If you have issues viewing or accessing this file contact us at NCJRS.gov.

81810

The National Institute of Law
Enforcement and Criminal Justice

death penalty: a statistical test. *Stanford*

rent Effect of Capital Punishment: Another
a University Department of Economics.
apital punishment: a comment. *Yale Law*

deterrence. Pages 65-84 in S. Rottenberg.
nent. Washington, D.C.: American Enter-

hia: American Law Institute.
T. (1972) *Delinquency in a Birth Cohort.*

NCJRS

OCT 21 1981

ACQUISITIONS

81810

MFI

From: Deterrence and Incapacitation: Estimating The Effects of Criminal Sanctions On Crime Rates, see NCJ# 44669

On the Feasibility of Identifying the Crime Function in a Simultaneous Model of Crime Rates and Sanction Levels

FRANKLIN M. FISHER *and* DANIEL NAGIN

I. INTRODUCTION

In recent years, considerable social science research activity has been directed toward empirically estimating the deterrent impact of criminal sanctions. With few exceptions, the analyses have found a negative and often statistically significant association between crime rates and sanction measures such as clearance rates,[1] interpretable as a measure of probability of apprehension given crime; the ratio of imprisonments to crimes, interpretable as a measure of probability of imprisonment given crime; and time served in prison, a measure of severity of punishment given imprisonment (e.g., Gibbs 1968; Ehrlich 1973; and Sjoquist 1973).

While these negative associations are consistent with the hypothesis that deterrence exists at a measurable level, several reviews (Green-

Franklin M. Fisher is Professor of Economics, Department of Economics, Massachusetts Institute of Technology, and Daniel Nagin is Assistant Professor of Policy Sciences, Institute of Policy Sciences and Public Affairs, Duke University.
NOTE: Contributions by Daniel Nagin were partially supported by PHS research grant no. 1 RO1 MH 28437-01 from the National Institute of Mental Health, Center for Studies of Crime and Delinquency.

[1]The clearance rate is the proportion of reported crimes that are eventually "solved." In general, crimes are solved by the arrest of a suspect.

berg 1977; Gibbs 1975; and Nagin, in this volume) have questioned these results on several grounds. The key issues raised by Nagin are:

1. The processes underlying the generation of data on crimes and sanctions offer alternative explanations for the observed inverse association between crime and sanctions. Variations, either across jurisdictions or over time, in police practices in the recording of offenses reported to them by the public or in the subsequent unfounding[2] of recorded offenses may in themselves generate an inverse association between published crime rates and any sanction variable using published counts of crime in its denominator (e.g., clearance rate, prison commitments per crime). Jurisdictions that record fewer reported crimes and/or unfound more recorded crimes will tend to have lower crime rates and higher measures of such sanction rates. Overt manipulation of clearance and crime reports will serve to generate an even larger negative association between crime rates and the clearance rate. High clearance rates and low crime rates are used as indicators of an effective police department. Police departments may use their discretion not to record or to unfound a reported offense to manipulate reductions in published crime rates. Concurrently, by offering suspects leniency if they admit to previously unsolved crimes, the police can also inflate clearance rates. The negative association between clearance rates and crime rates may simply reflect the varying intensity across jurisdictions with which such practices occur.

Similarly, the observed inverse association between prison commitments per crime and the crime rate may also be a reflection of the plea bargaining process. Plea bargaining will have the effect of understating in published statistics the actual number of prison commitments for more serious offenses because the commitments will be recorded for a less serious offense (e.g., assault charges may be disposed of as disorderly conduct). If plea bargaining is more prevalent in judicial systems that are overcrowded by increased crime, an inverse association between commitments per reported crime (a measure of probability of imprisonment) and crime rates will be induced that is not a reflection of deterrence.

2. The inverse association between crime and sanctions also re-

[2]An offense is said to be "unfounded" when (a) circumstances following the report show that no crime actually occurred (e.g., a reported theft is in fact a case of misplaced property) or (b) there is good reason to believe that no crime occurred (e.g., it is suspected that an offense is reported merely to implicate another individual in wrongdoing).

flects, at least in part, incapacitat
fects. In places where the probabil
time served is longer, a greater pr
will be incarcerated, *ceteris parib*
reduced by physically restraining
element from committing crimes.

3. Motivated by a belief that cr
one another, many recent analyse
tems in which crime is presumed t
presumed to affect crime. To sepa
priori restrictions must be impos
These restrictions have taken the f
cant exogenous variables from one
ing them in one or more of the
restrictions are made on the assu
causal effect on the dependent va
included but has no direct effect
equation from which it is excluded
error, then the estimated coefficier
effect of sanctions on crime as th
estimation procedures. The restri
generating function are often impla
doubts as to the interpretability of

The purpose of this paper is to
raised in (3) by addressing the qu
identify and estimate the deterren
tained hypothesis that crimes a
another.

When two factors x and y are sin
y on x and x on y cannot tell us the
of x on y and y on x, since their n
confounded in both of the resp
example, one cannot estimate the
demanded, q_D, by simply regressin
quantity supplied, q_S, which in equ
dures exist that provide methods
mutual effects of simultaneously
conditions are satisfied. It can be s
tions are not satisfied, then there
mated. Before discussing these n

fects, at least in part, incapacitation effects rather than deterrent effects. In places where the probability of imprisonment is larger and/or time served is longer, a greater proportion of the criminal population will be incarcerated, *ceteris paribus*. The crime rate will thereby be reduced by physically restraining a greater proportion of the criminal element from committing crimes.

3. Motivated by a belief that crimes and sanctions mutually affect one another, many recent analyses have postulated simultaneous systems in which crime is presumed to affect sanctions and sanctions are presumed to affect crime. To separate empirically the mutual effects, *a priori* restrictions must be imposed on the behavior of the system. These restrictions have taken the form of selectively excluding significant exogenous variables from one equation in the system while including them in one or more of the other equations in the system. The restrictions are made on the assumption that a variable has a direct causal effect on the dependent variable in the equation in which it is included but has no direct effect on the dependent variable in the equation from which it is excluded. If these exclusions are seriously in error, then the estimated coefficients are as unsuitable for inferring the effect of sanctions on crime as those estimated by nonsimultaneous estimation procedures. The restrictions used to identify the crime-generating function are often implausible, consequently raising serious doubts as to the interpretability of the estimated parameters.

The purpose of this paper is to pursue the identification problem raised in (3) by addressing the question of whether it is possible to identify and estimate the deterrent effects of sanctions under a maintained hypothesis that crimes and sanctions mutually affect one another.

When two factors x and y are simultaneously related, a regression of y on x and x on y cannot tell us the magnitude of the respective effects of x on y and y on x, since their mutual effects on each other will be confounded in both of the respective regression coefficients. For example, one cannot estimate the causal effect of price, P, on quantity demanded, q_D, by simply regressing q_D on P because P also affects the quantity supplied, q_S, which in equilibrium equals q_D. Statistical procedures exist that provide methods for identifying and estimating the mutual effects of simultaneously related variables provided certain conditions are satisfied. It can be shown, however, that if those conditions are not satisfied, then there is no way the effects can be estimated. Before discussing these methods, we shall first discuss the

reasons for believing that crime affects sanctions as well as that sanctions affect crime.

Economists have argued that for a given level of resources devoted to the criminal justice system (CJS), increased crime rates saturate the resources of the CJS. The effect of the over-utilization of CJS resources is a reduction in the level of sanctions delivered per crime, S. Specifically, if we define a relationship $S = h(C,E)$ that defines S as a function of crime rate, C, and CJS resources, E, then the resource saturation hypothesis would predict that $\partial h/\partial C < 0$ and $\partial h/\partial E > 0$.

A specific example of the resource saturation hypothesis is a predicted negative effect of crime rate on the clearance rate, holding E constant. Although the police will clear more crimes in absolute terms when crime rates increase, the percentage cleared (i.e., the clearance rate) will decrease (Figure 1).

The resource saturation hypothesis is explored in analyses done by Avio and Clark (1974), Carr-Hill and Stern (1973), and Ehrlich (1973). In each of these analyses the structural equation defining sanction level showed a negative and significant association of crime rate with the dependent variable, sanction level. However, because of problems related to identification of the sanction functions (in addition to those related to the identification of the crime function), their results indicating a negative effect of crime on sanctions must be regarded as tentative.

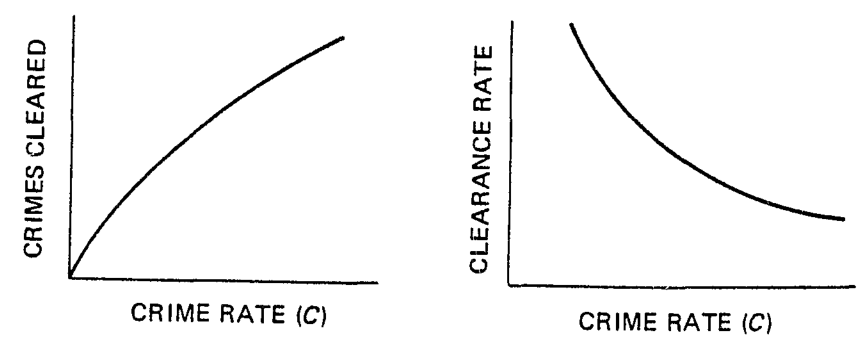


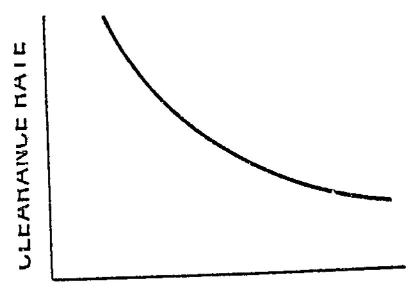
FIGURE 1 Relationship between number of crimes cleared and clearance rate per crime for a fixed level of resources under the assumption of decreasing marginal productivity for police resources.

s sanctions as well as that sanc-

;iven level of resources devoted creased crime rates saturate the over-utilization of CJS resources delivered per crime, S. Specifi- $C,E)$ that defines S as a function E, then the resource saturation : 0 and $\partial h/\partial E > 0$.

saturation hypothesis is a pre- n the clearance rate, holding E ır more crimes in absolute terms ıtage cleared (i.e., the clearance

is explored in analyses done by Stern (1973), and Ehrlich (1973). l equation defining sanction level sociation of crime rate with the owever, because of problems re- ı functions (in addition to those le function), their results indicat- tions must be regarded as tenta-



CRIME RATE (C)

imes cleared and clearance rate per crime on of decreasing marginal productivity

Blumstein and Cohen (1973) and Blumstein *et al.* (1976) have offered still another reason for believing that crime rates will negatively affect sanctions. They have hypothesized that society is willing to deliver only a limited amount of punishment. As crime rates increase, a relatively constant level of punishment is maintained by adjusting the standards defining criminal behavior, reducing the probability of sanctions being imposed or the severity of sanctions imposed or all of these. This might involve a general reduction in sanctions in response to an overall increase in crime or a more selective response that is limited to specific crimes. While Blumstein, Cohen, and Nagin have provided empirical support for the "limits on punishment" hypothesis, their results are also tentative and require further investigation.

Both the "resource saturation" and "limits on punishment" hypotheses predict a negative effect of crime on sanctions. Some have argued the plausibility of increased crime rates causing a toughening of sanctions. This hypothesis is raised, for example, by Forst[3] and Avio and Clark (1974). Empirical evidence supporting this position is scant.[4] Avio and Clark (1974) observed a positive association between crime rate and sentence length. The enactment of the New York Repeat Offender Law and the Massachusetts Gun Law also support the "toughening" position.[5]

The possibility of simultaneity between crime and sanctions, no matter what its cause, raises serious obstacles to empirical analysis and requires that simultaneous estimation be used to estimate the deterrent impact of sanctions in the simultaneous association of crime and sanctions. The separation of the two effects cannot be achieved unless *a priori* assumptions about the specific nature of the simultaneous relationship are invoked. These assumptions, which are called "identification restrictions," are the keystone of simultaneous equation estimation, for data alone are not sufficient for estimating the structural parameters of a simultaneous system "no matter how extensive and complete those observations may be" (Fisher 1966, p. 2).

In the next section, the identification problem will be discussed and its basic role in simultaneous equation estimation illustrated.

