Economic Models of Criminal Behavior,
J.M. Heineke (ed.)

CHAPTER TWO

ECONOMETRIC MODELS OF CRIMINAL BEHAVIOR: A REVIEW

John B. Taylor

Although published studies on the econometric approach to criminal behavior began to appear only 5 years ago, the field has expanded at such a rapid rate that, at least in quantitative terms it already appears to have taken a place beside such old stalwarts of applied econometrics as production and investment behavior. Most of these recent studies[1] represent a methodological advance over earlier empirical research on criminal behavior, in that they use formal multivariate and multiple equation econometric estimation techniques to isolate the deterrent effects and to sort out the two-way interaction between criminal behavior and the criminal justice system. Whether this methodological advance has improved our practical knowledge of the determinants of crime is as yet an open question, however, for a number of critics[2] have questioned the econometric methodology and the findings of these studies.

This paper reviews these recent developments in the econometrics of criminal behavior. Its intention is to outline a consensus of empirical results which have emerged from the new methodology, and to evaluate their statistical reliability from the viewpoint of policy analysis and future empirical work. While the review considers a number of general methodological questions, its emphasis is on empirical findings. Rather than summarizing all empirical work in the area, it focuses on several papers which are representative of different types of data and statistical technique.

Section I reviews the theory of criminal behavior which has served as a framework for much of the empirical work in this field. Section II considers cross-sectional studies which have used aggregate data from different geographical areas. Section III then examines a number of econometric issues specific to these models. Finally Section IV discusses the special empirical problems which have arisen in attempts to measure the deterrent effect of capital punishment.

I. Theoretical Background

Becker's (1968) utilitarian model of criminal behavior, along with a number of modifications and extensions[3] has served as the "microfoundation" for most of the econometric work on aggregative crime models. While the utilitarian appraoch to criminal behavior can be traced back at least to Bentham, the economic models introduced by Becker improve on earlier work by making explicit and emphasizing the inherent uncertainty associated with decisions to engage in criminal activity. Applying the expected utility approach to criminal decision making, Becker is able to derive a number of empirically testable propositions about criminal behavior.

To take the simplest case, suppose that criminals maximize the expected value of a utility function which depends on the net income derived from criminal activity across two uncertain states: getting caught or getting away with the crime. If Y is income (monetary as well as psychic) and F is the penalty if caught, then the value of the utility function U--an increasing function of net income--will be $U(Y-F)$ if the criminal gets caught and $U(Y)$ if not. If p is the likelihood of getting caught as perceived by the criminal, then $1-p$ is the likelihood of getting away, and the expected utility of engaging in criminal activity is simply

$$(1) \quad EU = (1-p)\ U(Y) + pU(Y-F).$$

If this expected utility is high enough relative to the expected utility from legal pursuits, then it is assumed that criminal behavior will be undertaken;[4] consequently, any change which increase EU will raise crime rates.

A number of the propositions implied by this formal analysis are intuitive--the skeptic might say obvious. For example, equation (1) shows that increasing the likelihood (p) and/or severity (F) of punishment will lower the expected utility of criminal activity and hence, according to the theory, lower crime rates.[5] While an empirical analysis based on such a theory must consequently incorporate and test for this derived deterrent effect, even the most theory-deprived elementary empirical study would entertain the possibility that punishment has an effect on crime. But the utilitarian model also leads in more substantive directions, and these have played an important role in guiding the recent empirical work. There are at least three such directions which are important for the empirical work discussed below.

First, a distinguishing characteristic of the utilitarian approach is the supposition that criminal behavior can be described in much the same way as conventional economic behavior, without particular reference to psychological theories. The hypothesis is that criminals act *as if* they are responding rationally to incentives and deterrents presented by their socioeconomic environment, including the criminal justice system. According to this theory, an empirical model of criminal behavior can be specified simply by obtaining suitable empirical counterparts for the relevant socioeconomic and deterrent variables--income distribution, age, urbanization, arrest rates, length of imprisonment, etc. The intermediate and frequently unobservable psychological process, which describes how these environmental inputs affect criminal behavior, can be by-passed, enabling one to specify the form of the empirical model directly. The statistical implication is that regression and simultaneous equation techniques, based on observable socioeconomic variables, can be used rather than statistical methods common to psychometrics, such as factor analysis, which emphasize unobservables. The recent proliferation of econometric methods in the area of criminal behavior, therefore is not a coincidence, but is a consequence of the micro-theoretic foundation used. This is not to say, of course, that econometric techniques are necessarily the preferred method. If there were doubt about the ability of utilitarian models to describe criminal behavior, and thereby place restrictions on the empirical specification, factor analytic methods--which tend to be more agnostic or symmetric in their use of restrictions--would be the preferred methodology. Indeed, some recent researchers in one of the traditional fields of economic analysis--business cycles--have questioned the conventional use of a priori restrictions and have experimented with factor analytic techniques as an alternative to the standard econometric methodology.[6]

A second important emphasis of the economic theory of crime, is the two-way interaction between criminals and the criminal justice system. Equation (1) is only part of the story. Criminals are not the only expected utility maximizers; society, in its operation of the criminal justice system, also maximizes utility by weighing the costs of operating the system against the costs to society of criminal behavior. The implication of this part of the theory is that the criminal justice system will respond to changes in criminal activity by adjusting expenditures on police, courts, or prisons. As a result, arrest rates, conviction rates, and severity of punishment will tend to depend on crime rates. Moreover, because the justice system itself is composed of utility maximizing individuals, there may be an interrelation between arrest rates, conviction rates and severity of punishment; for example, conviction rates will depend on severity of punishment if anticipation of potential sentences influences the verdicts of juries. Viewed in conjunction with the behavior of criminals, therefore, the models suggest that crime and punishment are jointly determined. Neither can be said to depend on the other, but together they depend on other variables which are exogenous to both.

This joint determination of the two sides of the law has fairly clear statistical implications. Care must be taken to insure that the equation of interest--say the crime rate equation--is identified by the a priori restrictions of the theory. In particular, the theory must insure that some exogenous factors influence the criminal justice system, but do not influence crime. Otherwise we could not hope to disentangle the effect of changes in the criminal justice system on crime, from the effect of changes in crime rates on the criminal justice system. And given that the equation is identified simultaneous estimation techniques which, at the least, generate consistent parameter estimates should be used rather than inconsistent regression methods such as ordinary least squares.

Finally, among the propositions of the microtheory which have had an important role in the empirical work is a prediction about the relative magnitude of the effect of certainty versus severity of punishment in deterring crime. A certain punishment has long been thought by criminal justice experts to deter criminals more effectively than a severe punishment. Becker's expected utility analysis interprets this conventional wisdom in terms of the potential criminal's taste for risk. If criminals are not risk averse then, according to the theory, a proportional increase in the severity of punishment cannot be more effective in deterring a criminal act than a proportional increase in the likelihood of punishment. This result has stimulated empirical researchers to test for the relative effects of the two deterrents, and consequently almost all empirical studies have included measures of certainty and severity. Moreover, these studies have used a functional form--linear in the logarithms--suitable for comparing proportional impacts (elasticities). As will be discussed later such a specification has had substantive effects on the empirical findings and, although the proposition has been questioned, it is worth considering its derivation here.

Using the notation introduced earlier, the statement that the likelihood of punishment is proportionally more effective than severity can be represented in terms of derivatives as

$$(2)\quad \left|\frac{d(EU)}{U}\Big/\frac{dp}{p}\right| > \left|\frac{d(EU)}{U}\Big/\frac{dF}{F}\right| .$$

The left hand side of this inequality is $[U(Y-F) - U(Y)]\,p/U$ and the right hand side is $pU'(Y-F)\,F/U$, as can be seen by differentiating equation (1). Hence, inequality (2) will hold if

$$(3)\quad [U(Y) - U(Y-F)]/F > U'(Y-F).$$

But this is simply the condition that the utility function be convex, or equivalently that the potential criminal prefers risk. Risk preference therefore implies that certainty of punishment is more effective than severity in lowering crime rates.

To summarize the discussion of this section, the microtheoretic framework appears to have had three important influences on the empirical work on crime: (1) it has enabled empirical researchers to place observable socio-economic and deterrent measures directly into crime rate equations with restrictions on the parameters, and thereby avoid statistical models developed for unobservable variables; (2) it has underlined the need for simultaneous equations methods in order to control for the effect of crime on the criminal justice system, and (3) through its emphasis on relative effectiveness of certainty and severity, it has motivated functional forms which include measures of both deterrents and which are linear in the logarithms.

