Deterrence: Evidence and Inference*

Isaac Ehrlich†

Because of space limitations and the short time I have been given to prepare my response to the two critiques of my work published in this issue of the Journal, I shall confine myself principally to the critique by Bowers and Pierce. I do that not because the paper by Baldus and Cole does not warrant a detailed reply but because an elaborate response to the central issues they raise is contained in a study of mine now in progress which deals critically with published research by Sellin and others. I choose to focus on the Bowers and Pierce piece also because, as I hope to show, their work largely misinterprets and misapplies the framework I have developed for testing the deterrence hypothesis. Addressing their work critically provides the opportunity to elaborate upon some pertinent aspects of my research, hopefully for the benefit of interested scholars.

The conclusions of my time series study of murder basically are two: (1) that previous research never adequately tested a set of direct and specific implications suggested by a general theory of deterrence and (2) that my empirical findings, while tentative and inconclusive by the very nature of observational statistics, are not inconsistent with rather sharp implications emanating from this theory, including the hypothesized deterrent effect of the conditional risk of execution. Bowers and Pierce tacitly accept the first conclusion and seek to evade the second, evidently on faulty grounds. In their efforts to obscure the empirical findings, they have selectively deleted observations, utilized an inferior regression specification, considered irrelevant variables and correlations, and revealed in the process misunderstanding of elementary statistical concepts, as I discuss in points III, IV, and V below. They do not provide evidence based on a systematic statistical analysis showing that capital punishment, or punishment in general, does not deter crime. Essentially, they make only the point that the observed deterrent effect of the risk of execution can be confounded

* This reply has been prepared in cooperation with Randall Mark.
† Associate Professor of Business Economics, University of Chicago; Research Associate, National Bureau of Economic Research.
3. See, however, the discussion in Parts V, VI, and VII supra in response to Baldus and Cole as well as Bowers and Pierce. See also notes 33, 44, 48 & 50 infra.

209
The Yale Law Journal

Vol. 85: 209, 1975

when insufficient regard is shown for proper methods of hypothesis
testing. Indeed, my principal response to Bowers and Pierce is that
they concern themselves only with making a point (i.e., confounding
a result) rather than testing a hypothesis. There is a fundamental dif-
ference between a systematic and statistically coherent test of all the
ramifications of a general hypothesis and an exercise in search of that
set of circumstances which—for purely technical reasons—may weaken
the effect of a single variable within a comprehensive model. Bowers
and Pierce do not analyze the effects of variables other than the con-
ditional probability of execution—variables such as estimates of ap-
prehension and conviction risks and unemployment and labor force
participation rates. Moreover, they do not address themselves at all to
the question of the optimal form of testing for the deterrence hypo-
thesis, given the technical limitations of observational statistics.

I shall respond in some detail only to the substantive issues raised
by Bowers and Pierce and to a few of the arguments advanced by
Baldus and Cole. Also, I shall point out only some of their more
serious errors. For expositional convenience, the discussion below is
ordered as seven general points.

I. Corroborating Evidence from Bowers and Pierce

First and foremost, the Bowers and Pierce work, however inadver-
tently, has lent considerable strength to the case for the deterrent ef-
fect of capital punishment, because their application of the theory and
econometric methods outlined in my paper over the entire period con-
sidered in my analysis produces results quite similar to my own. This
is noteworthy for several reasons.

A. Their data set is not identical to the set of data that I
utilized. Their comments indicate that at least four variables have been con-
structed differently in their work.

B. In addition, Bowers and Pierce do not accurately execute my

4 A detailed description of the data set that I utilized is contained in the memoran-
dum, I. Ehrlich, The Deterrent Effect of Capital Punishment: A Question of Life and
Death, American Economic Review (June, 1975): Sources of Data, May 1975 (on file with
Yale Law Journal). Mr. Bowers, among others, received a draft of this memorandum,
which noted some errata in the published paper, and which was completed in July,
1975, by Randall Mark.

5 Bowers & Pierce, supra note 1, at 194 n.18. They allege that one of these dif-
ferences, that concerning XPOL., is due to my having used an unspecific auxiliary
analysis for estimating missing values of this variable. Their allegation is false: my
memorandum on data sources, supra note 4, at 17-19, outlines in detail the auxiliary
procedure utilized. In addition, Bowers and Pierce utilize different values of P'cla than
did I for the years 1933, 1934, and 1935, because they did not choose to apply the pro-
cedure I outlined for estimating these values. Their reference to “Ehrlich’s conviction
rates” for these years is misleading. Bowers & Pierce, supra at 191 n.15.
regression analysis in which two measures of execution risk are used to test the hypotheses of the model. Their results in connection with these measures of execution risk are likely to be inferior to my own because they omit an observation; therefore their results are based on fewer degrees of freedom.\(^6\)

C. Furthermore, Bowers and Pierce may have used computational procedures different from those I utilized. That their regression results and coefficients of serial correlation always differ from mine is consistent with this view.\(^7\)

Yet despite these differences, the regression results which Bowers and Pierce report are “similar to Ehrlich’s.”\(^8\) More particularly, they confirm not only the apparent restraining effect of the conditional probability of execution on the murder rate but also my predicted ranking of estimates of elasticities of the murder rate with respect to the three deterrence variables—the probability of apprehension, the conditional probability of conviction, and the conditional probability of execution.\(^9\) This is at once a confirmation of the strength of my approach and a corroboration of the basic findings. The results stand out in contrast to previous allegations that no evidence exists suggesting that the death penalty may have a restraining effect on the frequency of murder in the population. How Bowers and Pierce attempt to obscure these results is the subject of the three following points.