[3]Private communication.
[4]However, to the extent that identification problems arise, empirical evidence either way must be viewed with caution.
[5]While this evidence is consistent with the "toughening" hypothesis, in each case the sanction pertains either to sentences or to statutory definition. It is not clear that these official declarations materially alter the level of sanctions actually delivered (e.g., actual time served). If criminals react primarily to cues on actual sanctions, then the "toughening" hypothesis would require evidence of a positive effect of crime on actual sanctions.

II. THE IDENTIFICATION PROBLEM

Simultaneous estimation procedures were developed because classical regression techniques are inadequate for estimating the structural equations in a simultaneous system. In particular, when two variables x_t and y_t are simultaneously determined as indicated by the system (1) shown below (such variables are referred to as endogeneous), then a simple regression of y_t on x_t will generate a biased and inconsistent[6] estimate of b, the parameter defining the direct effect of x_t on y_t, and likewise a regression of x_t on y_t will generate a biased and inconsistent estimate of d, the parameter defining the direct effect of y_t on x_t:

$$y_t = a + bx_t + \epsilon_t \qquad (1a)$$

$$x_t = c + dy_t + u_t \qquad (1b)$$

The respective regression coefficients are not consistent estimates of the structural parameters b and d because the mutual interaction of x_t and y_t makes it impossible to assume that either is independent of the stochastic disturbances ϵ_t and u_t. Since ϵ_t influences y_t, and since y_t influences x_t, it cannot be the case that x_t and ϵ_t are uncorrelated. Hence a regression of y_t on x_t will confound the effect of x_t on y_t with that of ϵ_t on y_t and will not produce a consistent estimate of b.[7]

[6]An estimator is said to be consistent if its probability limit exists and is the true parameter value. Intuitively, this is similar to saying that with a sufficiently large sample the parameter can be estimated with high probability with any desired precision. An estimator that is inconsistent will also, generally, be biased. The converse is often not the case.
[7]The respective covariances of x_t with ϵ_t and y_t with u_t can be shown to be:

$$\sigma_{x\epsilon} = \frac{1}{1-bd}\,[d\sigma_\epsilon^2 + \sigma_{u\epsilon}]$$

$$\sigma_{yu} = \frac{1}{1-bd}\,[b\sigma_u^2 + \sigma_{u\epsilon}]$$

where:

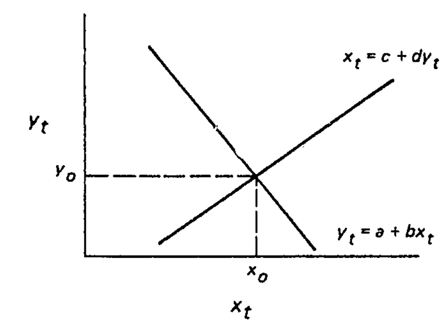
$\sigma_{x\epsilon}$ = covariance of x_t and ϵ_t
σ_{yu} = covariance of y_t and u_t
σ_ϵ^2 = variance of ϵ_t
σ_u^2 = variance of u_t
$\sigma_{u\epsilon}$ = covariance of u_t and ϵ_t

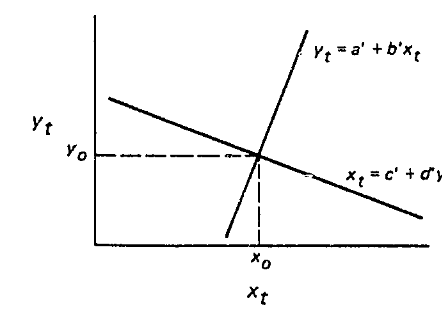
Since $\sigma_{x\epsilon} = 0$ and $\sigma_{yu} = 0$ are respectively necessary conditions for regression to produce consistent estimates of b and d, regression is an inappropriate estimation technique.

Indeed, in the present case, not o niques produce inconsistent parame estimator of those parameters exist: *estimate them from the data*. The ı which presents the non-stochastic c (1b).

Because x_t and y_t mutually affect o single equilibrium point (x_o,y_o). (If duced, then the equilibrium points v This single equilibrium point does n for estimating either of the two equat it. For example, the equilibrium $(x_o$ generated by the system shown in Fi

Indeed, there are an infinite numbe generated (x_o,y_o). There is no way to ı system from the others. Algebraicall any linear combination of equations (tical equilibrium (x_o,y_o). There is no v or (1b) from any such linear combina





EM

vere developed because classical
e for estimating the structural
n particular, when two variables
ed as indicated by the system (1)
rred to as endogeneous), then a
erate a biased and inconsistent[6]
the direct effect of x_t on y_t, and
nerate a biased and inconsistent
he direct effect of y_t on x_t:

$_t + \epsilon_t$ (1a)

$_t + u_t$ (1b)

s are not consistent estimates of
use the mutual interaction of x_t
hat either is independent of the
e ϵ_t influences y_t, and since y_t
hat x_t and ϵ_t are uncorrelated.
found the effect of x_t on y_t with
onsistent estimate of b.[7]

bility limit exists and is the true parame-
hat with a sufficiently large sample the
ity with any desired precision. An es-
e biased. The converse is often not the

ith u_t can be shown to be:

$r_{ue}]$

$r_{ue}]$

ssary conditions for regression to pro-
an inappropriate estimation technique.

Indeed, in the present case, not only will ordinary regression techniques produce inconsistent parameter estimates, but no consistent estimator of those parameters exists. *There is no consistent way to estimate them from the data.* The problem can be seen in Figure 2 which presents the non-stochastic components of equations (1a) and (1b).

Because x_t and y_t mutually affect one another, we will observe only a single equilibrium point (x_o, y_o). (If the stochastic terms were introduced, then the equilibrium points would be scattered about $[x_o, y_o]$.) This single equilibrium point does not provide sufficient information for estimating either of the two equations, (1a) and (1b), that produced it. For example, the equilibrium (x_o, y_o) could just as well have been generated by the system shown in Figure 3.

Indeed, there are an infinite number of such systems that could have generated (x_o, y_o). There is no way to use the data to distinguish the true system from the others. Algebraically, this amounts to observing that any linear combination of equations (1a) and (1b) will produce an identical equilibrium (x_o, y_o). There is no way of distinguishing the true (1a) or (1b) from any such linear combination.

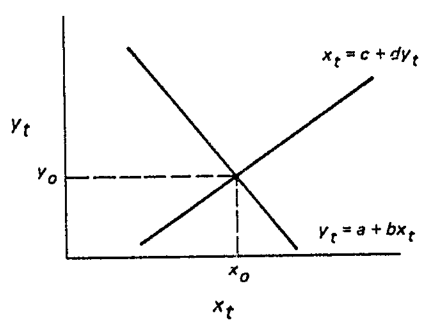


FIGURE 2 A simplified model of a simultaneous relationship between two variables.

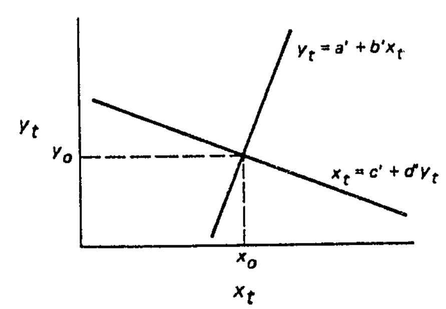


FIGURE 3 Example of an alternative system that generates the same equilibrium point as shown in Figure 2.

Nevertheless, estimating structural equations involving simultaneously related variables is often possible.[8] Under certain conditions, discussed below, simultaneous estimation procedures do provide methods for consistently estimating the true structural equations that generated the observed associations among the simultaneously related variables. However, the true system must satisfy these conditions if the identification problem just exemplified is to be avoided and consistent estimation is to be possible.

The necessary conditions for estimating the true structural equations involve the imposition of *a priori* assumptions about the behavior of the system. Most commonly, these take the form of assuming that variables whose values are determined outside the system ("exogenous variables") or values of endogenous variables determined in prior periods ("predetermined variables") directly affect one or more of the endogenous variables but not all of them. Such restrictions aid in the identification of the structural equation from which the exogenous or predetermined variable is excluded. The exclusion of a variable from one or more equations, however, does not aid in the identification of the structural equations that do include that variable.

To illustrate how such exclusions can identify a structural equation, consider again system (1). As the system is specified, neither equation is identified and neither can be estimated consistently by any method. As indicated earlier, the impossibility of estimating the system is a reflection of there being an infinite set of equation systems that could generate (x_o, y_o).

Suppose, however, that an exogenous variable, T_t, is suspected to have an effect on y_t, but is *known* to have no effect on x_t. Eq. (1a) could then be re-specified as:

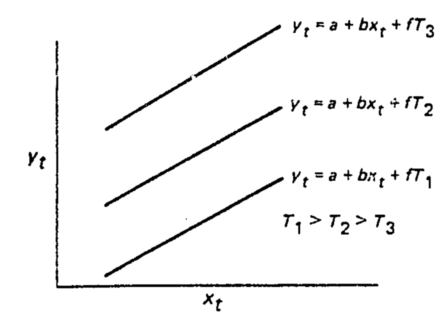
$$y_t = a + bx_t + fT_t + \epsilon_t \qquad (1a')$$

Additionally, assume for concreteness that $f < 0$.[9]

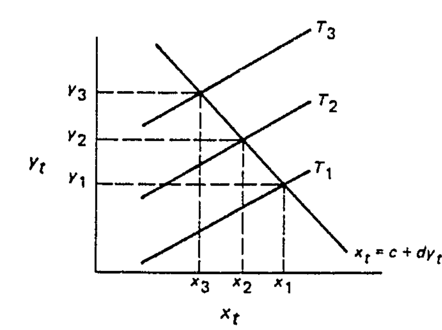
In Figure 4, the non-stochastic component of (1a') is presented as a function of x_t for three different values of T_t. Consistent with the assumption that $f < 0$, Figure 4 shows that for any given value of x_t, y_t is smaller for larger values of T_t.

[8]Ordinary least squares regression, however, remains inconsistent even though consistent estimators exist.

[9]An assumption of $f > 0$ would do just as well; an assumption, however, of $f = 0$ would leave both equations unidentified as before.



In Figure 5, eq. (1b) is superimpose
of T_t. The three points where (1a')
equilibrium values of x_t and y_t for the
If these three equilibrium points we
the structural equation (1b) for x_t wou
however, that (1a'), the structural equ
no variables included in (1b) are excl
The fact that (1a') is not identified
alternative set of structures for y_t wo
values of x_t and y_t. Again, there are
(1a') that would generate the obser
only a single version of (1b), the true
It is important to stress, however
predicated on f, the coefficient of T
were equal to zero, the situation w
single equilibrium point (x_o, y_o) woul
longer be identified.[10]



[10]If f is nearly equal to zero, then (1b) is s
movement in the equilibrium over variation
practice to estimate (1b).

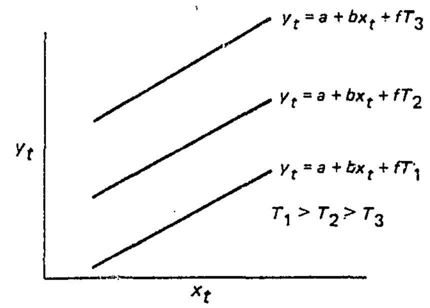


FIGURE 4 y_t as a function of x_t and an exogenous variable, T_t.

In Figure 5, eq. (1b) is superimposed on (1a′) for the different values of T_t. The three points where (1a′) and (1b) intersect indicate the equilibrium values of x_t and y_t for the three different values of T_t.

If these three equilibrium points were observed and connected, then the structural equation (1b) for x_t would be *uniquely* determined. Note, however, that (1a′), the structural equation for y_t, is still not identified; no variables included in (1b) are excluded from (1a′).

The fact that (1a′) is not identified can be seen in Figure 6, where an alternative set of structures for y_t would generate identical equilibrium values of x_t and y_t. Again, there are an infinite number of versions of (1a′) that would generate the observed equilibria; however, there is ·ly a single version of (1b), the true one, that could do so.

It is important to stress, however, that the identification of (1b) is predicated on f, the coefficient of T_t, being different from zero. If f were equal to zero, the situation would revert to that in Figure 2; a single equilibrium point (x_o, y_o) would be observed; and (1b) would no longer be identified.[10]

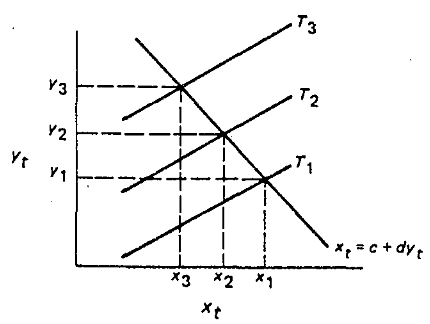


FIGURE 5 The identifying role of an exogenous variable, T_t, in a simplified model of a simultaneous relationship between two variables.