Assuming that summing data on individual criminal behavior into geographic crime rates causes no aggregation problems, the microtheory therefore suggests that aggregative statistical models of crime should have the following form:[7]

$$(4) \quad c = f_1(p,s,x,e_1)$$

$$(5) \quad p = f_2(c,s,x,e_2)$$

$$(6) \quad s = f_3(p,c,x,e_3)$$

where c,p, and s are the crime rate, perceived probability of punishment, and severity of punishment, respectively, in a geographic area during a given time period. The vector x consists of all relevant explanatory variables, and the e_i are the random disturbances in the model. In principle the theory generates restrictions on which variables enter the functions f_1, f_2, f_3 in order that some or all three equations can be identified. For reasons mentioned above, the functions are usually specified in log-linear form. The following three sections consider various approaches to estimating one or more of the equations in such a model.

II. Studies Based on Aggregative Cross-Section Data

This section reviews six representative empirical studies which have attempted to estimate crime rate relationships of the form of equation (4) using aggregative cross-section data. As can be seen from the summary in Table 1, the papers reviewed represent a fairly wide geographic range of data: states and cities in the U.S., provinces in Canada, police districts in the U.K., counties in California, and police precincts in New York City. The degree of aggregation varies substantially across the different studies. Some examine only the total felony crime rate, while others disaggregate by type of crime. Almost all of the studies utilize some type of simultaneous equation technique (usually two stage least squares), but report ordinary least squares estimates as well.

A. Ehrlich's study of states in the U.S.

Ehrlich (1973) reports crime rate equations in the form of equation (4) based on aggregative state data for three census years 1940, 1950, 1960. The dependent variable in his analysis is the number of reported crimes (C) of a particular type divided by the population of the state (N) in the same year. Seven different types of crime are examined: murder, rape, assault, robbery, burglary, larceny and auto theft. Two deterrent variables are included in the structural equations for each category of crime: the number of imprisonments (I) divided by the number of reported crimes (C), and the average time (T) served by offenders in state prison for that type of crime. In addition, three socioeconomic variables, which do not vary by type of crime, are included in the most frequently reported equations. The median income of families in the

TABLE 1

Summary of Crime Rate Equations based on Cross-Section Data[1]

Authors	Data	Time Period[2]	Level of Crime Aggregation	Deterrent Variables	Estimation Method[3]
Avio and Clark	Provinces in Canada	1970, 71, 72	Robbery, Burglary, Theft, Fraud	Arrest Rate,[4] Conviction Rate, Length of Prison Term	TSLS OLS
Carr-Hill and Stern	Police districts in England and Wales	1961, 62	Felonies	Arrest Rate[4]	FIML
Ehrlich	States in U.S.	1940, 50 60	Murder, Assault, Rape, Robbery, Burglary, Larceny, Auto Theft	Imprisonment, Length of Prison Term	TSLS OLS
Mathieson and Passell	Police precincts in New York City	1970	Murder, Robbery	Arrest Rate[4]	TSLS OLS
Phillips and Votey	Counties in California	1966	Felonies	Conviction Rate Prison-Probation Rate	TSLS OLS
Sjoquist	Cities in U.S.	1963	Aggregate of all Burglaries, Robberies and Larcenies over $50	Arrest Rate,[4] Conviction Rate, Length of Prison Term	OLS

1. These studies are discussed in Section II of the paper.
2. The Avio and Clark study is based entirely on data pooled over the three time periods; the other studies consider separate regressions for each time period.
3. In some of the studies the OLS estimates are based on transformed data, thereby approximating some type of weighted least squares.
4. Arrest rate refers to the reported clearance rate.

state (W), the percent of families below ½ of the median income (X), and the percent of nonwhite in the population (NW).

The theoretical models of crime suggest that, holding other effects constant, the two deterrent variables should have a negative effect on crime, while the three socioeconomic variables should have a positive effect. The income variables W and X are measures of the relative gain to criminal activity: in richer states (higher W), potential criminals can expect more loot and in states with a more skewed income distribution (higher X), that loot will look relatively more attractive to a larger fraction of the population. Ehrlich includes the variable NW for similar reasons: legitimate employment opportunities for the nonwhite population may be deficient.

Other explanatory variables are also considered in Ehrlich's analysis, but his basic empirical results focus on these five. All his results are based on a log-linear functional form specification which leads to the following statistical crime rate equation:

$$(7) \quad \ln\left(\frac{C}{N}\right) = b_o + b_1 \ln\left(\frac{I}{C}\right) + b_2 \ln T + b_3 \ln W + b_4 \ln X + b_5 \ln NW + e.$$

Equation (7) was fitted over most of the states for all crime categories in 1960, and for a more limited selection in 1940 and 1950.

The results for the deterrent variables are reported in Table 2 and are in fairly close accord with the forecasts of the theoretical model. In particular b_1 and b_2 are generally negative while b_3, b_4 and b_5 are generally positive and significant. The economic variables X and W tend to have greater effects on crime against property than on crimes against persons, as one would expect. In addition, the deterrent variables appear to be more effective in reducing crimes against persons than against property. Moreover, for all disaggregated crimes except burglary and sometimes larceny the impact of a more certain punishment (I/C) tends to have a greater impact on crime than a more severe punishment (T)--a finding which is in agreement with the views of most criminal experts, as well as with the theoretical model if potential criminals prefer risk. However, as Ehrlich points outs, the average length of prison term may significantly overstate the expected severity of punishment for a prospective criminal with a high discount rate. If so, the estimated impact of severity on crime rates would be understated. Note also that the differential impact is not evident for the aggregate equation.

The qualitative nature of these results does not appear to depend on the statistical estimation technique. In the 1960 cross section, both ordinary least squares (OLS) and two stage least squares (TSLS) estimation[8] techniques are reported. The direction of impact and statistical significance is generally the same for both estimation techniques. The quantitative effect is quite different, however, with the simultaneous estimation approach (TSLS) yielding estimated elasticities for the deterrent variables which are about twice as large as with OLS. The implicit simultaneous equation model which calls for techniques like TSLS is similar to equations (4) and (5) of Section I. The crime rate (C/N) and the estimated likelihood of punishment (I/C) are assumed to be simultaneously determined for reasons already mentioned. The severity of punishment, however, is not assumed to be jointly determined with the other variables, and consequently an equation like (6) is not considered.

The predetermined variables which are assumed to be excluded from the crime rate equation but not from the probability of punishment equation, and which consequently can be used for identification and for construction of the TSLS estimates are: population, an unemployment variable, the percent of the population of age 14 to 24, the percent of the population living in SMSA's, the ratio of males to females, average education, per capita expenditure on police, and dummy variable for southern states. In addition to these exogenous variable, the crime rate in 1959 was assumed to be a predetermined variable for estimation of the 1960 cross section. There does not appear to be much rationale assuming that these variables do not directly affect crime rates, but do directly affect the criminal justice system. In fact, the classification between included and excluded exogenous variables differs from other work of Ehrlich[9] and from that of others. Moreover, the use of a one period lag of a dependent variable as a predetermined variable in a cross-section study is questionable when one expects serial correlation to be significant relative to the cross-section variation. These problems have led to criticism of Ehrlich's use of simultaneous equation methodology;[10] if variables assumed to be excluded from the equation are not actually excluded then the equation might not be identified and the TSLS estimate would be meaningless. This issue is discussed in more detail in Section III.

The OLS and TSLS estimated elasticities for the two deterrent variables are shown in Table 2 for the various crime types in the 1960 cross section. For all crimes the estimated elasticity is about $-\frac{1}{2}$ when OLS is used and about -1 when TSLS is used. If the criminal justice system reacts to higher crime rates by increasing the likelihood of punishment, then the higher elasticity for TSLS is in accord with expectations. The simultaneous procedure isolates the negative deterrent effect from the positive criminal justice effect, and consequently is able to uncover a larger deterrent effect. This effect of TSLS is evident for all types of crime.

