidence in its estimate of the deterrent effect of capital punishment.

Ehrlich

Ehrlich, like Passell, performed a cross-sectional regression analysis on data from the various American states. However, he assumed a multiplicative form for the relationship between homicide rates and the variables that influence them (i.e., $H = CX_1^{a_1}X_2^{a_2} \ldots X_n^{a_n}$, which implies a linear relationship between log H and the log $X_i$'s). Representative of his results is the following model for 1950 for the 35 states that had positive execution rates:

$$\log H = -4.62 - .459 \log T - .687 \log P - .272 \log E + .496 \log NW + 1.2 \log X + 1.23 \log W + .274 \log A - .742 \log V \quad (11)$$

where

W = median income in 1949

X = fraction of families with income below W/2

V = percent of state residents who live in urban areas

Other variables are as defined before, except that E is based on all executions, not just those for homicide (i.e., number of executions for murder or rape/number of killings).

Under (11) a state's rate of homicide is proportional to $E^{-.272}$; thus, executions would appear to deter killings. For the 35 states considered, the average value of $E^{-.272}$ in 1950 was 2.06; the ratio $(E/2.06)^{-.272}$ would therefore be the magnification (or shrinkage) factor in a state's homicide rate associated with the deviation between its own frequency of executions and aggregate patterns. Calculations based on this observation lead to the approximation that, on the average, each execution in 1950 averted 9 homicides in the state where it occurred. Correspondingly, the roughly 100 annual executions in the U.S. near 1950 may each year have saved about 900 potential murder victims.

We proceed in the usual fashion to estimate the systematic error in (11), and obtain $\bar{K} = 2.66$. Since a typical $\sigma_i$ for the 35 predictions is slightly over 13, we obtain 1200 as an estimate of Ehrlich's overall model-based error.

The systematic error in Ehrlich's model, we can observe, exceeds the rather considerable effect he attributes to capital punishment. While the former does not dwarf the latter, it is still large enough to raise grave doubts about the accuracy of Ehrlich's results. This is especially so since Forst's model, which assigns no deterrent effect to executions, contains a slightly lower level of systematic error.* The disagreement between the "equally matched" studies of Forst and Ehrlich only underscores how unclear their models leave us on how the chips would fall if a perfect model could ever be found.

## IV. Final Remarks

It would be improper to jump from these results to the conclusion that no data analysis can reliably indicate the deterrent effect, if any, of capital punishment. While all three studies we examined included a series of models, each contained very little variation in the functional forms it considered. What the results really suggest is that highly accurate models of homicide levels include few if any that ultimately are calibrated with a linear regression computer package. It is certainly conceivable that homicide models arising from other perspectives might be useful even for the delicate task these studies attempted.

*Forst did, to be sure, consider later years when executions were rarer; yet, if Ehrlich's 9-1 ratio is extrapolated forward -- and Ehrlich has always assumed the effect of executions to be time-invariant -- Forst's "holding his own" against Ehrlich does not lose its significance.

Going beyond any specific policy question, we find it significant that it seems possible, through operations research reasoning, to develop quantitative standards of accountability in some research areas where they are now weak or non-existent. Such standards might well enhance the precision of discussion on the strengths and weaknesses of mathematical models that attempt to describe social phenomena. Attempting to develop them in a variety of public-sector research areas could, we believe, be a fruitful direction for future effort.

## References

1. A. Blumenstein, J. Cohen, D. Nagin (editors), Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime Rates, National Academy of Sciences, Washington, D.C. (1978).

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

3. B. Forst, "The Deterrent Effect of Capital Punishment: A Cross-State Analysis of the 1960's," Minnesota Law Review, 61, (1977), p. 743.

4. G. Goehrke, D. J. Kaplan, W. Walthall, "On the Deterrent Effect of Capital Punishment," unpublished masters' thesis, M.I.T. Sloan School of Management, (1977).

5. P. Passell, "The Deterrent Effect of Capital Punishment: A Statistical Test," Stanford Law Review, 61, (1975), p. 80.



## APPENDIX A
### On the Independence of Potential Killers

This paper's estimate of the role of fluctuations in homicide levels was based on the assumption that potential killers act independently (i.e., that $x_A$, the number of killings by A, is unaffected by the comparable quantity $x_B$ for B). We attempt here to show that the imperfections of that assumption are second-order effects that can reasonably be neglected.