[10]If f is nearly equal to zero, then (1b) is still identified but there will be very little movement in the equilibrium over variations in T_t. Thus, it may be very difficult in practice to estimate (1b).

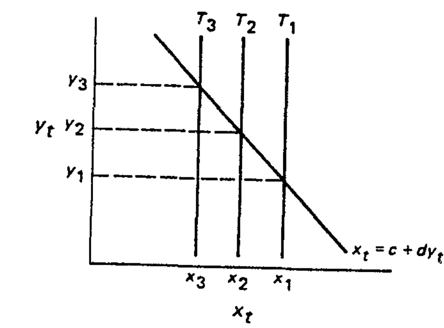


FIGURE 6 An alternative set of y_t functions that generates the same equilibrium points as shown in Figure 5.

When more than two variables simultaneously affect one another, the conditions for identification become somewhat more complicated (see Fisher 1966). Before outlining these conditions, a simplified model of the simultaneous relationship between crime and clearance rates will be examined to illustrate the importance of proper identification for making correct causal inferences.

Suppose, in system (1), x_t is the clearance rate in period t, and y_t is the crime rate in period t. Also, suppose that unbeknownst to us, clearance rates do not in fact affect crime rate (i.e., $b = 0$), but increased crime rates do decrease clearance rates (i.e., $d < 0$). Under the assumption of $b = 0$, a graphical characterization of the unobserved (and as was shown unidentifiable) system is given in Figure 7.

Suppose, however, that the average sentence in period t, T_t, *does* affect crime rates, with longer sentences reducing the crime rate. Thus, the augmented specification of the crime rate equation would be as in equation (1a′), which is repeated below:

$$y_t = a + bx_t + fT_t + \epsilon_t \tag{1a′}$$

The presumed effect of T_t on y_t is illustrated in Figure 8.

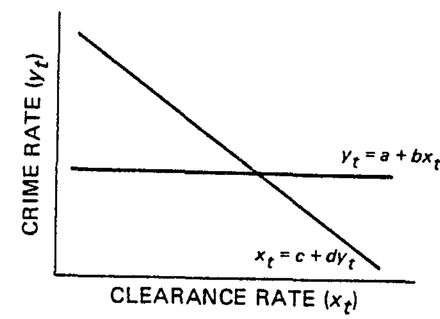


FIGURE 7 A simplified model of the relationship between crimes and sanctions in which sanctions do not affect crimes but crimes do affect sanctions.

FIGURE 6 An alternative set of y_t functions that generates the same equilibrium points as shown in Figure 5.

imultaneously affect one another, :ome somewhat more complicated ıese conditions, a simplified model veen crime and clearance rates will :tance of proper identification for

:learance rate in period t, and y_t is ıuppose that unbeknownst to us, t crime rate (i.e., $b = 0$), but in- rance rates (i.e., $d < 0$). Under the haracterization of the unobserved ʼstem is given in Figure 7.

ıge sentence in period t, T_t, *does* ıces reducing the crime rate. Thus, :rime rate equation would be as in ow:

$$+fT_t + \epsilon_t \qquad \text{(1a')}$$

ıstrated in Figure 8.

FIGURE 7 A simplified model of the relationship between crimes and sanctions in which sanctions do not affect crimes but crimes do affect sanctions.

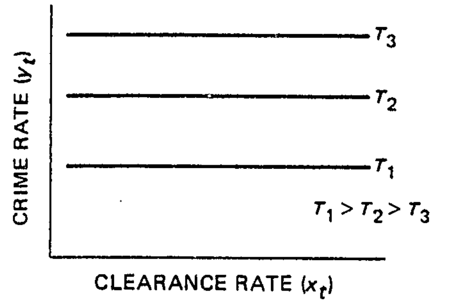


FIGURE 8 The crime rate as a function of the clearance rate and the average sentence (T_t).

In Figure 9, the clearance rate function is superimposed on the crime functions in Figure 8. As was shown previously, the *clearance rate function* is now identified. By connecting the observed intersections in Figure 9, the exact specification for the clearance rate function can be determined. The *crime function,* however, remains unidentified and it will remain unknown and unknowable to us that, indeed, higher clearance rates do not deter crime.

Suppose, however, it were arbitrarily assumed that sentence, T_t, affected clearance rates and not crime rates. Then the mechanics of simultaneous estimation would have allowed an equation for the crime rate to be estimated. That equation, however, would be identical to the one obtained by drawing a line through the equilibrium values of x_t and y_t. Thus, the estimated relation would actually be the relationship describing the effect of crime rate on clearance rates and not clearance rate on crime rates, and so would be completely wrong. In this case, we would conclude that clearance rates have a deterrent effect on crime when in fact they have none.

The very real possibility of making erroneous causal inferences when a model is identified through erroneous assumptions underscores the point that identification is not a minor technical point of estimation. If an equation is not identified, one cannot estimate it. If one tries to do so

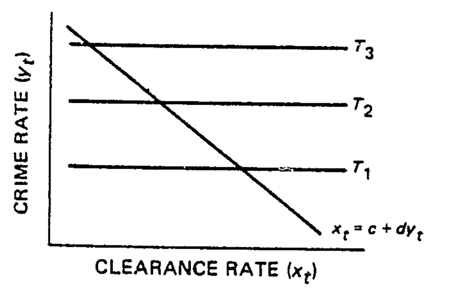


FIGURE 9 The identifying role of average sentence (T_t) in a simplified model of the relationship between crime and sanctions.

using false restrictions to identify the equation, one can draw completely erroneous conclusions from the estimated relationship.

It is thus essential that when exclusion restrictions are used for identification, the restrictions must be carefully justified on the *a priori* grounds that the excluded variables do not directly affect the value of the endogenous variable on the left side of the equation from which they are excluded. If a variable is excluded from an equation merely to facilitate estimation, then the coefficient estimates will remain inconsistent and thus unsuitable for inference about the behavior of the system. Moreover, identifying restrictions must be assumed *a priori* and the nature of the problem is such that restrictions needed to identify can never be tested using data generated by the model under investigation.[11]

In analyzing the mutual association of crime and sanctions, the possibility of making erroneous causal inferences about the causal effect of sanctions on crime is particularly high. Since there are good reasons for believing that crime has a negative causal effect on sanctions, we would expect to observe a negative association in the data between crime and sanctions even if sanctions do not deter crime. Such negative associations are well documented in the deterrence literature (e.g., Ehrlich 1973; Sjoquist 1973; Tittle 1969). Having observed the negative association, we are left with the delicate problem of determining the extent to which it is produced by the negative deterrent effect of sanctions on crime as opposed to the negative effect of crime on sanctions (if the latter effect is indeed negative).[12]

In view of the importance of the identification problem, we shall review some of the restrictions that have been used by some authors to identify the crime functions so that the validity of their findings on the deterrent effect of sanctions can be put into perspective. When evaluating the validity of such restrictions, one should keep in mind that crime-function restrictions presume that the variables involved affect either sanctions, police expenditures per capita (a variable commonly hypothesized to be simultaneously related to crime), or other endogenous variables included in the model, but do not directly affect the crime rate itself.

Ehrlich (1973) identified his crime function by excluding from it (but including elsewhere in his model) the following variables:

[11]However, other data generated in other ways (by experiment, for example) can be so used.

[12]Indeed, in a complex model, such an observed negative association could occur even if neither direct effect is negative because of relations among the disturbance terms.

1. The crime rate lagged one peri
2. Police expenditures per capita
3. Unemployment rate of civilian
4. Percent of males aged 14-24
5. Percent of population living in
6. Males per female
7. A southern regional variable
8. Mean years of schooling of po
9. Total population.[13]

In Carr-Hill and Stern (1973), tl excluding:

1. Total population
2. Proportion of reported crimes
3. A measure of the proportion of

Avio and Clarke (1974) estimate a ance rates, and police expenditures termined. The crime function is ider

1. Population density
2. The total population
3. Police expenditures lagged one
4. Motor vehicle registration per
5. Crimes against persons lagged

In all these papers, identification exclusion of socioeconomic variab variables from the crime function. It argument for the exclusion of the SE these SES and demographic correla determine which among them do ha but it is simply not plausible to assu have a direct effect on crime, while rectly affect either sanctions or polic

[13]In his Ph.D. dissertation, Ehrlich (1970) es above unemployment, age, and education va significant association between crime rate an tified in part by the exclusion of the remainin apparently arbitrary set of identification restr

[14]Indeed, Ehrlich's own theoretical model s does have such an effect.

1. The crime rate lagged one period
2. Police expenditures per capita lagged one period
3. Unemployment rate of civilian males aged 35-39
4. Percent of males aged 14-24
5. Percent of population living in SMSAS
6. Males per female
7. A southern regional variable
8. Mean years of schooling of population over age 25
9. Total population.[13]

In Carr-Hill and Stern (1973), the crime function is identified by excluding:

1. Total population
2. Proportion of reported crimes that are violent
3. A measure of the proportion of the population that is middle class.

Avio and Clarke (1974) estimate a model in which crime rates, clearance rates, and police expenditures per capita are simultaneously determined. The crime function is identified by excluding:

1. Population density
2. The total population
3. Police expenditures lagged one period
4. Motor vehicle registration per capita lagged one period
5. Crimes against persons lagged one period.

In all these papers, identification of the crime function relies on the exclusion of socioeconomic variables (SES) and lagged endogenous variables from the crime function. It is difficult to imagine any plausible argument for the exclusion of the SES variables. Intercorrelation among these SES and demographic correlates of crime makes it difficult to determine which among them do have a causal association with crime, but it is simply not plausible to assume that such SES variables do not have a direct effect on crime, while also assuming that each does directly affect either sanctions or police expenditures per capita.[14]

[13] In his Ph.D. dissertation, Ehrlich (1970) estimated a crime function that includes the above unemployment, age, and education variables and found a negative and generally significant association between crime rate and sanctions. This crime function was identified in part by the exclusion of the remaining variables listed above, a different but still apparently arbitrary set of identification restrictions.

[14] Indeed, Ehrlich's own theoretical model specifies that unemployment in particular does have such an effect.

Further, two of the analyses also use the exclusion of lagged endogenous variables to identify the crime function. For the estimation procedures employed, the use of such restrictions to identify rests crucially upon an assumption of no serial correlation in the stochastic disturbance terms in the equations, because these estimation procedures treat lagged endogenous variables as uncorrelated with current disturbances. If current and lagged disturbances are correlated, this assumption cannot be true. (This point will be discussed in greater detail below.) While methods exist to handle serial correlation, the analyses discussed do not use such methods. There are cogent reasons, which will also be discussed, for believing (a) the assumption of no serial correlation to be incorrect and (b) there is positive serial correlation in the disturbances for the type of data used in these analyses.

Assuming that crime and sanctions are simultaneously related, our conclusion is that it is most unlikely that the authors mentioned have successfully identified and consistently estimated the deterrent effect of sanctions. Consequently, one can have little confidence that the estimated sanctions coefficients are consistent. Moreover, the magnitude of the inconsistency seems likely to be substantial since the restrictions used to identify seem unlikely to be even approximately correct (see Fisher 1961). Consequently, the resulting parameter estimates cannot be used for causal interpretation.

A crucial question is then: Can the crime function ever plausibly be identified, i.e., can we ever hope to find variables that influence sanctions but have no direct effect on crimes? This question, which is the central topic of this paper, is the focus of the next section. The question of the feasibility of identifying the crime function requires an appreciation of some more generalized identification concepts. Thus, before we turn to the topic of feasibility, we shall develop these concepts.

III. SOME MORE GENERALIZED IDENTIFICATION CONCEPTS

The prior discussion has focused on the requirements for identifying the structural equations in a system where only two variables are simultaneously related. We shall now generalize to a situation where M variables simultaneously affect one another.