For analysis of the other cross-section studies it is important to mention a statistical difficulty which is evident in equation (7). The

Table 2

Estimated Impact of Deterrents on Crime[1]

	TSLS		OLS	
Ehrlich(U.S. - states)[2]	Certainty	Severity	Certainty	Severity
Murder	-.852	-.087*	-.341	-.140*
Rape	-.896	-.399	-.578	-.188*
Assault	-.724	-.979	-.275	-.180*
Robbery	-1.303	-.372*	-.853	-.223*
Burglary	-.724	-1.127	-.534	-.900
Larceny	-.371	-.602*	-.133	-.263
Auto Theft	-.407	-.246*	-.247	-.174*
All Crimes	-.991	-1.123	-.526	-.585
Sjoquist (U.S. - cities)[3]				
Robbery, Burglary and Larceny over $50			-.354	-.292
Carr-Hill and Stern (UK)[4]				
All Offences	-.66	-.28		
Avio and Clark (Canada)[5]	Arrest/Conv.		Arrest/Conv.	
Robbery	-.83/-2.8	.04	-.76/-2.9	.48
Break and Enter	-1.52/.67*	.08*	-1.3/.16*	.21
Fraud	-.71/-.59*	-.01*	1.71/-.14*	-.03
Theft B	-.79/-1.6	.24	-.67/-1.7	.08
Theft A	-.74/-1.8	-.50*	-.67/-1.9	.04
Phillips and Votey (Calif.)[6]				
Felonies	-.61	-.34	-.62	-.35
Mathieson and Passell(NYC)[7]				
Murder	-1.96		-.74	
Robbery	-2.95		-1.06	

1. Numbers refer to the estimated coefficient of the logarithm of variables representing certainty and severity of punishment in an equation for logarithm of the crime rate. Certainty is generally measured by arrests, convictions or imprisonment per crime. Sjoquists' study indicates that these measures give similar results (see Table 1, equation (1) and (2) of that study). Severity is measured by the average time served in prison. The numbers marked with an asterisk represent variables with t-ratio less than 2.

2. Ehrlich (1973), Tables 2,3,4, and 5.
3. Sjoquist (1973), Table 1, Equation 2.
4. Carr-Hill and Stern (1973), p. 303; the reported estimates are based on FIML.

5. Avio and Clark (1976), Tables 7 and D1. Reported results give the coefficient of the arrest rate and the conviction rate (given arrest) in the same equation. Both are reported here.

6. Phillips and Votey (1975), Table 1.

7. Mathieson and Passell (1976), Table 2 and 3. This study does not include a severity measure. Hence the estimated elasticity of the arrest rate is likely to be biased upward in absolute value.

number of crimes C in equation (7) is a variable which is generally measured with considerable error. This problem, which plagues all crime studies and which is acknowledged by researchers and critics alike, may tend to cause a negative spurious correlation between C/N and I/C. In principle it is impossible to distinguish this spurious correlation from the deterrent effects without additional information. Hence, the estimated elasticities in equation (7) may overstate the true deterrent effect. The TSLS estimates do not circumvent this pitfall, unless the measurement error is eliminated in the reduced form estimates. Since Ehrlich's reduced form (first stage) contains $(C/N)_{-1}$ on the right hand side of the (I/C) equation, the measurement error would appear to persist in the TSLS estimates. In order to make practical use of the estimated elasticities it is necessary, therefore, to assess the quantitative significance of this measurement error. Note that the other deterrent variable in Ehrlich's equation - average time in prison - does not have the same type of measurement error.[11] This problem is discussed in more detail in Section III.

B. Sjoquist's study of cities and towns in the U.S.

Sjoquist reports crime rate equations of the form of equation (4) for a 1968 cross section of 53 municipalities in the U.S. with 1960 populations between 25,000 and 200,000. In order to avoid spillover problems caused by the tendency for criminals to shift their activity to neighboring cities when their own law enforcement agencies crack down, only relatively isolated cities were examined. Municipalities with large neighboring municipalities were excluded from the sample.

Only one dependent variable was examined: the number of reported robberies, burglaries, and larcenies over $50 divided by the population. The deterrent variables included a severity variable - average prison sentence served - and three alternative variables representing the likelihood of punishment - arrests per crime, convictions per crime, and convictions per arrest. The third probability measure is a conditional likelihood and would therefore be expected to have a smaller effect on crime. The six socioeconomic variables are population, population density, retail sales, income, education, and the percent of the population which is nonwhite. A log-linear functional form was chosen for estimation. Hence, with the exception of the education variable the specification is very similar to that reported by Ehrlich. Note that Ehrlich excluded education from the crime equation for identification.

Sjoquist only reports OLS estimates for his model and these are reported for the deterrent variable in Table 2. The results are in agreement with the theory, with both deterrent variables having a negative effect. The estimated elasticity of the conviction rate is about -1/3 and is somewhat less than the average of Ehrlich's estimates, but is not appreciably different, considering that the focus is cities and towns rather than states. The elasticity of severity is slightly less than Ehrlich's OLS estimates.

The likelihood of punishment variable reported in Table 2 is the number of convictions per crime; similar results are reported for arrests per crime, but the convictions per arrest variable in insignificant. As Sjoquist points out, this may be due to the absence of the spurious correlation problem for this variable. The variable may also have less impact because it represents a probability conditional on an uncertain event (arrest). Alternatively, multicollinearity may be a problem.

C. Carr-Hill and Stern's study of police districts in the U.K.

Of the studies reviewed here, this is the only one which specifies and estimates a complete model using full information simultaneous equation techniques. Carr-Hill and Stern's data set pertains to urban and rural police districts in 1961 and 1966 in England and Wales. Their statistical tests suggest a significant structural break in the model between these two years, consequently the two samples are considered separately in the main analysis. Focusing on the 1961 results for urban areas, the crime rate, defined as the total number of reported offenses per capita (y) is assumed to depend on two other endogenous variables: the arrest rate (p) and the number of police per capita (c). In addition, two exogenous variables are assumed to influence crime: the proportion (a) of the population between 15 and 24, the value of real estate (t), and the severity of punishment (f). The latter is measured by the proportion of convicted criminals given custodial treatment.[12] No attempt is made to explain the severity of punishment within the context of the model, so that equation (6) of Section I is implicitly assumed to depend only on the exogenous variables.

The model is completed by specifying equations for the arrest rate and the number of police. In their 1961 model Carr-Hill and Stern assume that the arrest rate equation included police expenditures per capita, total population (n), the working class proportion of the population (s), the proportion of violent crimes (v), and the age variable (a). Finally, the number of police per capita (c) is assumed to depend on the arrest rate (p), the middle class fraction of the population (m), and the proportion of violent offenses (v). The specification for 1966 is similar but includes additional explanatory variables. (The notation corresponds to that of Carr-Hill and Stern, not to Section I above).

As with other research on crimes, all equations are specified in log-linear form. Using the above notation to represent the log of the variables, leads to the following model representation:

(8) $y = \alpha_1 p + \alpha_2 f + \alpha_3 a + \alpha_4 t + \alpha_0 + u_1$

(9) $p = \beta_1 c + \beta_2 a + \beta_3 s + \beta_4 n + \beta_5 v + \beta_0 + u_2$

(10) $c = \gamma_1 p + \gamma_2 m + \gamma_3 v + \gamma_0 + u_3$

The parameters of the above model were estimated using the Full Information Maximum Likelihood (FIML) technique, which gives more efficient estimations than two stage least squares because it incorporates the interaction between the equations. This efficiency advantage, of course, requires that all the equations are specified correctly. If not, then the estimates may not perform as well as two stage least squares or even ordinary least squares. The use of a full information method has an important practical advantage in any case, however, even if one is interested only in the crime rate equation. Such estimates require the investigators to state explicitly the equations which contain variables excluded from the crime rate equation. Hence, the identifiability conditions on the first equation can be evaluated more reliably.

The estimated values of the deterrent variables which result from the FIML techniques are reported in Table 2. The differential between the arrest rate elasticity and the severity elasticity is considerably greater than the estimated values reported by Ehrlich and Sjoquist, but this may be due to the different measures of severity used. Again, the sign of the differential indicates risk preference. The deterrent elasticity of the arrest rate is -2/3, which is between the TSLS estimate of Ehrlich and OLS estimates of Sojquist and Ehrlich. The estimated deterrent effect of severity is less than that of the other researchers, but is negative and significant. The "swag" variable t, has a positive effect as does the income variable used in the U.S. studies. Despite some differences, however, these results are generally consistent with the results for the U. S. and appear to lend further support for the theoretical model.

Perhaps the most provocative finding of Carr-Hill and Stern is that the number of police per capita has significantly negative effect on arrest rates ($\beta_1 < 0$). If the arrest rate is viewed as the output of the crime prevention production function, then this finding would seem to imply negative returns to additional law enforcement personnel. While negative returns are plausible over some range of production, it is difficult to believe that police districts would operate at such a point. Carr-Hill and Stern offer two intriguing explanations for this disturbing finding: first, they argue that with more police around, more crimes will tend to be reported that are hard to solve (presumably these would be minor crimes which the police would have less interest in solving as well); second, a larger police force will have a large deterrent effect on crimes that are easier to solve, thereby raising the number of crimes with inherently low arrest rates. The combination of these two effects is to alter the crime mix by increasing the proportion of crimes which are hard to solve relative to those that are easier to solve. Hence, if the magnitudes of the effects are large enough, the number of police will be negatively related to the arrest rate.