II. The Effects of Data Imperfections

Bowers and Pierce attempt to invalidate the results of my empirical investigation of the effects of deterrence variables (and these results

\(^6\) Bowers and Pierce suggest that all regressions of mine in which \(PXQ\_1\) and \(TXQ\_1\) are used as measures of the conditional probability of execution are erroneous. Bowers & Pierce, supra note 1, at 195 n.24. They argue that I mistakenly utilized \(PXQ\_1\) and \(TXQ\_1\) for the effective periods 1935-1969 and 1937-1969, respectively, whereas, they claim, the absence of data on \(PXQ\_1\) in 1933 would require starting the effective period one year later in each case. They then speculate that an “erroneous (probably zero)” value of \(PXQ\_1\) must have been inserted by the computer for 1933. \(Id.\) Contrary to their speculation, however, \(PXQ\_1\) in 1933 was estimated for use in all regressions with \(PXQ\_1\) and \(TXQ\_1\) over the periods 1935-1969 and 1937-1969, respectively. The value of \(PXQ\_1\) in 1933 is estimated as 7.16299 (percent). The estimate is based on published FBI data for the murder rate and for the murder clearance ratio in 1932 and on an estimated conviction risk in that year. The degrees of freedom associated with a regression are the excess of data points over the number of regression coefficients estimated.

\(^7\) The calculations I reported were computed via the Econometric Software Package, into which R.C. Fair’s three-round procedure is integrated. See J. Cooper, ECONOMETRIC SOFTWARE PACKAGE (1973). Bowers and Pierce provide no documentation of their computational procedures.

\(^8\) Bowers & Pierce, supra note 1, at 196.

\(^9\) \(Id.\) at 197.
alone) by posing, in effect, the extreme argument that data limitations prior to the 1960's preclude any empirical test of the deterrence hypothesis.\textsuperscript{10} I do not agree with this general conclusion, and, in addition, I find their analysis concerning particular variables to be superficial and not constructive.

With respect to the dependent variable, the murder rate, the figures I have used are based on the Federal Bureau of Investigation's revised estimates of annual total murders and nonnegligent manslaughters. For the purposes of the empirical investigation, the FBI data are conceptually superior to the homicide series published in the \textit{Vital Statistics of the United States} because the FBI category is defined to include only willful felonious homicides. Law enforcement officials, not health officials, bear the responsibility and undergo training for distinguishing willful felonious homicides from other homicides.\textsuperscript{11} Indeed, by definition, the homicide data of the \textit{Vital Statistics} explicitly include justifiable homicides and are likely to include some negligent manslaughters as well.\textsuperscript{12} Moreover, the revised homicide data of the FBI reflect that agency's unique opportunity to incorporate into its estimates whatever homicide data have been collected by health officials. In contrast, the homicide figures tabulated for the \textit{Vital Statistics} are never revised after the cutoff date for data collection for a given year.\textsuperscript{13} In addition, information on death certificates, on which the \textit{Vital Statistics} figures are based, sometimes may reflect classification of deaths only by medical cause rather than by external cause (such as accident or homicide), especially in those instances when death from homicidal assaults occurs later than the time of assault.

Because the \textit{Vital Statistics} category includes some nonfelonious homicides, one would expect its homicide counts to exceed the FBI

\textsuperscript{10} See id. at 187-92. This direct implication of their remarks notwithstanding, Bowers and Pierce subsequently claim that their regression results are "more adequate and consistent" when they examine only subperiods ending early in the 1960's. \textit{Id.} at 198. Yet by their own analysis, data imperfections clearly are less important in the 1960's than in earlier periods. \textit{Id.} at 190-92.

\textsuperscript{11} Although health officials are expected, when they can, to distinguish homicides from other violent "external" causes of death, namely accidents and suicides, they clearly experience much difficulty in drawing these distinctions. As late as 1970, for example, in which the \textit{Vital Statistics} reports 16,848 homicides, a total of 5,384 violent deaths are classified as due to undetermined cause, \textit{i.e.}, either accidental or purposeful. \textit{II U.S. PUBLIC HEALTH SERV., DEPT OF HEALTH, EDUCATION, AND WELFARE, VITAL STATISTICS OF THE UNITED STATES 1970, at 226.}

\textsuperscript{12} Homicides inflicted through legal intervention by police are included in the homicide count reported by the \textit{Vital Statistics}. Only beginning in 1949 can such homicides be separated from the total.

\textsuperscript{13} This failure to revise leads to an undercount of homicides pertaining to a given year because homicides occurring in that year that were not reported by the cutoff date are not added subsequently to that year's total.
estimates. Yet, in every year from 1939 through 1961, the more narrowly defined FBI revised estimates are higher, by more than 900 in some years. These comparisons suggest the importance of the technical considerations cited above that may lead to a significant underestimation of the relevant number of criminal homicides in the Vital Statistics. Tests which I have been conducting with independent bodies of data, and which I hope to report in the near future, indicate, nevertheless, that the effects of the deterrence variables—including the conditional probability of execution—on the homicide rate as reported by the Vital Statistics are qualitatively the same as those found in my time series investigation.