Suppose we specify the interrelationship of the M variables by:

$$
\begin{aligned}
y_1 &= a_{12}y_2 + a_{13}y_3 + \ldots + a_{1M}y_M + b_{11}x_1 \\
&\qquad + b_{12}x_2 + \ldots + b_{1N}x_N + \epsilon_1 \\
y_2 &= a_{21}y_1 + a_{23}y_3 + \ldots + a_{2M}y_M + b_{21}x_1 \\
&\qquad + b_{22}x_2 + \ldots + b_{2N}x_N + \epsilon_2 \qquad (2) \\
&\;\;\vdots \\
y_M &= a_{M1}y_1 + a_{M2}y_2 + \ldots + a_{MM-1}y_{M-1} + b_{M1}x_1 \\
&\qquad + b_{M2}x_2 + \ldots + b_{MN}x_N + \epsilon_M
\end{aligned}
$$

where:

y_i = the i^{th} endogenous variable ($i = 1, \ldots, M$)
a_{ik} = the coefficient defining the magnitude of the direct ("causal") effect of y_k on y_i
x_j = the j^{th} non-endogenous variable ($j = 1, \ldots, N$)
b_{ij} = the coefficient defining the magnitude of the j^{th}, non-endogenous variable's direct effect on y_i
ϵ_i = the stochastic component of the i^{th} structural equation.

As was shown previously, when variables are simultaneously related, the empirical observations of the system's behavior, no matter how well measured or extensive they may be, are not sufficient for consistently estimating the structural relationships. Consider the first structural equation in system (2). Estimation of the relationship would require generating $M - 1 + N$ parameter estimates. However, the limits of empirical information are such that only N independent pieces of information can be obtained from the data to estimate the $N + M - 1$ parameters of this equation. This corresponds to the fact that only the N non-endogenous variables, the x_j, can be varied independently. The M endogenous variables, the y_i, are determined (except for stochastic effects) once the x_j are set. If there were no stochastic effects, we could think of performing experiments (or having nature perform them for us) by setting the values of the x_j and observing the effect on the y_i. There would be, however, only N independent ways of setting the N non-endogenous x_j, and further experiments would be redundant.

In the stochastic case, the corresponding fact is that we are entitled to assume (at most) that each of the N non-endogenous x_j is uncorrelated with the disturbances, ϵ_i, and in particular with the disturbance from the first equation, ϵ_1. The y_i cannot be so uncorrelated.

If $M = 1$ so that there were no simultaneity, then these N zero correlations would suffice to allow the consistent estimation of the first (and only) equation by ordinary regression. In that case, only exogenous variables would appear on the right side of that equation and the N zero correlations would satisfy the necessary conditions for ordinary regression to generate a consistent estimator—namely, that the regressors be uncorrelated with the disturbance. Where $M > 1$ and there is simultaneity, these N zero correlations are not enough to recover the $M - 1 + N$ parameters of the first equation.

Another way of putting it is to say that analysis of the data can at most only tell us about the full effects (direct and indirect) of the x_j on the y_i (from the "reduced form" in which the equations are solved for the y_i only in terms of the x_j and ϵ_i). The direct effects of the x_j on the y_i (the b_{ij}) and the direct effects of the y_i on each other (the a_{ik}) cannot be recovered from the data without at least $M - 1$ additional independent pieces of information for each equation.[15] Such additional information must come from outside, *a priori* considerations.[16]

The situation is completely isomorphic to the logical impossibility of finding a unique solution to a system of linear equations in $M + N - 1$ unknowns, when only N independent equations are available. A unique solution can only be obtained if $M - 1$ additional independent equations, comparable to our restrictions, are imposed. The identification restrictions in simultaneous equation estimation provide the $M - 1$ additional restrictions that sufficiently augment the empirical information to allow the estimation of the structural equation.

The $M - 1$ additional equations in the system of linear equations in $M + N - 1$ unknowns are as important in specifying a unique solution as the N original equations. Similarly, the identification restrictions are as important in the determination of the coefficients as the observational information.

The additional $M - 1$ restrictions can be (but need not be) generated by assuming that $M - 1$ of the parameters in the equation are zero. The $M - 1$ restrictions could be generated if we assumed $a_{1i} = 0$ ($i = 2, \ldots, M$), which is to assume that y_1 is *not* simultaneously related to any of the other y_i's. Since the x_i's are assumed to be uncorrelated with ϵ_1, the coefficients of the first equation could then be consistently estimated by ordinary least squares.

Suppose, however, that we conclude that *a priori* considerations allow us only to assume that $(M - 1) - k$, *where* $0 \leqslant k < M - 1$, of the

[15] This is a necessary but not sufficient condition for identification. For a full discussion see Fisher (1966).
[16] See Fisher (1966) for a complete discussion.

a_{1i}'s are zero. We must still estimate $k + N$ parameters, which can still not be done using only the N pieces of empirical information available.[17] The additional k pieces of information can be generated if *a priori* considerations would allow us to assume plausibly that k of the N non-endogenous x_j do not enter the first equation but do enter one or more of the other equations (i.e., k of the $b_{1j} = 0$ but $b_{ij} \neq 0$ for some $i \neq 1$). By assuming that k of the b_{1j} are zero, it becomes unnecessary to estimate them. Thus the N pieces of empirical information can be used to estimate the remaining N parameters consistently. It must be emphasized, however, that the remaining N parameters will only be consistently estimated if the *a priori* considerations that led to the assumptions that $M - 1 - k$ of the a_{1i}'s and k of the b_{1j}'s were zero are correct.[18] Thus, any empirical conclusion hinges critically on the validity of those *a priori* premises.

When only $M - 1$ restrictions can be imposed and the equation in question is identified, it is said to be "just identified." This terminology derives from the fact that if we can generate only $M - 2$ restrictions, then the equation will not be identified (i.e., unidentified). Being short only a single restriction means that there exists more than one, and in general an infinite number of equations that are consistent with the data. All such equations are observationally equivalent to the true one. Thus, it must be remembered that from the perspective of the existence of a consistent estimator, one is no better off having $M - 2$ restrictions than zero restrictions. In either case, no consistent estimator will exist and no causal inference can be made about the equation for y_1. In some of the models to be examined in the next section, this point will return to haunt us.

Sometimes it is also possible to generate more than $M - 1$ restrictions and to identify the equation in more than one way. In such instances, the equation is said to be "over-identified" and, since we have more than N pieces of information to estimate less than N parameters, estimation, of course, remains possible.

Before turning to the next section on the feasibility of identifying the crime function, several important points must be made. In order of importance, they are: First, if an equation is just identified, then the restrictions used to identify it cannot be tested with the data being analyzed. The untestability of the restrictions follows from the fact that a model cannot even be estimated unless we assume they are true;

[17]In the earlier discussion, $M = 2$ and $k = 0$; thus, we needed only one identification restriction.

[18]Fisher (1961) shows that the magnitude of the inconsistency in parameter estimates is directly related to the degree of "correctness" of the identification restrictions.

[e.g., the clearance rate's specification (1b) cannot be estimated unless we assume that T_t does not enter (1b). Since we cannot estimate (1b) if T_t does enter it, then we cannot test whether it should enter (1b).]

A related point follows when a model is over-identified, that is, when there are alternative ways to just-identify it. One can estimate the model under a variety of subsets of just-identifying restrictions, with each of the resulting model estimates being contingent upon the validity of the just-identifying subset used. If one has little or no faith in the validity of any one of the subsets, then even if one gets the same results under each subset (for example, sanctions do not deter crime), then one cannot conclude that those results are valid.

Second, any additional restrictions beyond a set of $M - 1$ just-identifying ones can be tested. Those tests are, however, contingent upon the validity of the $M - 1$ just-identifying restrictions. If one has faith in the validity of these $M - 1$ restrictions, then one can have faith in the validity of the empirical tests of the additional over-identifying restrictions. But, if one has little faith in the validity of the just-identifying restrictions, one can have only little faith in the validity of the test of the remaining restrictions. One implication of this point is that if one generates a set of over-identifying restrictions—but in this set there does not exist a subset of just-identifying restrictions whose validity is unquestionable (or nearly so)—one cannot gain a valid test of the set of restrictions by exhaustively testing each restriction under the assumption that the remaining ones are correct.[19]

IV. ON THE FEASIBILITY OF IDENTIFYING THE CRIME FUNCTION

In this section, we shall examine the central issue of this paper: Can the crime function be plausibly identified? We shall proceed by first examining the simplest model in which a single crime type and sanction type are simultaneously related. Several categories of just-identifying restrictions, none of which are mutually exclusive, will be analyzed for their strengths and weaknesses. The single-crime-type, single-sanction-type model overly simplifies the real phenomenon of multiple crime types and multiple sanction types. However, to date no analyses have attempted to estimate models in which more than one crime and sanction type are simultaneously related. More important for our pur-

[19]There do exist methods for testing an entire set of over-identifying restrictions symmetrically; however, such tests are not very strong as indications of which restrictions are incorrect. See Fisher (1966, Chapter 6).

m (1b) cannot be estimated unless . Since we cannot estimate (1b) if vhether it should enter (1b).] el is over-identified, that is, when lentify it. One can estimate the just-identifying restrictions, with being contingent upon the valid- . If one has little or no faith in the n even if one gets the same results ictions do not deter crime), then ts are valid.

is beyond a set of $M - 1$ just- e tests are, however, contingent dentifying restrictions. If one has strictions, then one can have faith of the additional over-identifying aith in the validity of the just- only little faith in the validity of . One implication of this point is ntifying restrictions—but in this ist-identifying restrictions whose so)—one cannot gain a valid test ely testing each restriction under es are correct.[19]

ENTIFYING THE CRIME

entral issue of this paper: Can the ? We shall proceed by first exam- ngle crime type and sanction type categories of just-identifying re- y exclusive, will be analyzed for The single-crime-type, single- the real phenomenon of multiple es. However, to date no analyses which more than one crime and ted. More important for our pur-

t of over-identifying restrictions symmet- g as indications of which restrictions are

poses, such simple models will serve to highlight the strengths and weaknesses of some different categories of just-identifying restrictions. These points will remain valid in analyzing more complex models.

We shall then consider the more complex but more realistic models in which (a) a single crime type is simultaneously related to multiple sanction types and (b) multiple crime types and a single sanction type are simultaneously related. We shall not consider under a separate heading the most complex model in which multiple crime and sanction types are simultaneously related because the problematic feasibility of identifying such a model will become clear from the discussion of the preceding two model types. The principal focus of our discussion will be the identification of simultaneous models. The mutual association of crime and sanctions may, however, occur with time lags rather than simultaneously. In the Appendix we shall point out the difficulties with results based upon path models, which are a specific class of lagged models, and then discuss the difficulties likely to be encountered in estimating more general classes of lagged models.

None of the models that will be discussed will explicitly include SES variables. While SES variables should indeed be included in a specification of the crime function, we do not envisage the exclusion of SES variables being plausibly used as identification restrictions. Such exclusions would have to be predicated upon *a priori* considerations that allow one to assume that the excluded SES factor directly affects some other endogenous variable in the system but not crime. Currently we simply do not have a sufficiently well-developed and validated theory of the socioeconomic factors affecting crime and sanctions plausibly to assume that some SES factor can be excluded from the crime-generating model but included elsewhere in the system. Some new insight in this regard would, of course, be very useful.

The absence of explicit consideration of SES effects should not be interpreted as indicating that we believe these effects to be inconsequential; their effects are undoubtedly substantial, but the mechanism of their operation is simply not understood well enough plausibly to employ SES variables as identification restrictions. Thus, our discussion omits SES variables only for expositional convenience. Most models would include such variables, at least in the crime function. However, it is the exclusion of such variables from the crime function (but not from other equations) that would aid identification.[20]

[20]Naturally, no model is likely to include all relevant SES variables. Omitted SES effects become part of the disturbance terms. We shall later discuss at length the behavior of omitted SES factors on these stochastic components of the model since appropriate specification of such behavior is crucial to making consistent estimates of the parameters.

A. SINGLE-CRIME-TYPE, SINGLE-SANCTION MODELS

1. Models Using Expenditures as an Identifying Omitted Variable

Suppose we specify the following model:

$$C_t = f(S_t) + \epsilon_t^1 \quad (3a)$$

$$S_t = h(C_t, E_t) + \epsilon_t^2 \quad (3b)$$

where:

$f(S_t)$ and $h(C_t, E_t)$ are linear functions[21]
C_t = crime rate in t
S_t = sanctions per crime in t
E_t = criminal justice system (CJS) expenditures in t
ϵ_t^i = stochastic error ($i = 1,2$) whose properties are to be discussed.

In this model, which is also characterized by the flow chart in Figure 10, C_t is determined by S_t, and S_t is determined jointly by C_t and E_t. The CJS expenditures variable, E_t, enters the equation for S_t under the theory that increased resources devoted to the CJS, as measured by E_t, will decrease the resource saturation effect of any given level of crime, C_t (i.e., $\partial h/\partial E_t > 0$). As noted earlier, the resource saturation theory is one of the primary theories underlying simultaneous models of crimes and sanctions.