In outlining the policy implications of their empirical model, Carr-Hill and Stern focus on the reduced form coefficients, which is the correct approach from a simultaneous equation standpoint. The fact that the arrest rate is negatively correlated with the crime rate in a simultaneous equation model offers no immediate policy implications. These two variables are jointly determined and endogenous; according to the model, therefore, one can only affect the behavior of arrest rates by altering some exogenous variable. Moreover, if one is interested in the crime rate, then it is the relationship between the crime rate and the exogenous variables that is important. The relationships are all captured by the reduced form. Very few of the studies reviewed here utilize the reduced form for policy, despite the fact that the exogenous variables used in the first stage of a complete two stage least squares estimation may be sufficient for estimating an unrestricted reduced form.

D. Avio and Clark's study of provinces of Canada.

One of the most striking geographic characteristics of crime statistics in Canada, is the extremely high crime rates in the Yukon and Northwest Territories compared to the rest of the country. For example in 1966, the total crime rate in the territories was about 10 times greater than the average crime rate in the provinces. Such variation would appear to be invaluable in estimating the determinants of crime. Unfortunately, the study by Avio and Clark is unable to utilize these extreme variations because of insufficient data on socioeconomic variables. Moreover, the province with the second highest crime rate--Alberta--must also be eliminated from the sample. This data loss is especially unfortunate in this study because of the relatively small number of provinces in Canada. The basic model of Avio and Clark is estimated with a sample of only 8 cross-sectional observations corresponding to 8 provinces. Although they do pool this data for three years 1971, 1972, and 1973, the effective information is still small unless there were unusual fluctuations in crime during those three years. Moreover, pooling the cross-section data over three time periods potentially introduces serial correlation problems which cannot be adequately handled with such a small sample.

Despite these data problems Avio and Clark are able to obtain estimates of a crime rate equation for Canada which are similar to those reported above for the U.S. and the U.K. The basic equation is estimated in log-linear form for 5 different types of crime: robbery, break and enter, fraud, and two types of theft. Included in the equation are the ratio of arrest to crime (the clearance rate), the ratio of conviction to arrests, the average length of sentence, the percentage of families with incomes less that ½ of the median income (the same variable used by Ehrlich for income distribution effects), and the number of households with record players. The last variable represents the "victim stock" or the gains from criminal activity. The inclusion of two different deterrent ratios (the arrest rate and conviction rate) was first attempted by Sjoquist in the paper discussed above, and has since been used by other authors.

The estimated deterrent effects are reported in Table 2, and are generally negative and significant with the important exception of the length of sentence which is usually insignificant but sometimes perversely positive and significant. The significant differential between the coefficient on the arrest rate and the conviction rate indicates that the usual aggregation of the two into the ratio of convictions to crimes may be inappropriate (Sjoquist also finds a similar differential). Not reported in Table 2 are the elasticities for the income distribution and the wealth variable which are both positive and significant as the theory would suggest. The elasticity of the income distribution variable is greater than one for most crimes, which is also Ehrlich's finding for the U.S.

The results are very similar for the TSLS and the OLS estimates. However, one must question the advantages of TSLS estimation for such a small sample size. Even with the pooled regressions (where there may be little variability across three adjacent years) there are only 24 observations.

Avio and Clark's discovery that the length of sentence has no deterrent effect and may even increase crime rates is potentially important, for they construct the variable in a way which avoids a potential spurious correlation which may exist in the other studies. The problem is that when crime rates fall and consequently the number of new prisoners also fall, there is a period shortly thereafter during which the average length of sentence for prisoners automatically increases. With the supply of new prisoners down, there are fewer replacements for the recently released short-timers, and the prison population is heavily weighted with long-timers. This raises average sentence lengths. Avio and Clark circumvent the problem by obtaining data directly on sentences handed down by the judicial system with a correction for parole and remissions. Using this alternative data they find that sentence length does not matter. Unfortunately they do not report results using the standard variable on average sentence length. Hence, it is not possible to determine whether the findings of other researchers may be due to this spurious correlation problem.

E. Phillips and Votey's study of counties in California.

The final two studies in this section focus on more limited geographic area within the U.S.--the state of California and the city of New York. Phillips and Votey develop an empirical model for felonies in California in which the crime rate equation is embedded in a model of community operation of the criminal justice system. The three equations of the model include the crime rate, the conviction rate, and the labor cost for operating the criminal justice system.

The model is estimated using the standard log-linear functional form for 50 counties in California using TSLS and OLS. The estimated elasticities for the deterrent variable appear in Table 2 and are negative and significant with the usual differential between certainty and severity. Phillips and Votey use a slightly different measure of severity--the proportion of convicted criminals sentenced to state prison, probation with jail, or simply probation. The result is a much lower elasticity than the average length of sentence--about the same as the Carr-Hill and Stern estimates. Nevertheless the severity effect is negative and significant.

F. Mathieson and Passell's study of precincts in N.Y.C.

This is the smallest geographic area considered in this review. Except for the problem of overflow of crime from one precinct to another, N.Y.C. with its single police force would appear to be an excellent region for the study of criminal deterrents. Mathieson and Passell develop a three equation crime model and obtain estimates for the crime of robbery and murder in 65 precincts in New York City in 1971. The model consists of equations for the crime rate, the arrest rate, and the number of police assigned to a precinct. The crime rate equation includes the arrest rate, median family income, percent of families with income greater than $25,000, median family income of adjacent precincts, and a dummy variable for the business districts. It should be noted that no measure of severity of punishment is included in this model. Indeed it would be difficult to determine whether severity of punishment is systematically different across N.Y.C. where the court system overlaps many precincts. The estimated deterrent effect of the arrest rate is given in Table 2 for both the OLS and TSLS estimates of the crime equation. The estimated elasticities are negative and significant, but are substantially larger than those obtained by the other researchers. The TSLS estimate for the robbery equation yields an elasticity of about -3. One reason for these high estimated impacts might be the omission of a variable to measure the severity of punishment. Another might be the overflow problem: it is easy for criminals to shift their activity to another precinct when the police crack down in their usual district. This mobility would give a higher elasticity than obtained in the other empirical work where geographical distances are greater (recall that Sjoquist discarded cities and towns from his sample if they were adjacent in order to avoid this problem). If this overflow problem is substantial then one must be cautious when applying these elasticities to a region when mobility is more restricted. For example, one would expect a much smaller reduction in crime if arrest rates rose in all of New York City, than if they rose in a single precinct as in the estimated equation.

An interesting feature of the Mathieson and Passell study is the arrest rate equation. Arrest rates are assumed to depend on police manpower per reported crime and a measure of neighborhood stability--the percentage of families who have lived in the same neighborhood for more than 5 years. The neighborhood stability variable appears to have little effect, but police manpower per crime is usually positive and very significant. This finding seems reasonable but is at odds with that of Carr-Hill and Stern who found a negative relation between police and arrests. Is the NYC police system that much different than that of England or Wales? Aside from the different countries and type of data, there is an important difference between the two studies. Mathieson and Passell divide police by reported crimes, while

Carr-Hill and Stern divide by the population of the police district. Since the dependent variable is the ratio of arrests to reported crimes, a potential spurious correlation problem is introduced to the Mathieson and Passell equation which cannot exist for Carr-Hill and Stern. This spurious correlation may be caused by measurement error in crime rates and would generate a positive relation between arrest rates and police per crime. It is certainly possible that such measurement error accounts for the difference between the two studies. One might want to examine the rationale for deflating police expenditures by crimes rather than population, and whether an alternative deflator would alter the Mathieson and Passell results.

The Mathieson and Passell paper is alone among the papers considered here in its application of the estimated elasticities to a cost-benefit analysis of police expenditure increases. However tentative, the estimates such an analysis is instructive and is a check on the plausibility of the estimated results. The cost-benefit analysis is based on a quasi-reduced form (the arrest rate, but not police manpower, is treated as an exogenous variable) which as argued above is the correct approach in a simultaneous equation framework. Their finding is that a one percent increase in police, which would cost $4.2 million in 1971, would save New Yorkers $2.85 million. Further savings or greater actual deterrent effects would be necessary to have the estimated benefit cover the cost.