With respect to the empirical measures of the apprehension and conviction risks, the critics raise the issue of whether the FBI data are “unrepresentative” in the earliest years of the sample period. Yet I have reported results demonstrating that the basic findings of my investigation are observed even without the presence of these early years in the observation set. Moreover, their inference that the conviction data in 1936 and 1937 are biased because of a particular pattern observed in 1974 is purely conjectural and not founded upon any systematic analysis. Imperfections in data notwithstanding, my qualitative results in connection with the effects of apprehension and conviction risks over my entire sample period have been observed by others using the same FBI sources.

More basically, though, the critics’ attempt to discredit the empirical investigation on the basis of data quality is self-defeating to their own case. As is well known, “errors of measurement,” as they are termed in the econometrics literature, generally lead to underestimation of the true effects of an explanatory variable in a simple regression analysis when that variable is subject to random measurement imperfections. If the true variables of interest were grossly misrepresented by their empirical counterparts, then the statistical implementation of the theory should have failed to demonstrate any of the effects theorized. In particular, only a remarkable coincidence could then explain the fact that, as predicted by my theory, the findings show that the probability

14. Bowers and Pierce’s statement, supra note 1, at 189, that the FBI estimates differ from those of Vital Statistics by only about three percent for the period after 1939 is quite misleading in view of the magnitude of the year-to-year differences.

15. Bowers & Pierce, supra note 1, at 190; see id. at 190-92.

16. See, e.g., Ehrlich, The Deterrent Effect of Capital Punishment: A Question of Life and Death, 65 AM. ECON. REV. 397, 410 (1975) (equation 5 of Table 3 and equation 7 of Table 4).


of arrest (measured by the FBI's clearance ratios) had a proportionally larger impact on the murder rate than the conditional probability of conviction (derived from the FBI's statistics) and that the conditional probability of execution had the least effect. The FBI's reported figures on arrest and conviction in the 1930's surely were not tailored to my theoretical predictions 30-odd years later.

III. The Effects of Deleting Observations

Curiously, Bowers and Pierce's remarks in the beginning of Part II of their paper may convey the impression that I have artificially restricted the time span of my empirical investigation to test deterrence effects. The fact is that my empirical investigation was conducted for the longest time period for which necessary data were available at the time of my study. In contrast, Bowers and Pierce conduct the bulk of their analysis over arbitrarily restricted subperiods after deleting specific observations from the complete data set. Their assertion that "Ehrlich's regression model fits the data better without them" is based upon an erroneous method of inference, as I point out in Part V below. Here I shall address the consequences of discarding the information provided by the observations that they choose to omit.

Selective elimination of a sufficient number of observations from a regression analysis is a virtually foolproof method for reversing any single result derived from an original sample. Imagine, for example, a regression line verifying a negative association between the quantity demanded of corn and the price of corn. Since a majority of the data points typically will not lie on the regression line, the selective exclusion of data points can easily turn a significant negative relationship into an insignificant negative association or even a positive one. Such exclusions are particularly disturbing when the entire sample is relatively small. Indeed, the elimination of data points relating to murder in the 1960's—accounting for 17 percent or more of the full sample which I investigated—amounts to, in practice, the selective, nonrandom exclusion of observations crucial to an efficient estimation of the effects of key deterrent variables.

A. Omitting observations from the 1960's drastically reduces the variability in estimates of the conditional probability of execution and in the modified rates of change of these estimates, which are the actual regressors in the analysis. Whereas the rates of change in this variable had been quite stable for the preceding two decades, the objective (measured) risk of execution declined quite sharply starting about

The sharp movements in the rates of change of the conditional risk of execution are, of course, not my invention—they accurately reflect the objective trend in the *true* risk of execution in the 1960’s. In addition, murder clearance ratios, used as objective estimates of the probability of arrest for murder, also exhibit little variability over specific subperiods ending in the early 1960’s. Thus, variability in arrest and execution risks is particularly small between the late 1930’s and the early 1960’s.

It is a well known principle that a minimum amount of variability is necessary to perform a regression analysis. Indeed, an *efficient* sample designed for the purpose of estimating via a regression format the partial effects of specific explanatory variables on a dependent variable is one which maximizes the range of variability in the regressors. Imagine, again, the attempt to confirm the negative association between prices and quantities demanded of corn. The attempt would fail if the subperiod selected by a researcher for his regression analysis is one in which corn prices or their rates of change, whichever are relevant, are relatively stable. The appropriate inference, however, is not that the theory of demand fails to explain movements in corn purchases but rather that the selected subperiod cannot be utilized to estimate the partial effect of corn prices on quantities demanded. The importance of such considerations is apparent in my time series analysis not only in connection with deterrence variables but also in

20. The fact that this trend concerns only 20 states that imposed executions in 1960 has no specific relevance in connection with the estimated deterrent effect of execution, contrary to the unsystematic discussion of Bowers and Pierce, *supra* note 1, at 204. Only a subset, and typically a minority, of states enforced the death penalty in any given year. Many fewer than 20 applied the death penalty in every year from 1933 to 1967. Moreover, aggregating over states which had executions and states which did not implies that the estimated elasticity of the murder rate with respect to execution risk is likely to be biased downward relative to the true elasticity. *See* Ehrlich, *supra* note 16, at 408.