In this system, there are two endogenous variables, C_t and S_t. The crime equation includes one right-side endogenous variable, S_t. Estimation of eq. (3a) will thus require that one identification restriction be imposed. [Within the context of the identification rules laid out in the previous section, $M = 2$ and therefore we need $M - 1 = 1$ restriction to identify eq. (3a).]

In this system, E_t is not included in the crime function. This exclusion, which can be used to provide the necessary single identifying restriction to estimate eq. (3a), is predicated upon the theory that E_t affects crime only insofar as it affects the capability of the CJS to deliver sanctions. For sanctions delivered by the courts (e.g., conviction, im-

[21]In this analysis, we assume for simplicity that all functions are linear. Nonlinearities in the sanctions function can aid in identification, but only if one is sure of the functional form of the nonlinearity and sure that similar nonlinearities are not present in the crime equation. Such precise information on functional forms is seldom available and is certainly not so in this case. (See Fisher, 1966, Chapter 5, for extended discussion.)

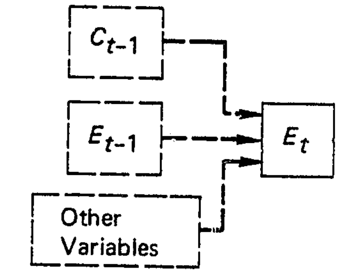


FIGURE 10 Diagram of model variable. The possibility that C_t. expenditures at t but are omitte aid in the identification of the la do not appear anywhere in the s effect captured beyond taking e loop. Another way of putting it ables from the crime equation d the sanctions equation since the equation either.

prisonment) or regulated by cor such an assumption seems reaso itures and S_t is defined as the cl E_t has no direct effect on C_t is su

The level of police expenditur police, since in two identical c penditures is likely to have a la have an independent deterrent e by clearance rate, because the prehension probability (which is ing when S_t refers to police-deli from multiple cues from his env observe the actual apprehension ure it roughly. One such measu fellow criminals with whom he Perhaps this frequency can be ap criminal's perception of appreh have to be based solely upon th estimates. He is likely, in makir bility, to react to additional cue intensity of the police presence.

To the extent that police visi apprehension probability and th

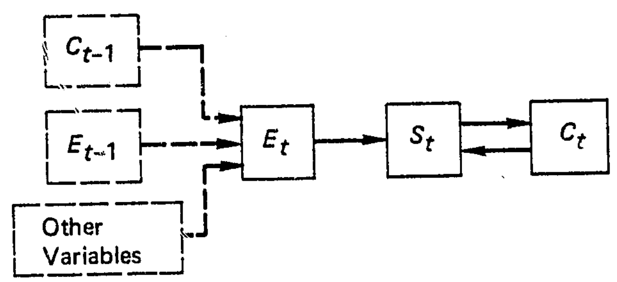


FIGURE 10 Diagram of model using expenditures as an identifying variable. The possibility that C_{t-1}, E_{t-1}, and other variables affect expenditures at t but are omitted from the crime equation does not aid in the identification of the latter. This is because these variables do not appear anywhere in the sanctions-crime loop and have no effect captured beyond taking expenditures as exogenous to that loop. Another way of putting it is that the omission of such variables from the crime equation does not help to distinguish it from the sanctions equation since the variables do not appear in that equation either.

prisonment) or regulated by corrections (e.g., time served in prison), such an assumption seems reasonable. However, if E_t is police expenditures and S_t is defined as the clearance rate, then the assumption that E_t has no direct effect on C_t is suspect.

The level of police expenditures is likely to influence the visibility of police, since in two identical communities, the one with greater expenditures is likely to have a larger police force. Police visibility may have an independent deterrent effect beyond S_t, where S_t is measured by clearance rate, because the potential criminal's perception of apprehension probability (which is the "true" measure of S_t we are seeking when S_t refers to police-delivered sanctions) undoubtedly derives from multiple cues from his environment. A potential criminal cannot observe the actual apprehension probability, but rather can only measure it roughly. One such measure is the frequency with which he and fellow criminals with whom he has contact experience apprehension. Perhaps this frequency can be approximated by the clearance rate. The criminal's perception of apprehension probability, however, does not have to be based solely upon these undoubtedly inaccurate frequency estimates. He is likely, in making his estimate of apprehension probability, to react to additional cues from the environment—such as the intensity of the police presence.

To the extent that police visibility provides an independent cue of apprehension probability and thus acts as an independent direct deter-

rent distinct from the indirect effect of an increased police presence on clearance rates and hence on crime, then E_t should appear directly in the equation for C_t. Such an appearance, however, would leave the crime function unidentified.

Putting such considerations aside and presuming the exclusion of E_t from the crime equation to be valid, that exclusion will identify the crime equation if either of the following statements is true:

1. Expenditures are fully exogenous. To assume that E_t is exogenous is to assume that neither C_t nor S_t in the current period or in prior periods affects E_t. An assumption of exogeneity seems untenable because it is likely that the level of crime affects the level of expenditures, at least across jurisdictions and probably over time. The observed positive association between police expenditures per capita and crime rate provides some evidence for the likelihood of such an effect (see, for example, McPheters and Strong 1975).
2. Expenditures are influenced only by lagged crime rates and are therefore predetermined, although not fully exogenous. This seems more reasonable than does full exogeneity. Due to the government budgeting cycle, the level of E_t is specified before the beginning of period t. That level, although probably influenced by the crime rate, is influenced by rates in prior periods, for example, C_{t-1}. Thus, E_t is a predetermined variable.[22]

Granting that E_t is predetermined, a further crucial assumption must be made about the behavior of the stochastic components, ϵ_t^i. We must specify the behavior of these stochastic terms over time. We could assume that the errors are independent over time, or we could make a less restrictive assumption that they are serially correlated. For example, we might assume that they follow a first-order autoregressive process, characterized by:

$$\epsilon_t^i = \rho_i \epsilon_{t-1}^i + \delta_t^i \tag{4}$$

where:

ρ_i = a parameter
δ_t^i = non-serially correlated disturbance term.

[22] It should be noted that if C_t does influence E_t directly, perhaps because the budget is adjusted in t in reaction to C_t, then E_t becomes determined simultaneously with C_t and S_t, and the crime function is no longer identified even if E_t does not appear in it. Some additional restrictions involving a nonendogenous variable are necessary.

Such assumptions about the serial relationships among the ϵ_t^j are critical for identification. In our previous discussion on the limits of the empirical information in a simultaneous system, we stated that the maximum number of independent pieces of empirical information available for consistently estimating each structural equation was N, where N equals the number of non-endogenous variables in the system. This was because of the assumption that there are N non-endogenous variables that are uncorrelated with the stochastic disturbances and thus that can be varied independently. If that assumption fails for J_1 of the non-endogenous variables, then the number of pieces of empirical information for consistently estimating each structural equation is reduced to N-J_1. In effect, an additional J_1 of the variables become endogenous.

When using predetermined variables for identification, the possibility that the disturbances are serially correlated must be given special consideration. If the ϵ_t^j are serially correlated [for example, a first-order autoregressive process as in eq. (4)], then the predetermined variables will necessarily be correlated with at least some of the stochastic components. In particular, E_t will be correlated with ϵ_t^1 because ϵ_t^1 is correlated with ϵ_{t-1}^1 and E_t is a function of C_{t-1}, which is in turn a function of ϵ_{t-1}^1.

When serial correlation among the disturbances is thought to be present, estimation still remains possible if one correctly specifies the specific structure of the presumed serial correlation. If one is not certain of the specific structure of the serial correlation, and one rarely is, then the less restrictive the assumption the better. For example, the first-order autoregressive assumption is less restrictive than assuming no serial correlation because the latter will occur for the special case of all the ρ_i zero. However, if the model is estimated under an assumption of no serial correlation, then the possibility of serial correlation of some specific type cannot be tested. Even less restrictive assumptions about the nature of the serial correlation (higher-order processes, for example) can be made, but some specific assumptions must be made.

Excepting a capital punishment analysis by Ehrlich (1975), all simultaneous analyses have employed estimation methods that generate consistent estimates only when there is no serial correlation of any kind among the disturbances. If the exclusion of a predetermined variable is used as an identification restriction, as with E_t in the model under consideration, the validity of using that restriction when using these methods turns on the assumption of no serial correlation. If the assumption is incorrect, then the parameter estimates will be inconsistent.

The assumption of no serial correlation among the disturbances is not only fundamental in cases like this; it reflects implicit assumptions about real effects stemming from factors influencing crime or sanctions that are captured in the disturbances because they are not explicitly included in the model. Deciding whether the assumption of no serial correlation can plausibly be maintained thus requires consideration of such factors.

In the crime function shown in eq. (3a), the variables not explicitly included would include all SES variables that affect crime. However, this is because of the simplistic nature of eq. (3a) adopted for expositional purposes. As already remarked, in practice, if eq. (3a) were to be estimated, some SES variables would be explicitly included. Nevertheless, some part of the stochastic disturbance, ϵ_t^1, would still consist of SES effects. It is impossible to include all the SES variables influencing crime both because we do not know all of them or cannot measure them and because there are likely to be many of them, each with a small effect. In addition, if included SES variables affect crime in ways only approximated by our choice of functional form in eq. (3a), then departures from that approximation influence the disturbance term.

From this perspective on the factors generating ϵ_t^1, is it reasonable to assume no serial correlation in ϵ_t^1? The answer, we believe, is no. Many of the SES variables influencing ϵ_t^1 change only gradually over time. Thus, if the realized values of these variables in period t are such that the disturbance is positive in period t, it is likely that their realized values in period $t+1$ will lead to a positive disturbance as well. Hence we should expect positive serial correlation in ϵ_t^1. One possible characterization might be the first-order autoregressive process shown in eq. (4), with $\rho_1 > 0$.

When using data with a cross-sectional component, the most common type of data utilized in deterrence analyses, the likelihood of serial correlation is particularly high because there is likely to be particularly wide variation in the values of excluded variables across the sampling units (usually states). Put simply, locations whose actual crime rate is higher than predicted by the systematic part of the equation in one year are likely to remain so in the next year.

The implausibility of an assumption of no serial correlation requires that estimation be done under a less restrictive assumption about the serial correlation of the stochastic terms if inconsistency is to be avoided. We shall not address the question of what sort of assumption on the nature of the serial dependence is plausible. The question deserves further attention, but it can be said that the less restrictive the assumption, the better. One possibility, given enough data, would be to allow for an autoregressive relationship of order γ, where:

$$\epsilon_t^i = \sum_{j=1}^{\gamma} \rho_{ij}\epsilon_{t-j} + \delta_t^i \tag{5}$$

Estimation under any assumption of serial dependence, however, requires the use of data with a time-series component. For example, the γ^{th} order autoregressive assumption would require that the time-series component in the data be at least $\gamma + 1$ periods. Pure cross-sectional data cannot be used.

To summarize, we conclude that the exclusion of the expenditures variable cannot be used plausibly to identify the crime function, at least with cross-sectional data. To do so at best requires the very implausible assumption of serial independence in the stochastic components. To estimate a model under any assumption of serial dependence requires time-series data and thereby precludes the possibility of using only cross-sectional data.

Moreover, as we have seen, the use of the expenditures restriction, no matter what one assumes about the nature of the serial dependence, hinges upon the assumption that E_t does not directly affect crime. If S_t and E_t are defined in terms of court-related activities only, this seems plausible. If E_t and S_t pertain to the police, however, then the assumption that E_t does not directly influence C_t is questionable. Expenditures on police will be closely linked to the visibility of police in the community, and police visibility may indeed be a very important factor in deterring crime. Further, if expenditures on courts and expenditures on police vary together, then one may simply be fooling oneself about identification in specifying and estimating a model in which E_t relates only to courts.