III. Problems of Econometric Methodology

Like most applied econometric work, a number of recurrent statistical problems are evident in the above studies. And also like most applied econometric work, these problems have been examined closely by both critics and proponents of the econometric approach to empirical work on crime. In this

Section, two econometric problems--measurement error and identification--are selected for further discussion from the longer list of problems that were only touched on in Section II. These two methodological problems have generated considerable debate in the criminal deterrence literature.

A. The Identification Problem.

In the above discussion of Ehrlich's model I noted the lack of a formal theory for excluding some variables from the crime rate equation--age, for example--while including such variables in the other equation in the simultaneous model. The lack of theory is evident in most of the empirical work reviewed above, especially when the models are compared. Some studies include variables that other studies exclude, and visa versa. In some case this inconsistency is evident in different studies by the same author. Since the proper exclusion of predetermined variables is crucial for identifying structural equations, the lack of formal theory has led some critics to the conclusion that the models of crime are not adequately identified and consequently that the statistical results are unreliable: if the crime equation is not identified then the resulting estimate may reflect the behavior of the criminal justice system rather than of criminals. For example, in their examination of the identification problem in models of crime, Franklin Fisher and Daniel Nagin conclude that "... it appears very doubtful that work using aggregate cross-sectional data can ever succeed in identifying and consistently estimating the deterrent effect of punishment on crime."

While one cannot deny the apparent inconsistency in the choice of zero restrictions for identification, a reasonable argument can be made that the crime models are identified when one takes a broader view of the identification problem. For the sake of illustration consider a two observation sample of crime rate (c) arrest rate (a) pairs. Suppose that one observation is taken from a region--the Wild West of the 1880's, say--with very high crime rates and low arrest rates, and that the other observation is taken from a region--the Established East--with very low crime rates and high arrest rates. These hypothetical observations are represented by the points WW and EE in Figure 1 and 2, for two alternative crime rate functions.

In Figure 1 the two observations lie on a negatively sloped crime rate function, at the intersection of this function and two levels of the criminal justice function. The criminal justice function has the same slope but its level in the Wild West is much lower than in the Established East--perhaps because of different tastes or more likely because of the high cost of law enforcement in an area with low population density. (The slope is shown to be positive, but the same argument would hold if it were negative.) As the figure is drawn, the crime rate function is clearly identifiable. Moreover, even if one could not find observable exogenous variables to explain the different levels of the criminal justice system (tastes, for example), the crime equation could still be identified. In fact OLS would give a fairly accurate estimate of the deterrent effect. The important qualitative characteristic of Figure 1 is that the criminal justice function is shifting much more than the crime rate function. The difference in shifts may be entirely due to unobservable factors and may therefore merely reflect that the variance of the disturbance term in the criminal justice function is much greater than the variance of the disturbance term in the crime rate function. In effect, the crime rate function can be approximately identified by a priori restriction on the variance-covariance matrix of the residuals of the strucural equation.[13] The a priori assumption is that all shifts in the crime rate function occur within a very narrow band, while shifts in the criminal justice system are

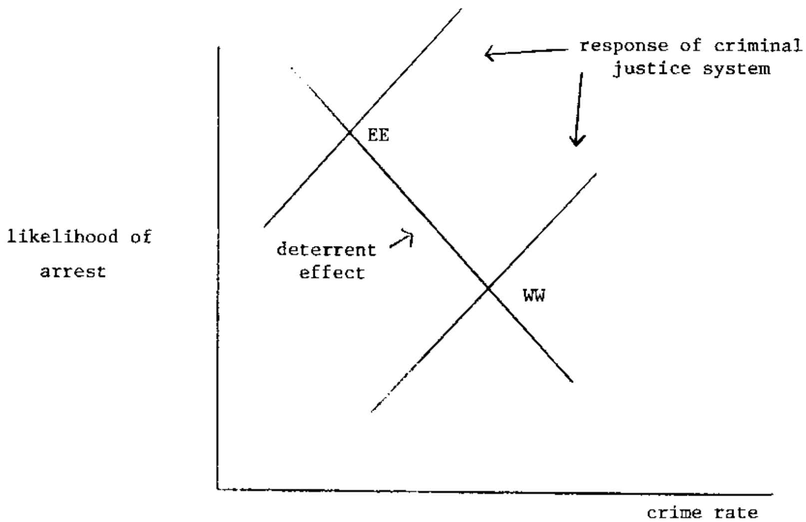

Figure 1

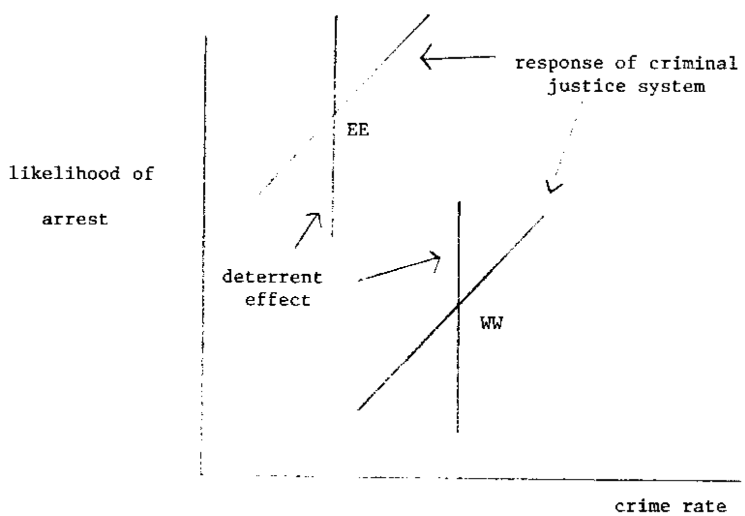

Figure 2

spread out. Drawing such bands around the lines in Figure 1 would illustrate the approximate nature of the identification.

A much different picture is represented in Figure 2, where the same two observations on arrest rates and crime rates are assumed to be caused by shifts in both the crime rate function and the arrest rate function. The crime rate function in Figure 2 has no real deterrent effect, yet an OLS estimate, or a TSLS estimate based on improperly excluded exogenous variables, would generate a negative deterrent. Those who criticize the econometrics of crime for lack of identification, are likely to have Figure 2 in mind. Note that the observations in Figure 2 could be the result of roughly equal variability of the unobservable disturbances in the crime rate function and arrest rate functions. The a priori condition on the relative variability, which could be used for approximate identification in Figure 1 cannot be used in Figure 2.

Whether one can practically identify a crime rate equation using cross-section data, therefore, seems to depend on whether a priori variance conditions of the kind illustrated in Figure 1 are plausible. However, a satisfactory answer to this question would require further research, for example, one might attempt to show that the variance of the unobservable shifts in the criminal justice system are substantially larger than the corresponding variance of the crime rate equation. In the Wild West versus Established East example, the "wide open spaces" would make law enforcement costly, and the high crime rates are likely to be due to the resulting low level of law-enforcement (the migration of illegal activity across geographical regions should be included here). The more recent high crime rates in the Yukon and Northwest Territories of Canada are likely to be the result of similar conditions. In general one might argue that human response to deterrents is more homogenous over cross-sections, than is the criminal justice system. If so, the identification problem associated with models of crime is not nearly as severe as an inspection of zero restrictions would suggest.

B. Measurement Error and Spurious Correlation

A second major econometric difficulty with models of crime is evident in all the studies reviewed above. With the dependent variable defined as the ratio of reported crimes (C) to population (N), and with the major explanatory variable defined as the ratio of arrests (A) to reported crimes, any measurement error in the number of reported crimes will generate a spurious correlation between the crime rate and the arrest rate. The problem is easiest to illustrate in the context of a crime rate equation which depends only on the arrest rate:

$$(11)\quad \ln\left(\frac{C}{N}\right) = \beta \ln\left(\frac{A}{C}\right) + u$$

where u is a random disturbance. Letting lower case letters represent the logarithms of the variables and subscripts to indicate observations, equation (11) becomes

$$(12)\quad c_i - n_i = \beta(a_i - c_i) + u_i$$

If c is measured with error ε, then the OLS estimate of β is given by

$$(13)\quad \hat{\beta} = \frac{\Sigma_i(a_i - c_i - \varepsilon_i)\ (c_i + \varepsilon_i - n_i)}{\Sigma_i(a_i - c_i - \varepsilon_i)^2}$$

which, after substitution of $c_i - n_i$ from (12) can be shown to converge stochastically to

(14) $$\frac{\beta M - \sigma^2}{M + \sigma 2}$$

where M is the average of $(a_i - c_i)^2$ in the limit, and σ^2 is the variance of ε. (The measurement error and the disturbance term u are assumed to be uncorrelated) Except for the special case where the true values of the elasticity β is equal to -1, (14) will not be equal to β when there is measurement error. Hence, $\hat{\beta}$ will not be a consistent estimator of β.