The inferences drawn by Bowers and Pierce, *supra* note 1, at 201-04, concerning murder rate and execution changes in particular states during the 1960’s, are based on a faulty methodology. Their analysis of yearly changes in murder rates and executions in specific states over the six-year period 1962-1967 fails to control for the different trends in apprehension and conviction risks or in any other relevant determinants of murder across these states. It considers the absolute number of executions rather than the theoretically relevant conditional risk of execution in each state. It is based on comparisons of contemporaneous changes in executions and murder rates, comparisons which are not capable of identifying the relevant causal relationships in specific cases. Furthermore, the overwhelming trend in all the states compared over this subperiod was toward a complete cessation of executions. The occasional deviations from that trend in each state in a single year cannot serve as a statistically meaningful basis for the classification of states into those with increasing or decreasing executions. These fundamental shortcomings in Bowers and Pierce’s independent analysis pertaining to the 1960’s further underscores the importance of applying the *full* econometric framework utilized in my study over as long a period as the data permit in order to isolate the deterrent effect of the death penalty.
relation to the effect of the unemployment rate on the frequency of murder. While the association between the latter two variables is found to be positive over the entire sample period, this association weakens substantially when the subperiod relating to the 1930's is excluded from the observation set. The significance of the 1930's in connection with movements in the unemployment rate is well known.

B. A related consideration is that the estimates of the objective risk of execution show a strong time trend over the subperiod ending in the early 1960's. Specifically, the graph plotting the logarithms of these estimates over the time period from the late 1930's to the early 1960's appears to be nearly a negatively sloped straight line. Over the same subperiod the murder rate in the United States also exhibits a continual, systematic negative trend. Because of the significant negative trend in estimates of both the frequency of murder and the conditional risk of execution, the estimated effect of the latter on the former over this specific subperiod may simply reflect the effect of pure time trend.

The differences between regression results based on the full sample and those based on subperiods ending in the early 1960's, including those reported by Bowers and Pierce in their appendix B, are consistent with the arguments developed above. Not only does the effect of execution risk appear to become quite weak as data points are deleted, but the effect of apprehension risk also becomes weak due to lack of variability. As indicated above, the effect of unemployment is also sensitive to the deletion of specific subperiods. More importantly, the negative partial effect of the time trend on the murder rate, which was verified over the entire period considered in my study, becomes "insignificant" in Bowers and Pierce's regression results for subperiods ending in the early 1960's, although over those specific subperiods the murder rate continually declined. Hence the effect of pure trend should have been found to be even more pronounced over these periods. The fact that the effects of both the risk of execution and time trend weaken suggests a high degree of competition or multicollinearity between the two variables over subperiods ending in the early 1960's.

C. The absence of variability and the presence of multicollinearity


22. The term "significance" here is used heuristically only to indicate relatively low standard errors. Standard tests of significance do not apply in connection with simultaneous equation estimation procedures.

23. For an explanation of this statistical term, see J. Johnston, supra note 18, at 159-67.
are hardly unique to my study but occur frequently in time series regression analyses. The conventional remedy to such problems is to extend the sample size so that variability may be enhanced and the separate effects of highly related explanatory variables may be extricated and identified. I have pursued this procedure from the outset in my efforts to extend the sample size into the early 1930's and up to the late 1960's. Bowers and Pierce pursue just the reverse course. Following my own report of weak results obtained from subperiods ending in 1963, they go on to perform most of their independent estimations over subperiods in which little meaningful analysis can be conducted.

D. As for the unfounded claim that somehow "significant" deterrent effects of execution risk and other punishment variables are "present" only in a short span of time relating to the 1960's and are not to be found in earlier (or later) periods, I plan to demonstrate in the near future, through evidence based on state-wide data in earlier years, that the deterrent effects of certainty and severity of punishment—including punishment by execution—are not uniquely associated with a specific set of data points.24

IV. The Merits of the Logarithmic Format

In the empirical implementation of my general theory of participation in illegitimate activities,25 I consistently have emphasized a logarithmic-linear specification of the relevant equations. However, as my following comments will show, while the logarithmic-linear specification appears to be preferable on analytical and experiential grounds, the qualitative results from my time series study do not depend exclusively on this functional form.