2. *Models Using Prison Cell Utilization*

In the system shown below, C_t is again a function of S_t and S_t is a function of C_t. Additionally, S_t is specified to be a function of prison-cell utilization, U_t, defined to be the ratio of the prison population in t, P_t, to total prison cells in t, K_t.

$$C_t = f(S_t) + \epsilon_t^1 \tag{6a}$$

$$S_t = h(C_t, U_t) + \epsilon_t^2 \tag{6b}$$

where:

P_t = the prison population in period t
K_t = prison cell capacity in period t
$U_t = P_t/K_t$

As before, SES variables are omitted for expositional convenience. To our knowledge, no deterrence investigation has included U_t in the equation for sanctions. The rationale for its inclusion again involves a resource utilization argument and, indeed, this model can be taken as a simple example in which the resource saturation hypothesis is made explicit. As prisons become increasingly crowded, pressure will be exerted to reduce utilization, U_t. In the short term (*e.g.*, a year) this reduction can only be accomplished through a reduction in prison population, P_t, since expansion of existing cell capacity, K_t, would require considerably more time.[23]

One recent example of this effect of resource saturation at work is Federal Judge Frank Johnson's order to the Alabama Corrections Department to release a sufficient number of prisoners to alleviate prison overcrowding (see *Criminal Justice Bulletin* 1976). Judge Johnson's order resulted in the reduction of both the probability of imprisonment given conviction and time served given imprisonment.

In this single-sanction and single-crime-type model with only two endogenous variables, identification of the crime function requires that one restriction be imposed; the absence of U_t, prison cell utilization in t, from eq. (6a) provides the necessary restriction. To see this, consider a log-linear specification of eqs. (6a-b):

$$\ln C_t = B_0 + B_1 \ln S_t + \epsilon_t^1 \tag{6a'}$$

$$\begin{aligned}\ln S_t &= \gamma_0 + \gamma_1 \ln C_t + \gamma_2 \ln\left(\frac{P_t}{K_t}\right) + \epsilon_t^2 \\ &= \gamma_0 + \gamma_1 \ln C_t + \gamma_2 \ln P_t - \gamma_2 \ln K_t + \epsilon_t^2 \end{aligned} \tag{6b'}$$

In addition, if we specifically define S_t to be the probability of imprisonment given a crime and assume that an imprisoned individual is incarcerated for a single period,[24] P_t will be:

[23] To the degree that crime does influence K_t by leading to more prison cell construction, that effect is longer-term, perhaps 5 to 10 years.

[24] This model is clearly an oversimplification. In general, prison terms are often considerably longer than a year, so that the prison population is not solely a function of the current values of C_t, S_t, and N_t but also depends on past incarcerations. This makes no essential difference to the points under discussion, however, save that past incarcerations could be used as an omitted predetermined variable in identifying the crime function under the assumption of no serial correlation.

$$P_t = C_t S_t N_t \tag{6c}$$

$$\ln P_t = \ln C_t + \ln S_t + \ln N_t \tag{6c'}$$

where:

N_t = total population in t[25]

Substituting eq. (6c') in eq. (6b') and rearranging terms:

$$\ln S_t = \frac{\gamma_0}{1-\gamma_2} + \frac{\gamma_1+\gamma_2}{1-\gamma_2}\ln C_t + \frac{\gamma_2}{1-\gamma_2}\ln (N_t/K_t) + \frac{\epsilon_t^2}{1-\gamma_2} \tag{6b''}$$

The exclusion of $\ln (N_t/K_t)$ from eq. (6a') provides the necessary restriction for identification.[26]

The validity of this identification procedure hinges upon the assumption that U_t does not directly affect crime. This assumption will fail if potential criminals have information on crowding in prisons and view the level of U_t as a partial measure of the severity of punishment. If, indeed, U_t has such an effect, then it should be included in the crime equation and the exclusion of N_t/K_t cannot be used validly to identify the crime function.

3. *Inertia Model: Lagged Sanctions*

In the system shown below, the equation for S_t includes S_{t-1}. Its inclusion could be argued on the grounds that sanctioning practice, being bound by tradition, will adjust slowly to changes in the crime rate or indeed to any other factors influencing sanctions. As a result, S_t will be

[25]The variable N_t is entered because C_t is expressed as crimes per capita, while P_t is the total number of prisoners.

[26]It might appear that we might separate $\ln (N_t/K_t)$ into two variables by writing $\ln (N_t/K_t) = \ln N_t - \ln K_t$ and then use the exclusion of both $\ln K_t$ and $\ln N_t$ from the crime equation to achieve not merely identification but over-identification. This apparent achieving of something for nothing does not succeed, however. Perhaps the easiest way to see this is to observe that the restrictions stating that both $\ln K_t$ and $\ln N_t$ do not appear in the crime equation can be written equivalently as the restriction that the coefficient of $\ln K_t$ in that equation is zero plus the restriction that the coefficient in that equation of $(-\ln N_t)$ is equal to that of $\ln K_t$. This latter restriction, however, is also satisfied in the sanctions equation and hence does not help at all in telling the two apart; if we used only that restriction we would not have identification. (This is an example of the fact referred to in a previous footnote that counting restrictions provides only a necessary, not a sufficient condition for identification.) To put it another way, $\ln K_t$ and $\ln N_t$ do not independently affect $\ln C_t$ and $\ln S_t$. There is only one piece of useful information to be gained from using them, not two.

influenced by sanctions in prior periods, assumed for illustration to be represented sufficiently by S_{t-1}. Since S_{t-1} does not appear in the crime equation, the crime function is identified with some assumption on the nature of the serial dependence of the ϵ_t^i.

$$C_t = f(S_t) + \epsilon_t^1 \tag{7a}$$

$$S_t = h(C_t, S_{t-1}) + \epsilon_t^2 \tag{7b}$$

While this rationale for including S_{t-1} in the specification of S_t is highly plausible, it is not plausible at the same time to exclude S_{t-1} from the crime equation. To do so assumes that potential criminals are not influenced by sanctions in prior periods. Such an assumption has little plausibility as a crucial identifying restriction, since it implies that historical sanction levels have no influence on perceptions of current sanctions even though they do influence current sanctions themselves.

For example, suppose a rational criminal has information indicating that a certain offense was not being prosecuted as vigorously as it had been previously. Should he disregard his information on sanction levels in prior periods and base his decision solely upon the new information on sanctions? There are several reasons that a rational criminal might still continue to consider prior information on sanctions.

First, unlike stock market prices, daily quotations of sanction levels are not available and the information that is available derives from very uncertain sources, including the criminal's own experience, the experience of his criminal peers, news reports, or even the published statistics utilized by deterrence researchers. When current information is poor, considering information from the past, even if it is also uncertain, is very sensible in making estimates of the current status.

Second, even if current information on a variable is good, information on prior levels provides important information on the stability or trend of the sanction over time. If, for example, potential criminals are not risk neutral, then they will want information on the distribution of potential sanctions. Prior periods may provide such useful information. Moreover, past information on sanctions may provide useful information on trends in sanctions that may also be of value to a rational criminal.

In view of the implausibility of assuming that S_{t-1} affects S_t but not C_t, we do not believe that identification can be validly achieved in this way.

B. A SINGLE-CRIME-TYPE, MULTIPLE-SANCTION MODEL

Our focus has been on simple models in which only a single sanction and single crime type are simultaneously related. We now turn to a model in which a single crime type is simultaneously related to several sanction types.

In this model we attempt to capture some of the interrelationships between crime and the CJS subsystems—police, courts, and corrections. These interrelationships derive from a model of the CJS put forward by Blumstein and Larson (1969) that characterizes the CJS as a flow process. A very simplified version of their conceptualization is shown in Figure 11.

Society generates crime, which is an input into the first of the pictured subsystems—the police. The police arrest suspects, some of whom are charged, while others are subsequently released without charge. The charged individuals are inputs to the courts subsystem. The courts in turn adjudicate the charges and some of those charged

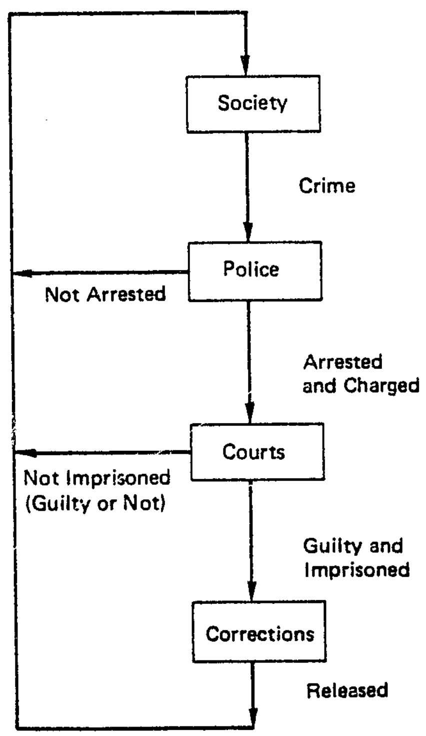


FIGURE 11 A simplified flow model of the criminal justice system.

are found guilty and imprisoned and turned over to the corrections subsystem. Others are not imprisoned, either because the charges do not lead to indictment or, if indicted, the indictment is dismissed or the defendant is acquitted—or, possibly, the defendant is convicted but not imprisoned. Finally, those individuals who are imprisoned are subsequently released to society either on parole or after having served their sentence.

The actions of each of the subsystems have implications for the possible penalties confronting a potential criminal; similarly, the amount of crime in the society has implications for the magnitudes of the flows through the subsystems.

In the models to be discussed, we attempt to capture these interrelationships between crimes and sanctions. Let us introduce the following notation:

C_t = total crimes in t
P_t^A = probability of apprehension and charge given a crime in t
$P_t^{G/A}$ = probability of conviction given charge in t
$P_t^{I/G}$ = probability of imprisonment given conviction in t
T_t = time served in period t
E_t^{Po} = police expenditures in t
E_t^{J} = judicial expenditures in t
E_t^{Pr} = prison expenditures in t
A_t = number of charges in t
G_t = number of convictions in t
I_t = number of imprisonments in t
U_t = prison utilization in period t
$\mu_t, \epsilon_t^i, \nu_t^i$ = random disturbances

$$C_t = f(P_t^A, P_t^{G/A}, P_t^{I/G}, T_t) + \mu_t \tag{8a}$$

$$P_t^A = g_1(E_t^{Po}, C_t) + \epsilon_t^1 \tag{8b}$$

$$P_t^{G/A} = g_2(E_t^J, A_t) + \epsilon_t^2 \tag{8c}$$

since $A_t = P_t^A C_t$ (ignoring sampling variation)

$$P_t^{G/A} = g_2(E_t^J, P_t^A C_t) + \epsilon_t^2 \tag{8c'}$$

$$P_t^{I/G} = g_3(E_t^{Pr}, G_t, U_t) + \epsilon_t^3 \tag{8d}$$

since $G_t = P_t^{G/A}\, P_t^A C_t$

$$P_t^{I|G} = g_3(E_t^J, P_t^{G|A}P_t^AC_t,U_t) + \epsilon_t^3 \quad (8d')$$

$$T_t = g_4(E_t^{Pr}, U_t) + \epsilon_t^4 \quad (8e)$$

$$E_t^{Po} = h_1(C_{t-1}, E_{t-1}^{Po}) + \nu_t^1 \quad (8f)$$

$$E_t^J = h_2(A_{t-1}, E_{t-1}^J) + \nu_t^2 \quad (8g)$$

$$E_t^{Pr} = h_3(U_{t-1}, E_{t-1}^{Pr}) + \nu_t^3 \quad (8h)$$

A crucial feature of this model is the distinction among the different types of sanctions. By differentiating among such sanctions as the probability of apprehension and charge, the probability of conviction given charge, the probability of imprisonment given conviction, and time served given imprisonment, the effect of each type of sanction can, at least theoretically, be measured. Different categories of sanctions are possible or greater refinement in the number of sanction types could be made. The crucial point, however, is that, *a priori*, there are good reasons for believing that the magnitude of the deterrent effect associated with each sanction type may be different. For example, the disutility of a conviction given charge is likely to be greater than the disutility associated with charge, since the stigma of conviction is greater than that associated with only being charged.

The likelihood of differential deterrent effects associated with different sanctions has both important technical implications for estimation and significant policy implications. For the purpose of estimation, if two types of sanctions, for example P^A and $P^{G|A}$, have different effects, then it is inappropriate to estimate a single parameter for the conglomerate effect of $P^G = P^AP^{G|A}$. Further, from a policy perspective, we would not want to aggregate the two, since it may be useful to know the relative magnitudes of the separate effects. By comparing effects with costs, we can determine where resources should be allocated. If, for example, identical increases in expenditures on police and courts would achieve the same percent increase in P_t^A and $P_t^{G|A}$, respectively, then crime reduction would be pursued more efficiently by allocating the additional expenditures to the sanction with the larger deterrent effect.