Expression (14) illustrates two things about the impact of measurement error. First, the measurement error will create bias only if it is large relative to the variability of the true arrest rate. Even if measurement error is great (σ^2 is large), substantial variability in the arrest rate (large M) will prevent the measurement error from generating faulty results. Second, the measurement error has two types of impact; a multiplicative effect $M/(M + \sigma^2)$ which is less than one and therefore tends to bias β towards zero, and an additive effect $-\sigma^2/(M + \sigma^2)$ which will reduce the estimated value of β. If the true value of β is negative (as the deterrent theory would suggest), then these two effects work in opposite directions. The multiplicative effect reduces the estimated elasticity in absolute value and the additive effect raises its absolute value and the additive effect raises its absolute value. Since the multiplicative effect is more powerful the larger is β while the additive effect remains constant, the total effect can generate either positive or negative effects. If $0 > \beta > -1$ the total effect raises the estimated impact, if $\beta = -1$ the effects cancel each other out. In sum, the measurement error drives the estimate of β towards -1.

Even in this simple case, the measurement error can cut both ways. However, if the estimated elasticities are any guide, the error seems to exaggerate the estimated deterrent effect. Only in a few cases are the estimated elasticities much greater than one, though such large estimates are the rule in the

Mathieson/Passell study. But even these results are no guarantee that the true values are greater or less than -1.

In the multiple regression model the situation is more complicated, and one cannot obtain simple conditions such as (14) to estimate the extent of inconsistency. Hence, what is generally thought to be spurious negative correlation could turn out to be spurious positive correlation. In a recent paper Forst, Filatov, and Klein performed some experiments on a multivariate murder equation (which is discussed in more detail in Section IV), and found that the spurious correlation tends to be negative in such a model. More detailed Monte Carlo experiments would be useful in order to establish whether this finding is true of the cross-section models discussed here. Without such a study it is difficult to estimate the extent of the measurement error effects.

Finally, in evaluating the relevance of such econometric problems for the reliability of the estimates in crime models, it is useful to examine their impact in other areas of applied econometrics work. A reasonable reliability criterion in the econometrics of crime is the state of the art of applied econometrics in general. With regard to measurement error, situations analogous to equation (11) arise in many econometric studies, and completely satisfactory solutions to the problem are yet to be found. Perhaps the best example is in labor economics where labor supply equations are estimated by a regression of hours of work on the wage rate. In most cases the wage rate is determined by dividing individual weekly earnings by the number of hours worked per week. The resulting regression thus takes the form:

$$(15) \qquad \ln H = \alpha + \beta \ln \left(\frac{E}{H}\right) + u$$

Clearly if there is measurement error in H this will generate a spurious correlation which is analogous to that of equation (11). Note an important theoretical difference, however. Microeconomic theory[14] suggests that β will be <u>positive</u> in equation (15) and <u>negative</u> in equation (11). Thus, as long as $|\beta| < 1$ the measurement error works against the economic theory of labor supply and in favor of the economic theory of criminal behavior. It is interesting in this regard that very few empirical researchers have been able to confirm the economic theory of labor supply--hence the fame of the backward bending supply curve. In contrast, very few empirical researchers have been able to disconfirm the economic theory of criminal deterrents.

IV. Special Problems in Measuring the Deterrent Effect of Capital Punishment

The preceding review of cross-sectional empirical work on criminal behavior indicates a striking similarity of results, despite widely different data sets. Moreover, the statistical and econometric problems evident in most of this work do not seem to be substantially more serious in many other areas of applied econometrics. In sum, the general statistical finding that punishment deters crime appears to be reasonably robust across geographical regions and statistical technique. This is not to say, however, that all specific statistical findings about particular types of crimes and punishments are equally robust. On the contrary, a number of statistical results in this area have been shown to be particularly sensitive to the chosen data set and the way the data set is employed to obtain statistical estimates. An important example concerns the deterrent effect of capital punishment. Statistical results about the use of the death penalty have generated considerable controversy in a number of scholarly journals, and have been submitted as evidence before the Supreme Court of the United States.[15] This Section of the paper discusses some special problems in the econometrics of capital punishment.

A. Time Series Evidence

Ehrlich (1975) attempted to measure the deterrent effect of capital punishment by estimating a murder rate equation for the United States--analogous to the various crime rate equations discussed earlier--in which one of the explanatory variables was the ratio of the number of executions for murder to the number of convictions for murder. Two other deterrent variables--measures of the clearance rate for murder and the conviction rate for murder--were also included in the regression, a procedure very similar to that employed by Sjoquist for property crimes in the U.S. Because of data limitations, however, Ehrlich was not able to include the length of prison sentence as an explanatory variable. Five socioeconomic variables were also included in the equation: the labor force participation rate, the fraction of the population between 14 and 24 years of age, an estimate of real per capita income, the unemployment rate, and a time trend. As with the studies reviewed earlier, the equation was specified in log-linear form.

The equation is estimated using an annual time series for the U.S. from 1935-1967 and consequently involves econometric problems and techniques different from those which arose in the cross-section models described earlier. A two-stage estimation procedure proposed by Fair (1970) was used to deal both with the simultaneous equations problem and serial correlation of the residual, the latter being frequently troublesome in time series work. The first stage involves a regression of the deterrent variables (those which are assumed to be endogenous) on current and lagged values of the predetermined variables, lagged values of all the endogenous variables, and a number of exogenous variables excluded from the murder equation (police and total government expenditures per capita, and the fraction of non-whites[16] in the population). The second stage uses an iterative technique to estimate the first order serial correlation coefficient. Ehrlich reports several versions of this equation most of which show significant deterrent effects for the clearance rate and the conviction rate. But contrary to earlier[17] empirical work Ehrlich was also able to report a significant deterrent effect from the execution rate. The estimated coefficient of the execution rate in this murder equation was used by Ehrlich to estimate a trade-off between execution and lives of potential victims of the order of magnitude of 1 to 8.

Passell and Taylor (1977)[18] disputed Ehrlich's estimated trade-off on a number of grounds. First they showed that Ehrlich's model does not remain stable over the sample period, and thereby violates an essential assumption for time series analysis. Observing that murder rates appear to behave differently after the early 1960's, they tested and rejected the hypothesis that the murder equation had the same structure from 1935 to 1962 as from 1963 to 1969. This in itself casts doubt on the estimated coefficients of such a model. Moreover, they found that the model's structure appeared to differ in a very important way: when estimated from 1935 to 1962, no deterrent effect of capital punishment appeared. The elasticity was -.008 and insignificant for the earlier period as compared to -.062 and significant for the whole sample period. And when the sample period was extended to 1964, the coefficient was still insignificant. Hence, Ehrlich's estimated deterrent effect appears to be based soley on the observations from the late 1960's. This dependence of the results on the last few observations is disturbing, because one suspects that a number of factors other than the diminished use of the death penalty may have been causing murder rates to rise. Crime rates in general rose in the late 1960's even though the death penalty was not a factor in the decision to commit most of these crimes.[19]

As an illustration of the importance of this general increase in crime rates, Forst, Filatov, and Klein (1977) included an index of other crime in Ehrlich's equation and found that the deterrent effect of capital punishment becomes insignificant. The magnitude and significance drops even further when an index of violent crime is added to the equation. In addition, one of the many possible explanations for this rise in crime is the increased proportion of youth in the 1960's as a result of the post World War II baby boom. While Ehrlich attempted to control for the age effect, Passell and Taylor (1977) used a narrower age group (18-24) rather than (14-24) and excluded members of the armed forces overseas. They found that the inclusion of this age variable reduced the estimated impact of the capital punishment variable and made it statistically insignificant. Hence, the deterrent effect estimated by Ehrlich is not robust to modifications of the other independent (control) variables which were known to behave differently in the late 1960's. This calls into question his heavy reliance on these years to obtain the deterrent effect. Ehrlich has responded to this criticism by stressing the importance of using all available information: when observations from the late '60's are "thrown away" it is not surprising that the estimated variances of the estimates rise. His critics argue that these observations cannot be added to the earlier period for statistical inference if the model has changed its structure.