The logarithmic-linear form can be justified on practical grounds if the elasticities of a dependent variable with respect to a set of explanatory variables are assumed to be constant to the first order of approximation. A logarithmic-linear specification is a superior regression format when the magnitude of the errors in the data are thought to be proportional to the level of the variables that the data purport

24. Further independent evidence on this point is provided by J. Yunker, The Deterrent Effect of Capital Punishment: Comment, Oct. 1975 (unpublished manuscript, on file with Yale Law Journal). Yunker's analysis, conducted over a sample period extending through 1974, yields results indicating that the murder rate declines as the risk of execution rises. The statistical methodology he uses, however, is different from my own.

to measure. It would be rather implausible, for example, to assume that the magnitude of underreported and misreported crime is independent of the level of reported crime. This assumption is implicitly invoked by use of the linear specification in the natural values of the variables, the functional form stressed by Bowers and Pierce. More plausibly, one may assume that the magnitude of reporting errors is proportional to the level of the relevant statistics. Furthermore, in the case of murder investigation, the dependent variable of interest is the rate of capital murder rather than the rate of criminal homicide, the actual measure utilized. Previous researchers have assumed that the capital murder rate and the total homicide rate are proportionally related. It would be convenient at least, then, to use a logarithmic-linear specification which enables a direct estimation of elasticities. Moreover, similar considerations also apply to the observed probabilities of apprehension, conviction, and execution, each of which is based upon data relating to reported willful felonious homicides rather than the true level of capital murders. All these considerations suggest that the efficient functional form underlying the murder supply function is likely to be one that utilizes the logarithms of the dependent variable and the key independent variables. And indeed, "prior information" accumulated through my past work on crime has led me to emphasize the logarithmic-linear regression format because of its observed relative efficiency.

Bowers and Pierce could have tested statistically for the optimal functional specification of the estimated supply-of-offenses function. In particular, they could have examined the efficiency of the logarithmic-linear specification relative to a specification that is linear in the natural values of all variables. In research now in progress, I have conducted statistical tests of optimal transformations based on a likelihood ratio method. As I plan to demonstrate, the conclusion emerging from these tests is that the logarithmic-linear format not only is decisively superior to the format using the natural values of the variables but that the former generally cannot be rejected as the optimal form within the class of single-parameter power transformations. Bowers and Pierce's demonstration that their regression results tend to deteriorate when running the regressions with natural numbers shows only that they evade the question of which of the two transformations is more appropriate. Hence they not only prefer testing the deterrence

27. For an elaboration of this point, see Ehrlich, Illegitimate Activities, supra note 25, at Section IV, App. 1.
hypothesis for subperiods of sharply limited usefulness, but they also prefer an inferior regression format.

As I noted in my paper, the basic results from that study were found to be unaffected qualitatively by the choice of functional form. The regression equation I report below was performed with the antilogs of the same set of variables used to derive the results reported in my published paper, and indicates that the qualitative deterrent effects of apprehension, conviction, and execution are not exclusively dependent on a specific functional form. Although the specification in natural numbers appears to be clearly inefficient in view of prior information and other tests, time series regression estimates derived via that format in this equation nevertheless indicate the existence of the expected deterrent effects.

V. Basic Statistical Errors in My Critics' Work

Part II of Bowers and Pierce's critique refers to "errors" in my analysis. The fact is that I have learned of no single error in either my theoretical analysis or the statistical methodology used to implement the theory. In contrast, the work by Bowers and Pierce is riddled with errors and demonstrations of misunderstanding of basic statistical principles. A few examples will illustrate.

A. The "$R^2$" statistic is essentially irrelevant in connection with a two-stage least squares regression analysis or the related three-round procedure used in my study. Yet, with no qualification whatever, Bowers and Pierce cite this statistic in conjunction with every estimated equation which they report. More seriously, they use the "$R^2$" statistic as a basis for inference.

B. The "$R^2$" statistics, as well as standard errors of the regressions, computed for subsamples having different ending dates are relied on

The estimated equation above pertains to the effective period 1935-1969. The murder rate is represented by $q^o$; the risks of apprehension, conviction, and execution are represented by $P^a$, $P^c$, and $P^x$, respectively; other symbols can be interpreted from Table 2 of my published paper, Ehrlich, supra note 16, at 409, and the numbers in parentheses denote ratios of estimated regression coefficients to their standard errors. The estimate for $\beta$ is -.119. Not only is the direction of the effect of every variable the same as is found with the logarithmic-linear specification, but the predicted ranking of the elasticities of the measures of the three deterrence variables also is observed.

J. Yunker, supra note 24, performs regressions with natural numbers only. His regression analysis also indicates the existence of a deterrent effect associated with the risk of execution in the United States.

Bowers & Pierce, Deterrence, supra note 21, appendix B.
by Bowers and Pierce to draw the startling inference that regression results are "more adequate and consistent for the periods with earlier ending dates."\(^3\) Their lack of regard for, indeed their apparent lack of awareness of, the diminishing degrees of freedom associated with smaller and smaller subsamples is astonishing. Even if the regression analysis were based on the ordinary, or classical, least squares procedure, comparison of \(R^2\) statistics and standard errors of the regressions across subsamples with successively smaller degrees of freedom would be improper: as Bowers and Pierce surely know, when the number of degrees of freedom is zero, the \(R^2\) statistic necessarily is unity. It should be noted that the deletion of observations from the full set of observations led Bowers and Pierce to estimate regression equations in which the number of observations is as low as 22 while the number of parameters estimated in the reduced-form regression analysis is 19. They apparently fail to recognize that estimated regression coefficients based on successively smaller degrees of freedom become increasingly imprecise. In particular, estimates based upon so few degrees of freedom as those Bowers and Pierce stress are unlikely to be adequate for rejecting the hypothesis of no deterrent effects.\(^3\)