The second crucial feature of the system, which has significant implications for estimation, is the simultaneous relationship of C_t with each of the sanction variables, due perhaps to resource saturation considerations. Thus, given police resources, E_t^{Po} (which are themselves affected by the number of past crimes), the probability of arrest, P_t^A,

depends on the current number of crimes, C_t, facing the police.[27] Further, although C_t only affects P_t^A directly, the levels of C_t also affect the workload of the courts and corrections subsystems "downstream" from the police. The probability of conviction given charge, $P_t^{G|A}$, is likely to be affected by the workload of the courts, A_t, but A_t will be determined by the product of C_t and P_t^A. Since C_t is also hypothesized to be affected by $P_t^{G|A}$, $P_t^{G|A}$ and C_t will be simultaneously related.

Similarly, the probability of imprisonment given conviction, $P_t^{I|G}$ is affected by G_t, the number of convictions in t. Since G_t is the product of C_t, P_t^A, and $P_t^{G|A}$, $P_t^{I|G}$ is simultaneously related to C_t. Time served, T_t, and $P_t^{I|G}$ are also hypothesized to be affected by the utilization of prison capacity, U_t, because we expect utilization to have its predominant effect on judges and parole boards who most directly control the size of the prison population. Since U_t is affected by the size of the prison population, which is just the number of currently imprisoned criminals (and thus depends on C_t, P_t^A, $P_t^{G|A}$, and $P_t^{I|G}$), T_t will also be simultaneously related to C_t.

As the model is specified, none of the sanctions is in a direct simultaneous relationship with any other (e.g., P_t^A directly affects $P_t^{G|A}$, but $P_t^{G|A}$ does not directly affect P_t^A). In terms of the problem of identifying the crime function, the validity of this assumption about the interrelationship among the sanctions is not relevant; the model could be generalized to allow such direct simultaneous relationships without altering our conclusion about the identifiability of the crime function (8a).

The crime rate, C_t, is determined by four sanction variables, all of which are presumed to be simultaneously related to C_t. Therefore, at least four independent restrictions are necessary to identify the crime function. Four such restrictions are provided by the exclusion of E_t^{Po}, E_t^J, E_t^{Pr}, and U_t (prison cell utilization).

The requirements for plausibly using these restrictions to identify the crime function have already been discussed. The key issues are worth restating. Since the expenditures variables are predetermined rather than exogenous [eqs. (8f)-(8g)], it is dangerous to assume no serial correlation in the ϵ_t^i. Some more general assumptions about the nature of that serial dependence are necessary; whatever the explicit assumption, data with a time-series component will be needed. The restrictions involving the exclusion of the police expenditure variable, E_t^{Po},

[27] In earlier sections, C_t was crimes per capita. Defining C_t as total crime instead of the crime rate would not affect our conclusion for this model; all state variables to be discussed, A_t, G_t, E_t^{Po}, E_t^J and E_t^{Pr} could be normalized by total population and thereby be redefined as rates.

crimes, C_t, facing the police.[27]
$_t^A$ directly, the levels of C_t also
s and corrections subsystems
probability of conviction given
y the workload of the courts, A_t,
uct of C_t and P_t^A. Since C_t is also
$^{G|A}$ and C_t will be simultaneously

onment given conviction, $P_t^{I|G}$ is
ions in t. Since G_t is the product
usly related to C_t. Time served,
be affected by the utilization of
t utilization to have its predomi-
ds who most directly control the
U_t is affected by the size of the
number of currently imprisoned
$P_t^{G|A}$, and $P_t^{I|G}$), T_t will also be

he sanctions is in a direct simul-
.g., P_t^A directly affects $P_t^{G|A}$, but
ms of the problem of identifying
assumption about the interrela-
evant; the model could be gener-
us relationships without altering
of the crime function (8a).
y four sanction variables, all of
usly related to C_t. Therefore, at
necessary to identify the crime
rovided by the exclusion of E_t^{Po},

these restrictions to identify the
ussed. The key issues are worth
iables are predetermined rather
dangerous to assume no serial
al assumptions about the nature
y; whatever the explicit assump-
nt will be needed. The restric-
olice expenditure variable, E_t^{Po},

Defining C_t as total crime instead of the
r this model; all state variables to be
lized by total population and thereby be

and U_t are particularly vulnerable to criticism, since being a measure of the intensity of the police presence in the community and the severity of punishment, respectively, it can be argued convincingly that each should also be included in the crime function. However, since the four restrictions are just-identifying and thereby necessary for estimation, we cannot test the validity of the restrictions involving E_t^{Po} and U_t, even assuming away the serial correlation problem just discussed.

In this multiple-sanction model, identification of the crime function requires the joint use of both the expenditures and cell-capacity identification restrictions, whereas in the one-sanction models, either one was sufficient to just-identify. The necessity of using both categories of restrictions to identify the crime function points to still another problem. As the number of endogenous sanctions increases, the difficulties in identifying the crime functions increase also. In the context of a multiple-crime-type model, which will be discussed next, this difficulty can become fatal to identification.

C. A MULTIPLE-CRIME-TYPE, SINGLE-SANCTION MODEL

Our discussion thus far has been limited to the consideration of single-crime-type models. We now consider the problem of identifying each of the crime equations in a multiple-crime-type model. A multiple-crime-type formulation is of interest because each crime type will incrementally impact a single set of CJS resources. An examination of their joint effect has important implications for identification.

A two-crime-type, single-sanction characterization of such a phenomenon is given below, along with the model's equivalent flow diagram, in Figure 12.

$$C_t^1 = f_1(S_t^1) + \epsilon_t^1 \quad (9a)$$

$$C_t^2 = f_2(S_t^2) + \epsilon_t^2 \quad (9b)$$

$$S_t^1 = g^1(E_t, C_t^1, C_t^2, S_t^2) + \epsilon_t^3 \quad (9c)$$

$$S_t^2 = g^2(E_t, C_t^1, C_t^2, S_t^1) + \epsilon_t^4 \quad (9d)$$

$$E_t = h(E_{t-1}, C_{t-1}^1, C_{t-1}^2) + \epsilon_t^5 \quad (9e)$$

where:

C_t^i = crimes of type i per capita in t

S_t^i = sanctions per crime of type i in t

E_t = CJS expenditures in t.

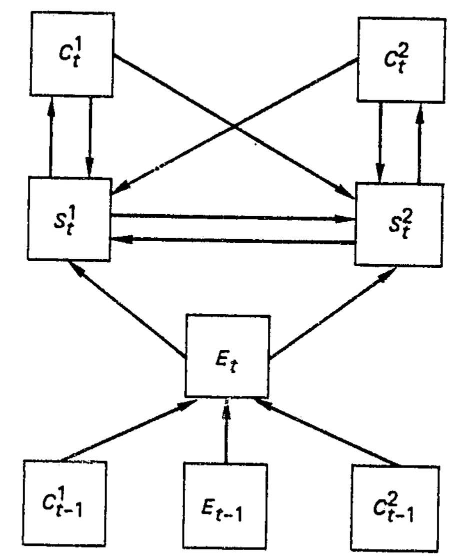


FIGURE 12 Flow diagram of multiple-crime-type, single-sanction model.

As indicated by eqs. (9c) and (9d), S_t^i is a function of total resources available to the CJS (E_t), the demands placed on these resources by each of the crime inputs (C_t^i, $i = 1,2$), and the level of the sanction imposed for the other crime type. The resource saturation theory would predict that increases in E_t would act to increase S_t^i ($\partial g^i/\partial E_t > 0$), increases in the prevalence of either crime type would act to reduce S_t^i ($\partial g^i/\partial C_t^j < 0$, $j = 1,2$) and increases in S_t^j would decrease S_t^i, $i \neq j$ ($\partial g^i/\partial S_t^j < 0$) because the additional resources required to increase S_t^j would be drawn from those used to maintain S_t^i.

Alternative theories of the effects of crime on sanctions might make different predictions, but the crucial point is that sanctions for each crime type are influenced by the level of both types of crime, because each crime type impacts the common set of CJS resources.

Considering eqs. (9a)-(9d) as the simultaneous system and treating E_t as predetermined by eq. (9e), the number of endogenous variables, M, is 4. Hence, at least three restrictions are necessary for the identification of each crime function. One such restriction is provided by the exclusion of E_t from eqs. (9a) and (9b) under assumptions outlined previously. A second is provided by the assumption that crime of one type has no direct effect on crime of the other type. The final restriction necessary for identification of each crime function, however, rests additionally upon the assumption that sanctions for one crime type do not influence the level of crime for the other crime type (e.g., S_t^1 does not

affect C_t^2). In the context o
glary), the possibility of suc
indeed consistent with the t
hypothesis—namely, that t
If such cross-effects exist

C_t^1

C_t^2

These more general versi
identified; there are now
Since estimation requires th
tions on each crime equatio
tional restriction be impose
sanction model, the prison
imposed.

This, however, is really o
problem in a multi-crime-t
crime type (e.g., robbery) w
affected by S_t^1, S_t^2, and S_t^3
variables (M) by two (C_t^3
restrictions on each crime
directly affect C_t^1 or C_t^2). T
identified case of M = 4 w
only four restrictions, and i
cation of the crime functio
seems even more difficult t

The difficulties in findin
acute when multiple sancti
example, S_t^i were divided i
single-crime-type, multiple-
of the three crime types all
the number of endogenous
tion restrictions would be r
tions, in addition to the a
crime appears in each such

In general, a model with
require $n \times m$ non-automa
tions. Hybrid versions of
restrictions. For example,
effects only exist among su

affect C_t^2). In the context of property crimes, (e.g., larceny and burglary), the possibility of such a cross-effect is quite conceivable and is indeed consistent with the basic principle that underlies the deterrence hypothesis—namely, that behavior is influenced by incentives.

If such cross-effects exist, then the two crime functions become:

$$C_t^1 = f^1(S_t^1, S_t^2) + \epsilon_t^1 \tag{9a'}$$

$$C_t^2 = f^2(S_t^1, S_t^2) + \epsilon_t^2 \tag{9b'}$$

These more general versions of the two crime functions are no longer identified; there are now only two, not three restrictions on them. Since estimation requires the imposition of three identification restrictions on each crime equation, identification would require that an additional restriction be imposed. For this simple two-crime-type, single-sanction model, the prison cell utilization identification might also be imposed.

This, however, is really only an illusory solution to the identification problem in a multi-crime-type setting. The addition of still another crime type (e.g., robbery) with S_t^3 affecting C_t^1, C_t^2 and C_t^3, and C_t^3 being affected by S_t^1, S_t^2, and S_t^3 would increase the number of endogenous variables *(M)* by two (C_t^3 and S_t^3) but would increase the number of restrictions on each crime equation by only one (because C_t^3 does not directly affect C_t^1 or C_t^2). Hence we would have moved from a just-identified case of $M = 4$ with three restrictions to one of $M = 6$ with only four restrictions, and identification would fail. In general, identification of the crime functions in a multi-crime, single-sanction model seems even more difficult than in the single-crime-type case.

The difficulties in finding sufficient restrictions become even more acute when multiple sanctions are introduced into the model. If, for example, S_t^i were divided into the four sanction types discussed in the single-crime-type, multiple-sanctions model and the sanctions for each of the three crime types all had cross-effects on the other crime types, the number of endogenous variables would be 15. Thus, 12 identification restrictions would be required to estimate each of the crime functions, in addition to the automatic restrictions that only one type of crime appears in each such function.

In general, a model with *n* crime types and *m* sanction types would require $n \times m$ non-automatic restrictions to identify the crime functions. Hybrid versions of the model would require fewer additional restrictions. For example, one might plausibly assume that cross-effects only exist among subsets of crime types (perhaps distinguishing

between property and violent crimes). From a practical perspective, however, such an approach offers little help since, for example, even a two-sanction model for the four index property crimes (i.e., robbery, burglary, larceny, and auto theft) would require eight non-automatic restrictions to identify each of the separate crime functions.

In view of the difficulty in generating plausible restrictions, the estimation of the generalized multi-crime-type, multi-sanction model including cross-effects of the sanctions does not appear feasible. To the extent that the generalized model is viewed as the only plausible characterization of the simultaneous association between crime and sanctions, an argument as to the impossibility of valid identification is even more compelling than in the case of the simplified models discussed earlier.