A second element of Passell and Taylor's criticism of Ehrlich is the dependence of his results on the log-linear specification. Though there is little a priori reason to prefer the log-linear to the linear functional form, most models of crime have been specified to be log-linear in order to interpret the coefficients as elasticities. Becker's (1986) derivation of the relative importance of severity versus certainty may be one reason for the interest in proportional effects, as mentioned above, but this does not justify the use of this log-linear form. However, the choice of log-linear form appears to be very important in the use of capital punishment. Passell and Taylor have shown that if the murder equation is estimated using the linear form, then the deterrent effect become insignificantly different from zero, while the other variables in the equation remain significant. Forst, Filtov and Klein have replicated this finding.

Of course, there is a wide range of possible transformations of the dependent and independent variables which also might be considered; the linear and log-linear represent only two possibilities. Passell and Taylor (1977) experimented with the Box-Cox (1964) transformation of the form $y_t = (X_t^{\alpha} - 1)/\alpha$ of all the variables in the murder equation. They found that the deterrent effect was significant for some values of α and insignificant for others. They also noted that the value of α which gave the largest correlation between the fitted and actual murder rate was .19, and that this value resulted in a significant deterrent effect of capital punishment (t - ratio = -3.89). Ehrlich has used this later finding as support for his results. However, the important aspect of an examination of alternative functional forms is to show that the statistical significance of the deterrent effect is not robust to arbitrary changes in structure. In fact, the finding that the value of α which gives the best fit, also gives a significant deterrent effect is not surprising from a curve-fitting point of view. Only the execution variable is sensitive in a substantial way to the functional form in this model. Hence, if any value of α gives a significant effect for the execution variable, so also will the one which gives the best fit. With the other variables insensitive to the functional form, the value of α is chosen, in effect, so that the execution variable will be significant.[20] The finding that the "best" α also gives a significant deterrent effect does not therefore invalidate a sensitivity analysis which shows how the deterrent effect is very sensitive to the value of α which is chosen.

The third point made by Passell-Taylor concerns the use of coefficients of a structural equation for policy analysis, as implicit in the trade-off calculated by Ehrlich between executions and the lives of potential victims. As was discussed in Section II of this paper, an important contribution of the economic analysis of crime is the emphasis on the joint determination of crime and punishment. The structural equations of the model represent this joint determination while the reduced form shows how each of the dependent variables are functions of the exogenous variables. The policy implication of this simultaneous determination is that the reduced form rather than the structural form should be used to calculate trade-offs. To illustrate the importance of this result in a simple example, Passell and Taylor (1976) show how the trade-off between executions and murders depends crucially on what is assumed to happen to another endogenous variable: the arrest rate. If the arrest rate is held constant then Ehrlich's 1 to 8 trade-off is obtained; but if the number of crimes cleared through arrest is held constant then the estimated coefficients imply that the trade-off reverses: more executions will increase the number of murders.

In a separate paper, Ehrlich and Gibbons (1977) have discussed this example in great detail. They argue that it is illogical to assume that arrests will be held constant, and just as illogical to use such an assumption to derive a positive trade-off between executions and murders. But the point of this simple example is not to show that more executions will cause more murders. Rather, it illustrates that policy analysis through a structural equation is misleading: the results are highly sensitive to what happens to other dependent variables in the model.

In sum, the time series evidence does not permit one to accept the hypothesis that capital punishment has a significant effect on murder. Statistical findings that the death penalty acts as a significant deterrent are not robust to changes in functional form or period of observation; and the reported trade-offs between executions and murders make arbitrary assumptions about the behavior of other endogenous variables in the model.

B. Cross-Section Evidence

In addition to these time series analyses of capital punishment, a number of cross-section studies of states in the U.S. have recently attempted to measure the deterrent effect of the death penalty using econometric techniques. The equations estimated in these studies are very similar in form to the crime rate equations discussed in Section II, except that a measure of the perceived likelihood of punishment by death is included in the equation. A major advantage of the cross-sectional analysis is that a measure of the length of prison term for murder is available and can therefore be entered into the equation. This Section briefly reviews three of these studies--Passell (1975), Ehrlich (1977) and Forst (1977).

Each of the studies is based on a cross-section for one or more recent Census years: 1940 and 1950 for Ehrlich, 1950 and 1960 for Passell, and 1960 and 1970 for Forst.[21] Ehrlich's and Passell's estimates are based on regressions for each year, while Forst's are based on regressions for the difference between the 1970 and 1960 observations. The results from the cross-section do not show[22] as much sensitivity to the functional form as the time series estimates, though "statistical significance" is usually greater with a nonlinear specification. Forst considers only the linear form while Passell and Ehrlich consider various values of the Box-Cox transformation discussed in the preceding section. The log-linear form is not practical because of the large fraction of states with zero executions.

The central results of the three studies are not consistent. In particular Forst and Passell find no evidence of a deterrent effect of capital punishment while Ehrlich does find evidence of a significant deterrent effect. Because of a basic similarity among the econometric specifications of the three studies, however, it is fairly easy to trace down the source of this major discrepancy: Ehrlich includes in his full sample regressions a dummy variable which distinguishes executing states from nonexecuting states as well as the ratio of executions to convictions in that state. Passell also considers this dummy for nonexecuting states, but only as an alternative to the ratio of executions to convictions. Ehrlich's results tend to confirm that the simultaneous inclusion of this dummy and the execution variable is responsible for the discrepancy between the various studies, for he reports that the deterrent effect vanishes when this dummy is excluded from the nonlinear regressions. Hence, Ehrlich's estimated deterrent effect appears to require that both variables be in the regression.

The reasons for this requirement are intuitively clear. In the U.S. sample, states which used the death penalty during the sample years tended to have high murder rates, even after one controls for other deterrents and socioeconomic factors. Hence a regression which includes either a dummy for executing states or the ratio of executions to convictions, but not both, will be unable to uncover a negative deterrent effect. The estimated coefficient will be positive or at least insignificantly different from zero. The only way to obtain a strong negative effect is to place both variables in the regression. Consequently if one is to place any confidence in the estimated deterrent effect, one must have a good theory as to why both variables should be in the model.

Ehrlich offers two possible explanations: (1) potential murderers in nonexecuting states believe that the probability is greater than zero that they will receive the death penalty (and greater than the probability would be in executing states should executions suddenly drop very close to zero in those states) and (2) certain variables are missing from the equation which would explain the higher murder rate in executing states. It is, of course, very difficult if not impossible to determine whether either of these possible explanations is true. Hence, the theoretical grounds for placing both variables in the regression are not strong.

As with the time series analysis, the results from the cross section, therefore, are inconclusive. Arguments can be made in support of including both variables or of including only one. Hence, unlike other deterrent variables, such as arrest rates, conviction rates, and prison terms, which show strong effect in reducing crime, this research suggests that one cannot reject or accept the claim that capital punishment is a deterrent.

V. Concluding Remarks

A central theme of this review is that in general the econometric problems associated with recent studies of crime are not noticeably more serious than in other areas of applied econometrics--especially when one "seasonally adjusts" for its relatively late development--and, moreover, that the findings are in agreement with a priori theoretical expectations and are robust across widely different geographical cross-sections.[23] However, this general statement must be qualified in particular cases, especially if the results are to be used seriously in policy analysis. The most important qualification concerns the death penalty, where a significant deterrent effect is not usually found.

Such a conclusion does not, of course, imply that the important econometric and measurement problems in this area have been solved. Indeed much additional research is required to clarify a number of important issues before the results can be used with any confidence. To mention a few: the theoretical rationale for including or excluding variables and for using a particular functional form, the estimation of efficient trade-offs using complete models, the quality of the data and the quantitative importance of measurement error, and a reconciliation of time-series and cross-section results.

Columbia University

FOOTNOTES

1. I refer here to the group of studies which began to appear in published form in 1973; for example, Carr-Hill and Stern (1973), Ehrlich (1973) and Sjoquist (1973). Research had begun in the 1960's, however, as evidenced by Ehrlich's 1970 Columbia dissertation and other unpublished work. This paper is a selective review in the sense that it focuses on a number of studies which are representative of different facets of the literature; it does not provide an exhaustive summary or bibliography of all empirical work on criminal behavior. See Nagin (1977), Cook (1977) or Palmer (1977) for additional references. It should be noted that in selecting cross-section studies for review, I have chosen <u>one</u> paper to represent each data source examined, even though several papers may have analyzed the same data. For example, Ehrlich's (1973) study of states in the U.S. is discussed in Section II, while Forst's (1977) study which has somewhat different findings is not. Anyone interested in pursuing empirical work on a given body of data should, of course, consult these other studies as well.
2. See Fisher and Nagin (1976), Cook (1977), and Nagin (1977), for example.
3. See Ehrlich (1973), Block and Lind (1975) and Block and Heineke (1975). A similar analysis is also found in Fleisher (1966). The modifications and extensions of Becker's theoretical model have, in some cases, qualified his results; see especially Block and Heineke (1975). Since our main interest is with empirical issues, we only consider the rudimentary theoretical model here.
4. The analysis can formally incorporate the allocation of time to criminal activity. Corner solutions can then represent legal activity.