C. At the end of their article, Bowers and Pierce report results from regressions in which they replace the conditional probability of execution—the theoretically relevant variable—with the absolute number of executions.\(^3\) A finding of a positive association between homicide rates and the number of executions is cited as evidence of the "brutalizing" effects of capital punishment. Bowers and Pierce do not recognize, however, that the positive association may be expected on purely technical grounds. Where there are no homicides there can be no convictions or executions. In contrast, when murder rates rise, with arrest, conviction, and execution risks constant or rising, the number of convictions and executions also would rise. A positive (or zero) association between criminal homicides and the number of executions thus is hardly surprising. Indeed, more than 20 years ago this relationship was noted by Schuessler, who recognized that this correlation does not constitute a test of the deterrence hypothesis.\(^3\)

\(^32\). Bowers & Pierce, supra note 1, at 198.
\(^33\). Unfortunately, on this point, as others, Baldus and Cole, supra note 2, at 180 & n.25, 184-85, uncritically accept the inferences drawn by Bowers and Pierce and other critics and, thus, implicitly commit the same errors. Without providing any valid grounds for their assertions, Baldus and Cole even more emphatically assert their judgment that the subperiod analysis provides estimates somehow superior to those obtained from the full sample.

For a definition of the term "degrees of freedom," see note 6 supra.
\(^34\). Bowers & Pierce, supra note 1, at 205 n.45; Bowers & Pierce, Deterrence, supra note 21, at 24-26.
\(^35\). Schuessler, The Deterrent Influence of the Death Penalty, 284 Annals 59 (1952). The Schuessler work is cited by Bowers and Pierce, Deterrence, supra note 21, at 2 n.5,
D. Although I have focused in this discussion on errors contained in the Bowers and Pierce paper, I should point out that the Baldus and Cole paper also is seriously flawed. For example, Baldus and Cole criticize my study because "the second-stage regression does not hold fixed" the cluster of variables denoted as $X_2$ in Table 2 of my published work. However, these variables serve in the role of "omitted exogenous variables" in the murder supply equation. The remark shows that Baldus and Cole do not understand the simultaneous equation estimation framework underlying my study. Exogenous variables such as those summarized under $X_2$ must not be held constant in estimating the structural murder supply function. If they were, the two-stage least squares procedure would become meaningless. Their effects, however, are integrated appropriately in the estimation procedure through their incorporation in the reduced form regression analysis. This error by Baldus and Cole betrays quite a fundamental misunderstanding of the methodology which they have undertaken to evaluate. Indeed, in their comparisons of my research with that of previous researchers, they do not note the prime benefit of utilizing the simultaneous equation framework—that it attempts to identify the direction of the causal relationship between the frequency of murder and the deterrence variables. The direction of that relationship is not self-evident. For example, in times of high murder rates, states may be more inclined to convict and execute offenders, or even reinstate the death penalty, than in times of low murder rates. The simple association between these variables thus may give the appearance that, for example, an increase in the risk of execution "leads" to an increase in the murder rate, even though the opposite is true. A simultaneous equation estimation framework is designed to identify causal relations and thus avoid potentially biased results.

VI. The Gap Between Evidence and Inference in Previous Research

Bowers and Pierce and, in greater detail, Baldus and Cole, rely heavily on previous research, especially that of Thorsten Sellin, as well as by Baldus and Cole, supra note 2, at 183 n.45, who also fail to recognize the technical association between homicide and execution rates. It might be noted that Schuessler, at 59-60, went on to attempt a test of the deterrence hypothesis by measuring the simple correlation coefficient between homicide rates and a measure of the risk of execution in 41 executing states over the period 1937-1949. Although he found the correlation coefficient to be —.26, indicating that the murder rate falls as execution risk rises, and although this negative association turns out to be significant statistically at the five percent level for a one-tail test, neither Bowers and Pierce nor Baldus and Cole mention this result.

36. Baldus & Cole, supra note 2, at 182 n.41.
evidence against the deterrence hypothesis. In this section I shall briefly comment on the research by Sellin and others concerning the deterrence hypothesis.

The principal shortcoming of Sellin’s research and related work is that the approach taken and the methods applied do not permit a systematic test of the main implications of the general theory of deterrence, a theory which posits that potential offenders respond to incentives. The shortcoming is basic because the implications following from the general deterrence hypothesis are what Sellin was attempting to challenge empirically. Yet his work neither develops nor tests the full range of implications following from the theory that he attempts to reject; nor is a competing theory developed and tested. In addition to implying that punishments in general, and executions in particular, may deter crime, the general deterrence hypothesis provides the testable expectation that punishments imposed only with relative infrequency will have less impact than those imposed with a high degree of certainty. The theory of deterrence predicts not only the direction of effects of, for example, equal percentage changes in the relevant apprehension, conviction, and execution risks, but also the relative order of magnitude of these effects. Although the basic premise of these and other testable implications is that, on balance, potential offenders respond to incentives, Sellin never devised an analytical framework for rejecting this general hypothesis. More significantly, to my knowledge Sellin never reported any parametric or nonparametric statistical tests that could justify his rather strong conclusions. In all, there exists a considerable gap between the limitedly useful evidence provided by Sellin and others following his methodology, and the rather emphatic inferences drawn.