The apparent infeasibility of identifying the generalized model hinges upon the assumption that the sanctions for C_t^i directly affect C_t^j. It may be that such cross-effects are, at most, very weak. The difficulty is that, using aggregate, non-experimental data, we cannot test for this. Moreover, a model estimated simply assuming no cross-effects would always remain suspect for having assumed that cross-effects are not operating.

V. CONCLUSION

Identification is the *sine qua non* of all estimation and especially of simultaneous equation estimation. It establishes the feasibility of determining the structure of a system from the data generated by that system. Without identification, estimation is logically impossible.

Researchers who have employed simultaneous estimation techniques to study the deterrent effect of sanctions on crime have failed to recognize fully the importance of this issue. The restrictions that they (implicitly or explicitly) use to gain apparent identification have little theoretical or empirical basis.

In this paper we have examined a variety of plausible approaches to the identification of the crime functions in a system in which crime rates and sanction levels are simultaneously related. Our conclusions with regard to the feasibility of identification, while not wholly negative, are certainly soberly cautious. In particular, it appears very doubtful that work using only aggregate cross-sectional data can ever succeed in identifying and consistently estimating the deterrent effect of punishment on crime. If we are to know that effect and, particularly, if we are to rely on that knowledge for policy purposes, that knowledge must come from analyses of a different sort. In particular, analyses

). From a practical perspective, e help since, for example, even a ‹ property crimes (i.e., robbery, ıuld require eight non-automatic arate crime functions.

ıg plausible restrictions, the esti- ɛ-type, multi-sanction model in- does not appear feasible. To the ; viewed as the only plausible association between crime and ssibility of valid identification is se of the simplified models dis-

ing the generalized model hinges s for C_t^i directly affect C_t^j. It may st, very weak. The difficulty is al data, we cannot test for this. ıssuming no cross-effects would umed that cross-effects are not

all estimation and especially of ɛstablishes the feasibility of de- ·om the data generated by that :ion is logically impossible.

simultaneous estimation tech- anctions on crime have failed to issue. The restrictions that they ›parent identification have little

riety of plausible approaches to ns in a system in which crime ɛously related. Our conclusions ication, while not wholly nega- In particular, it appears very e cross-sectional data can ever estimating the deterrent effect ow that effect and, particularly, ›olicy purposes, that knowledge nt sort. In particular, analyses

using aggregate non-experimental data must have a time-series component in the data (i.e., pure time-series or a time-series, cross-section), and the estimation procedures must account for the possibility of serial correlation in the stochastic components of the specification.

TECHNICAL NOTE: LAGGED MODELS OF THE MUTUAL ASSOCIATION OF CRIME AND SANCTIONS

The principal focus of this paper is the estimability of simultaneous models of crime and sanctions. In a simultaneous formulation, the mutual interaction is assumed to occur contemporaneously during the period of observation. For an observation period of a given length, a necessary requirement for a phenomenon to be simultaneous is that the impact of the actions taken by the system's actors (e.g., criminals and the CJS) be transmitted sufficiently fast so that each actor can react to the actions of the other actors within the observation period. Thus, a critical parameter is the length of the observation period. If the period is sufficiently short, then any mutual association can be modeled as non-simultaneous, whereas, if the period is sufficiently long, all such associations can be made simultaneous. In the context of the mutual association of crimes and sanctions, in which observations are generally made annually, the association is simultaneous if within a 1-year period potential criminals receive cues on the current level of sanctions being delivered by the CJS and if the level of crime in the current period also works to influence the sanctions delivered by the CJS.

If information does not flow this quickly, an alternative characterization of the mutual association involves lags. In the single-crime-type, single-sanction model, such a characterization could take the form

$$C_t = a + bS_{t-1} + \epsilon_t \tag{10a}$$

$$S_t = c + dC_{t-1} + \mu_t \tag{10b}$$

If the parameters of this model are to be estimated consistently by regression, the disturbances, ϵ_t and μ_t, must not be serially correlated.[28]

[28]The parameters of one of the equations could be consistently estimated if there is not serial correlation in that equation's disturbance. In general, however, if ϵ_t and μ_t are correlated either with their own past values or with each others' past values, consistency will not be present. In such general cases, the covariances of S_{t-1} and C_{t-1} with ϵ_t and μ_t will be complex expressions involving both the serial correlation behavior of ϵ_t and μ_t and their covariance.

In our prior discussion, we elaborated upon the reasons for believing that there is, in fact, serial correlation. Hence, we would have very little confidence in any causal inferences drawn from parameter estimates that are generated by ordinary least squares.

Our pessimism about using regression is reinforced by the fact that in the simplest case, where there is only serial correlation in ϵ_t, the serial correlation will result in an over-estimate of the deterrent effect of sanctions. Suppose that ϵ_t follows a first-order autoregressive process with parameter ρ. Let σ^2 denote the variance of ϵ_t. Additionally, assume that $d < 0$ (i.e., increases in C_{t-1} decrease S_t). Under these plausible conditions, if $\epsilon_{t-2} > 0$, then C_{t-2} will be larger than predicted by the structural component of eq. (10a). This larger-than-predicted value of C_{t-2} will drive down the value of S_{t-1}, since $d < 0$. In addition, since $\epsilon_{t-2} > 0$, ϵ_t will tend to be positive because cov $(\epsilon_t, \epsilon_{t-2}) = \rho^2\sigma^2 > 0$. With $\epsilon_t > 0$, C_t would be larger than that predicted by the structural component of eq. (10a). We would then observe large values of C_t being associated with small values of S_{t-1}, even if $b = 0$. This negative association, however, would drive the estimate of b to a negative value.

Attempts to analyze models of the type given by eqs. (10a) and (10b) have been limited to the sociological literature on deterrence (Logan 1975, and Tittle and Rowe 1974). In these analyses, S_t is defined as arrests per crime. Tittle and Rowe found a negative and often significant path coefficient between S_{t-1} and C_t, a result that is consistent with the deterrence hypothesis, while Logan found no such association.

The path coefficient estimate of the association between S_{t-1} and C_t is estimated in a way that is analytically equivalent to regression estimation of b in the model shown in eq. (10a). Therefore, these path coefficients suffer from all the ambiguities that we have discussed.

Models in which the mutual association between crime and sanctions occurs with a lag, however, are attractive because they offer an intuitively attractive characterization of this mutual association. Information on the sanctioning behavior of the CJS is probably transmitted very slowly through the kinds of cues that have been discussed. An assumption that information lag on sanctions is greater than a year may, therefore, be plausible in most instances.[29] Under such an assumption that C_t is a function of sanctions in prior periods, we could maintain the assumption that C_t affects S_t [e.g., C_{t-1} is replaced by C_t in eq. (10b)],

[29]In specific instances where official statements are published announcing changes in sanctioning practice (e.g., the case in which the District Attorney of San Francisco announced that prostitution would no longer be prosecuted), the assumption of a 1-year lag would be untenable.

and the model would remai be a catch. For such a mode regression, there not only m μ_t must be uncorrelated.

Thus, whatever the specif structure, estimation must u bility of serial correlation a terms if the estimated coeff estimate of the causal effect

REFERENCES

Avio, K., and Clarke, S. (1974) P Prepared for the Ontario Econo
Blumstein, A., and Cohen, J. (1973 *Criminal Law and Criminology*
Blumstein, A., and Larson, R. (196 *tions Research* 17(2):199-232.
Blumstein, A., Cohen, J., and Nagi ment process. *Journal of Crimin*
Carr-Hill, R. A., and Stern, H. H. (of recorded offenses in England 318.
Criminal Justice Bulletin (1976) Al Editorial Services, Cincinnati.
Ehrlich, I. (1970) Participation in Il toral dissertation, Columbia Uni
Ehrlich, I. (1973) Participation in ille tigation. *Journal of Political Eco*
Ehrlich, I. (1975) The deterrent effec *American Economic Review* 65(
Fisher, F. M. (1961) On the cost of estimation. *Econometrica* 29(2):
Fisher, F. M. (1966) *The Identificati* Hill Book Company.
Gibbs, J. B. (1968) Crime, punishn *Quarterly* 48(4):515-30.
Gibbs, J. B. (1975) *Crime, Punishme*
Greenberg, D. (1977) Crime, deterre *Modeling the Criminal Justice S*
Logan, C. (1975) Arrest rates and de
McPheters, L., and Stronge, W. B crime. *National Tax Journal* 27(4
Sjoquist, D. (1973) Property crime *American Economic Review* 83(
Tittle, C. (1969) Crime rates and lega
Tittle, C., and Rowe, A. (1974) Cert deterrence hypothesis. *Social Fo*

the reasons for believing
:e, we would have very
wn from parameter esti-
'ares.
ıforced by the fact that in
ɔrrelation in ϵ_t, the serial
f the deterrent effect of
r autoregressive process
nce of ϵ_t. Additionally,
:crease S_t). Under these
· be larger than predicted
his larger-than-predicted
. since $d < 0$. In addition,
because cov $(\epsilon_t, \epsilon_{t-2})$ =
an that predicted by the
:hen observe large values
S_{t-1}, even if $b = 0$. This
ıe estimate of b to a nega-

:n by eqs. (10a) and (10b)
re on deterrence (Logan
nalyses, S_t is defined as
egative and often signifi-
result that is consistent
t found no such associa-

ıtion between S_{t-1} and C_t
valent to regression esti-
). Therefore, these path
ıt we have discussed.
ween crime and sanctions
:ause they offer an intui-
ıal association. Informa-
ırobably transmitted very
:n discussed. An assump-
:r than a year may, there-
such an assumption that
s, we could maintain the
ılaced by C_t in eq. (10b)],

blished announcing changes in
ict Attorney of San Francisco
:ed), the assumption of a 1-year

and the model would remain non-simultaneous—but there would still be a catch. For such a model to be consistently estimated by ordinary regression, there not only must be no serial correlation, but also ϵ_t and μ_t must be uncorrelated.

Thus, whatever the specific nature of the model employing a lagged structure, estimation must use methodologies that allow for the possibility of serial correlation and non-zero covariance in the stochastic terms if the estimated coefficients are to be plausibly regarded as an estimate of the causal effect of sanctions on crime.

REFERENCES

Avio, K., and Clarke, S. (1974) Property Crime in Canada: An Econometric Study. Prepared for the Ontario Economic Council.

Blumstein, A., and Cohen, J. (1973) A theory of the stability of punishment. *Journal of Criminal Law and Criminology* 64(2):198-207.

Blumstein, A., and Larson, R. (1969) Models of a total criminal justice system. *Operations Research* 17(2):199-232.

Blumstein, A., Cohen, J., and Nagin, D. (1976) The dynamics of a homeostatic punishment process. *Journal of Criminal Law and Criminology* 67(3):317-34.

Carr-Hill, R. A., and Stern, H. H. (1973) An econometric model of the supply and control of recorded offenses in England and Wales. *Journal of Public Economics* 2(4):289-318.

Criminal Justice Bulletin (1976) Alabama Prison. Newsletter published by Anderson Editorial Services, Cincinnati.

Ehrlich, I. (1970) Participation in Illegitimate Activities: An Economic Analysis. Doctoral dissertation, Columbia University.

Ehrlich, I. (1973) Participation in illegitimate activities: a theoretical and empirical investigation. *Journal of Political Economy* 81(3):521-67.

Ehrlich, I. (1975) The deterrent effect of capital punishment: a question of life and death. *American Economic Review* 65(3):397-417.

Fisher, F. M. (1961) On the cost of approximate specification in simultaneous equation estimation. *Econometrica* 29(2):139-70.

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

Gibbs, J. B. (1968) Crime, punishment and deterrence. *Southwestern Social Science Quarterly* 48(4):515-30.

Gibbs, J. B. (1975) *Crime, Punishment and Deterrence*. New York: American Elsevier.

Greenberg, D. (1977) Crime, deterrence research, and social policy. In Stuart Nagel, ed., *Modeling the Criminal Justice System*. Los Angeles: Sage Publications.

Logan, C. (1975) Arrest rates and deterrence. *Social Science Quarterly* 56(3):376-89.

McPheters, L., and Stronge, W. B. (1974) Law enforcement expenditures and urban crime. *National Tax Journal* 27(4):633-44.

Sjoquist, D. (1973) Property crime and economic behavior: some empirical results. *American Economic Review* 83(3):439-46.

Tittle, C. (1969) Crime rates and legal sanctions. *Social Problems* 16(Spring):408-28.

Tittle, C., and Rowe, A. (1974) Certainty of arrest and crime rates: a further test of the deterrence hypothesis. *Social Forces* 52(4):455-62.

END