5. We are assuming here that any wealth or income effects are dominated by substitution effects.

6. See Sargent and Sims (1977).

7. Aggregation analysis does not appear to be any different than in other areas of applied econometrics and consequently is not discussed here. Note that some of the studies discussed below also consider the endogeniety of expenditures in police.

8. Ehrlich also reports results from a modification of the TSLS estimate to take account of the correlation between different crime types. These seemingly unrelated regression (SUR) estimates are very similar to the TSLS results and are not discussed separately here.

9. Ehrlich (1975) excluded NW from the crime rate equation and uses it as an instrument for TSLS, exactly opposite to the assumption of his 1973 paper.

10. Fisher and Nagin (1976).

11. Avio and Clark argue that another type of measurement error may exist with length of sentence variables.

12. Note that there is some problem in distinguishing between severity and certainty. Ehrlich's measure of certainty is the ratio of imprisonments to crimes, which except for the denominator, is similar to the measures used by Carr-Hill and Stern for severity.

13. Fisher (1966, Chapter 3) examines the possibility of approximate identification using restrictions on the variances, and establishes some general criterion for what he defines as a "nearly identified" equation. In their study of identification in crime models, Fisher and Nagin (1977) do not utilize the concept, however.

14. In both cases we are assuming that substitution effects dominate income and wealth effects.

15. The paper by Ehrlich (1975) and the critique by Passell and Taylor (1975) were submitted as evidence for Fowler vs. North Carolina in a Brief for the United States as Amicus Curiae and in a Reply Brief for the Petitioner respectively.

16. Note that this treatment of the nonwhite fraction is different from Ehrlich (1973); see footnote 9 above.

17. Sellin (1959).

18. The first version of the Passell and Taylor critique appeared as a Columbia Economics Workshop Discussion Paper in February 1975. The less technical aspects of this critique are published in Passell and Taylor (1976). A revised version using alternative data was published in the American Economic Review in 1977 and is discussed here. A number of other authors have reported similar results; in particular Bowers and Pierce (1975) and Baldus and Cole (1975). See also the discussion in Forst, Filatov and Klein (1977).

19. Care must be used in determining which crimes are not influenced by the death penalty. For example, while burglary was not punishable by death during the sample period, the prospective burglar might consider the chance of a murder during the act and the chance that this would be treated as a capital offense.

20. This statement is strictly true in a simple regression with only one explanatory variable. It is approximate in a multivariate regression model if the estimated coefficients of all variables except the one of interest are insensitive to functional form.

21. Forst's cross section included only 32 states because of the lack of conviction data in 1970; Ehrlich's samples are 43 and 44 for the two years respectively, while Passell's are 47 and 44 for his two years.

22. An exception to this generalization is one of Ehrlich's regressions in 1950 where the linear form yields a very insignificant deterrent effect.

23. It should be noted that some researchers have not found significant deterrent effects. See Forst (1977), for example.

REFERENCES

Avio, K. L. and C. S. Clark (1976), "Property Crime in Canada: An Econometric Study," Ontario Council Economic Research Studies, Ontario Canada.

Baldus, D. C. and J. W. Cole, (1975),"A Comparison of the Work of Thorsten Sellin and Isaac Ehrlich on the Deterrent Effect of Capital Punishment," Yale Law Journal, 85, 187.

Becker, G. S. (1986), "Crime and Punishment: An Economic Approach" Journal of Political Economy, 78, 2 (March/April), 526-36.

Block, M. K. and J. M. Heineke (1975), "A Labor Theoretic Analysis of the Criminal Choice," American Economic Review, 65, 3 (June), 314-325.

Block, M. K. and R. C. Lind (1975), "Crime and Punishment Reconsidered," Journal of Legal Studies, 4, January, 241-47.

Bowers, W. J. and G. L. Pierce (1975),"The Illusion of Deterrence in Isaac Ehrlich's Research on Capital Punishment," Yale Law Journal, 85, 170.

Box, G. and D. Cox, (1964),"An Analysis of Transformations," Journal of the Royal Statistical Society, Series B, 26, January, 211-43.

Carr-Hill, R. A. and N. H. Stern (1973), "An Econometric Model of the Supply and Control of Recorded Offense in England and Wales," Journal of Public Economics, 2, 289-318.

Cook, P. J. (1977), "Punishment and Crime: A Critique of Current Findings Concerning the Preventative Effects of Punishment," Law and Contemporary Problems, 41, 164-204 (Winter).

Ehrlich, I. (1973), "Participation in Illegitimate Activities: A Theoretical and Empirical Investigation," Journal of Political Economy, 81, (May/June), 521-65.

__________.(1975),"The Deterrent Effect of Capital Punishment: A Question of Life and Death," American Economic Review, 65, 3 (June), 397-417.

__________.(1977), "Capital Punishment and Deterrence: Some Further Thoughts and Additional Evidence," Journal of Political Economy, 85, 4, 741-788.

Ehrlich, I. and J. C. Gibbons, (1977), "On the Measurement of the Deterrent Effect of Capital Punishment and the Theory of Deterrence," Journal of Legal Studies, 6, (January), 35-50.

Fair, R. C. (1970), "The Estimation of Simultaneous Equation Models with Lagged Endogenous Variables and First Order Serially Correlated Errors," Econometrica, 38, May, 507-16.

Fisher, F. M. (1966), The Identification Problem in Econometrics, McGraw-Hill (New York).

Fisher, F. M. and D. Nagin (1976), "On the Feasibility of Identifying the Crime Function in a Simultaneous Model of Crime Rates and Sanction Levels," in Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime, National Academy of Sciences.

Fleisher, B. (1966), The Economics of Delinquency, Quadrangle, Chicago.

Forst, B., Filatov, V. and L. R. Klein (1976), "The Deterrent Effect of Capital Punishment: An Assessment of the Estimates," in Deterrence and Incapacitation: Estimating the Effects of Criminal Sanctions on Crime, National Academy of Sciences.

Forst, B. (1977), "The Deterrent Effect of Capital Punishment: A Cross-State Analysis of the 1960's," Minnesota Law Review, 61, 5(May),743-767.

________.(1977), "Participation in Illegitimate Activities: Further Empirical Findings," Forthcoming in Policy Analysis.

Mathieson, D. and P. Passell (1976), "Homicide and Robbery in New York City: An Economic Model," Journal of Legal Studies, 6, 83-98.

Nagin, D. (1977), "General Deterrence: A Review of the Empirical Evidence," Forthcoming in Management Science.

Palmer, J. (1977), "Economic Analysis of the Deterrent Effect of Punishment: A Review," Journal of Research in Crime and Delinquency, 14, (January) 4-21.

Passell, P. (1975), "The Deterrent Effect of the Death Penalty: A Statistical Test," Stanford Law Review, 28, 1(November), 61-80.

Passell, P. and J. Taylor (1976), "The Deterrence Controversy: A Reconsideration of the Time Series Evidence," in Capital Punishment in the United States, H. A. Bedau and C. M. Pierce (eds.) AMS Press, 359-371.

________________. (1977), "The Deterrent Effect of Capital Punishment: Another View," American Economic Review, 67, 3 (June), 445-51.

Phillips, L. and H. L. Votey (1975), "Crime Control in California", Journal of Legal Studies, 4, 2(June), 327-350.

Sargent, T. J. and C. A. Sims (1977), "Business Cycle Modeling Without Pretending to Have Too Much A Priori Economic Theory," in New Methods in Business Cycle Researc: Proceedings from a Conference, Federal Reserve Bank of Minneapolis.

Schuessler, K. F. (1952), "The Deterrent Influence of the Death Penalty," Annals of American Academy of Political and Social Science, November 54-62.

Sellin, T. (1959) The Death Penalty, Philadelphia.

Sjoquist, D. L. (1973), "Property Crime and Economic Behavior: Some Empirical Results," American Economic Review, 63, 3 (June), 439-446.