Two examples may illustrate this conclusion. Sellin has compared homicide rates in “abolitionist” and “retentionist” states, as defined by the legal status of the death penalty. But Sellin never accounts for the extent of the actual enforcement of the penalty in retentionist states. Yet by his own analysis, the suggested relevant variable is the risk of execution, not merely its legal status, for “were [the death penalty] present in the law alone it would be completely robbed of its threat.” The point is far from being subtle. In a number of “retentionist” states whose homicide rates are compared to neighboring abolitionist states, the execution risk was negligible throughout the

37. Baldus & Cole, supra note 2, passim; Bowers & Pierce, Deterrence, supra note 21, at 1 n.4, 2 n.6, 19 n.24.
entire period investigated by Sellin. For example, in Vermont, New Hampshire, South Dakota, and Nebraska, the death penalty was imposed only rarely after 1920; Massachusetts had no executions after 1947. Not surprisingly, with no allowance for actual execution risks, Sellin’s simple graphical comparisons of homicide rates across different states do not appear to uncover considerable differences. But could such observations justify his inference: “The conclusion is inevitable that the presence of the death penalty—in law or practice—does not influence homicide death rates?”

This deficiency is even more glaring in tests attempted by Bowers in another work to determine the effect of the moratorium on executions in the United States. Following Sellin’s methodology, Bowers compares the levels of murder rates in nine arbitrarily chosen mixes of neighboring states in the four years preceding and subsequent to the judicial moratorium. He reports similar patterns in most groups. However, the plain fact is that none of the states in eight of the nine groups had a single execution throughout the period. And in the ninth group, Bowers creates a dubious distinction between New York, classified as abolitionist, and New Jersey and Pennsylvania, classified as retentionist, although New York ceased all executions in 1963—the same year as New Jersey and one year after Pennsylvania. That such comparisons are used as a basis for inference about the deterrent effect of capital punishment taxes one’s imagination.

A second general shortcoming characterizing previous research on this issue by Sellin and others is the absence of any systematic standardization of data so that the effect of execution risk can be isolated from the effects of other factors that not only may influence the murder rate but also are expected to be systematically related to the risk of execution. Clearly aware of the general problem, Sellin has emphasized the need to compare states that are “as alike as possible.” However, his assumption that neighboring states satisfy this prerequisite is unacceptable. Pairs of neighboring abolitionist and retentionist states, such as Illinois and Wisconsin, Michigan and Indiana, or Massachusetts and Rhode Island, differ in their economic and demographic characteristics, in their crime rates and law enforcement activity, and presumably also in their medical services available to
victims of aggravated assaults. In addition, as my analysis has shown, variation in the legal status of the death penalty occasionally may be a result, rather than a cause, of changes in murder rates and thus may give rise to an apparent positive association between the two variables. For these reasons, the true effect of the death penalty on the murder rate cannot readily be inferred from simple comparisons of the sort performed by Sellin and others.

VII. The Proper Measurement of the Restraining Effect of Capital Punishment

My critics allude to my having inappropriately or incompletely estimated the deterrent effect of capital punishment because I neglected the interdependencies among apprehension, conviction, and execution risks. The fact is that I have stressed and integrated these interdependencies both in my theoretical analysis and in the empirical investigation. Both pairs of authors suggest that the only proper means of estimating the deterrent effect of the death penalty is by allowing apprehension and conviction risks to "vary" with execution risk rather than by holding them "constant," thereby estimating the "total" effect of the penalty. In fact, I have treated empirically both the "partial" effect of execution risk—by controlling for the apprehension and conviction risks—and the observed "total" effect—by estimating the effect of execution risk through a reduced form regression analysis controlling only for exogenous and predetermined variables. The former procedure is without question the only proper way of verifying

44. The perfunctory analysis of Baldus and Cole, supra note 2, at 177 n.30, on this point shows little appreciation of the issues involved. At Table III their list of data purportedly for one year is a mix of data drawn from two different years (1960 and 1966). Yet the issue on which they are leveling evidence is the extent to which Sellin "controlled for other factors" in each and every year of the 40-odd years considered in his studies. There still exists substantial variation in the variables Baldus and Cole compare within the five groupings of states they consider. They have not presented data for the sixth grouping considered by Sellin. The data sources they cite for the construction of their permanent income measures for the states are not capable of generating such statewide measures. Their inferences about the impact of specific variables on the murder rate in each state are not based on any systematic analysis. In three of their five cases, a retentionist state has the lowest homicide rate within a particular grouping—a fact they choose not to mention. By Baldus and Cole's analysis this fact might be viewed as evidence for deterrence. I am not suggesting, however, that such inference can be justified given the fragmentary nature of the analysis. More generally, I question the logical basis for preferring a statistical method that provides only indirect and incomplete control for specific variables expected to have an impact on the crime rate over a regression analysis attempting to control these variables directly.