It is clearly possible that both $x_A$ and $x_B$ are simultaneously affected by conditions that prevail in the community, e.g., the pattern of punishment for homicide. But such systematic factors are ideally already reflected in the probability distributions assigned to $x_A$ and $x_B$.* By independence we mean that the probability of $x_A$ given that $x_B = i$ is independent of i, for all A and B.

There are three major reasons one might suspect $x_A$ and $x_B$ to be related; we consider them separately below.

(1) Some homicides are committed in revenge for other homicides.

This is true, but the number of solved killings in which such vengeance was the motive is extremely small. Indeed, in its breakdown of the causes of homicide, the F.B.I.'s Uniform Crime Report doesn't even include such a category.

(2) Some killings are stimulated by the publicity given to others.

While this contention sounds plausible, the weight of evidence is against it. The most publicized killers in recent years -- the Boston Strangler, the Son of Sam, the Zodiac Killers, the Manson "Family," the Zebra Killers, the Hillside Strangler -- have not, so far as is known, spawned even one "copycat killing" despite the extraordinary publicity they aroused. One might argue that the dependence relationship is more indirect: that increases in the frequency of homicide make it less unthinkable as a way of dealing with "problems" and might thus stimulate additional killings. But even if this is so, homicide in the U.S. has risen gradually rather than abruptly

*Since homicide studies are based on patterns of past years, their authors presumably have actual values for the variables they believe affect homicide levels.

and perceptions about its frequency are probably based far more on the experience of several years than on any short period. Thus, assuming any appreciable correlation between A's and B's actions on this "climate of violence" grounds seems farfetched.*

(3) <u>If A kills B, B's capacity for homicide is somewhat reduced</u>.

Quite so. But that this effect is second-order is shown by considering an extreme case in which A and B are literally mortal enemies. Suppose that

$$x_A = \begin{cases} 1 \text{ w.p.}^{**}\ P_A \\ 0 \text{ w.p. } (1 - P_A) \end{cases}$$

$$x_B = \begin{cases} 1 \text{ w.p. } P_B \\ 0 \text{ w.p. } (1 - P_B) \end{cases}$$

and that, if either kills, his victim will be the other (i.e., $x_A + x_B \leq 1$). $x_A + x_B$ is the total contribution of these two people to the aggregate homicide level; it is easy to show that, if one falsely assumes that $x_A$ and $x_B$ are independent, his estimate of the mean of $x_A + x_B$ is correct while his variance estimate errs by $2P_AP_B$, which is negligible if both $P_A$ and $P_B$ are assumed small. Hence, the practical effect of the deviation from independence is insignificant.

One could argue that perhaps $P_A$ and/or $P_B$ is not small, and thus the argument is invalid. But this is tantamount to saying that there are individuals for whom the probability of committing murder in a one-year period is substantial; those advancing this view must explain why we lack any capacity, despite decades of research, to identify them.

All things considered, there is little basis for assuming that our independence assumption -- while admittedly an idealization -- introduces any serious errors into the analysis in the paper. Even if it were somehow replaced by its perfect counterpart, it seems highly unlikely that any of this paper's conclusions would change.

---

*More precisely, we are assuming that neither can greatly affect the other's perception of the level of killing.

**w.p. = with probability



## APPENDIX B

### Reconstructing the Data Bases of the Passell, Forst, and Ehrlich Studies

In gathering data to calculate $\bar{K}$-values, we tried as much as possible to use the precise definitions given and the data sources cited. In a few cases, however, the definitions of particular variables were a big ambiguous. We therefore took additional steps to ensure that we did not misrepresent an author's model. Forst's paper includes a table that lists the average value over his 32 states for each of his variables; we thus calculated the corresponding averages for our own data. The two sets of results were almost identical. For Passell and Ehrlich, who did not provide such detailed information, we actually performed regression analyses on our data similar to those they described. Comparisons of coefficients revealed no serious discrepancies. Indeed, we calculated residuals (i.e., $\tilde{r}_i$'s) both with the published coefficients and those we obtained; only trivial differences arose. In all, we are confident that no misunderstandings or accidents distorted our results.

END