46. Bowers & Pierce, Deterrence, supra note 21, at 6 n.9; Baldus & Cole, supra note 2, at 177-83 & n.42.

the hypothesized deterrent effect of execution risk, for the theory predicts that an increase in the conditional risk of execution at given levels of apprehension and conviction risks would have a deterrent effect on the incentive to commit murder. For the quite distinct purpose of evaluating the overall desirability of capital punishment as an instrument of policy, both estimation procedures, in principle, can provide useful guidance. Indeed, both are reported in my study. Although the “total” effects associated with a few measures of execution risk appear somewhat smaller than their estimated “partial” effects, the quantitative differences are generally small.

It should be pointed out, in this connection, that my critics misconstrue the theoretical predictions relating to the “total” deterrent impact of the risk of execution. While unwarranted movements in execution risk are expected to induce opposite movements in apprehension and conviction risks, the magnitudes of the resulting effects will not necessarily offset the effect of the initial change in execution risk. Moreover, the theory implies that no such compensatory movements are expected when the initial shift in execution is viewed as warranted. For example, if there were universal agreement that reinstatement of capital punishment under specific conditions were socially optimal, then there is no compelling reason to expect that juries would be less inclined to convict offenders charged with capital crimes. Levels of apprehension and conviction risks could also be maintained through an appropriate allocation of resources to specific law enforcement activities. Thus estimates of “partial” effects of execution risk, as well as of apprehension and conviction risks, provide useful information from a policy viewpoint.

Concluding Remarks

The basic issue underlying my theoretical and empirical investigation of murder has not been merely the deterrent efficacy of the death penalty but the more general issue of offenders’ responsiveness to incentives. My time series analysis of the trend of murder in the United

---

48. In reference to these possible compensatory changes in apprehension and conviction risks, both Bowers & Pierce, Deterrence, supra note 21, at 6 n.9, and Baldus & Cole, supra note 2, at 182 n.42, cite an exercise in an unpublished paper by P. Passell & J. Taylor, The Deterrent Effect of Capital Punishment: Another View, Feb. 1975 (Discussion Paper No. 74-7509, Columbia Univ. Dept. of Economics) (on file with Yale Law Journal), as evidence that the “total” effect of execution risk on the murder rate is likely to be positive. Ironically, neither pair of authors apparently understands the exercise by Passell and Taylor which has nothing to do with the “compensatory” changes discussed above. In fact, I believe that the exercise by Passell and Taylor is internally inconsistent and irrelevant, but since their paper has not been published yet I will not address their exercise here.
States should not be evaluated as an isolated experiment but rather as part of the more general research into offenders' behavior. Viewed in this more general context, the new research by myself and other social scientists in recent years has lent considerable support to the proposition that, in the aggregate, potential offenders respond to both negative and positive incentives. The growing body of new research by economists and sociologists has indicated the existence of deterrent effects of severity and certainty of punishment as well as other systematic regularities attributable to the effects of incentives. These conclusions have been demonstrated in studies using data from different times and different places. My research on the deterrence effect of capital punishment has produced results compatible with this new evidence.

The bulk of my critics' analyses challenges neither the theoretical formulation of the deterrence hypothesis nor the statistical techniques used in the theory's empirical implementation. Neither Bowers and Pierce nor Baldus and Cole present valid tests that reject the deterrence hypothesis. I hope that their work does not baffle the lawyer pondering the merits of using an economic approach to law or of using statistical techniques to study legal questions. The statistical methodology is quite useful and relevant when appropriately applied. Contrary to the inferences of Bowers and Pierce, their study does not present statistically meaningful evidence that the risk of execution has no deterrent effect, let alone that it has a "brutalizing" effect, on the frequency of murder in the population.

In this reply, I have elaborated upon some aspects of my research which, because of space limitations, have not been fully discussed in my published paper on the death penalty. Needless to say, my discussion here is not a substitute for the published paper. The paper discusses the analytical framework underlying the general deterrence hypothesis, which is the main issue of concern in my research. It also


50. Baldus and Cole do not present evidence of their own. They rely mainly upon Sellin's work as evidence, which I discuss briefly at pp. 221-24 supra. Their discussion at Section IV merely repeats other authors' inferences and judgments concerning the deterrence hypothesis which may prove to be unjustified or premature. It is beyond the scope of this discussion to deal with these other works.
elaborates upon the methodology needed for meaningful tests of that hypothesis. The paper also stresses the limitations of the empirical investigation and the tentative nature of the findings. Indeed, as I stated, conclusions based on studies of historical data necessarily are qualified in view of the difficulties of measuring efficient empirical counterparts of relevant theoretical constructs and in view of intrinsic limitations of statistical inference. However, data imperfections are likely to work against, not in favor of, the theorized deterrent effect of punishment and the effects of incentives in general. It is thus remarkable that the evidence of my time series analysis and additional research repeatedly has proven not inconsistent with rather sharp predictions emanating from the deterrence hypothesis. I have not claimed, however, that my research settles the issue of the deterrent effect of capital punishment. Nor have I advocated the use of capital punishment. As I stressed in my paper, the issue of deterrence is but one of a myriad of issues relating to the efficiency and desirability of capital punishment as a social instrument for combatting crime. The study of the deterrent effect of capital punishment is of considerable independent importance in connection with the hypothesis that potential offenders on the whole respond to incentives. Research on this issue undoubtedly will benefit in the long run from legitimate attempts to use more efficient data and statistical techniques than those heretofore employed in studies of capital punishment.