DETECTION CONTROLLED ESTIMATION*

JONATHAN S. FEINSTEIN
Stanford University

I. Introduction

It is widely acknowledged that many violations of laws and regulations remain undetected. This problem of incomplete detection seriously complicates statistical analysis of the factors associated with noncompliance since the data record only detected violations, which typically are not representative of all violations. The difficulties incomplete detection poses for statistical analysis have been recognized for many years. Thus Lambert Quételet, the father of the statistical analysis of crime rates, wrote in 1815 that "our observations can only refer to a certain number of crimes known and adjudicated, compared to a total number of unknown crimes which have been committed."[1]

Nonetheless, the empirical literature has largely ignored the problem. For example, one common practice has been to specify models that refer to all violations and then substitute a measure of detected violations into the data analysis as a proxy variable; recent examples include Clotfelter's analysis of income tax evasion, Witte's study of recidivism, and McCormick and Tollison's account of sports fouls.[2]

This article presents a general statistical methodology for addressing the problem of incomplete detection, organized around the principle of

* I acknowledge graduate student support from the National Science Foundation. Ed Keating provided outstanding research assistance. I am grateful to my advisors Franklin Fisher and Daniel McFadden for their advice and encouragement and to Craig Alexander, Alessandra Casella, Hide Ichimura, Jeremy Stein, Glen Sueyoshi, Leslie Sundt, Salvador Valdes-Prieto, the referee, and the editor for comments.

[1] Quoted by Thorsten Sellin & Marvin Wolfgang, The Measurement of Delinquency (1978).

[2] Charles Clotfelter, Tax Evasion and Tax Rates: An Analysis of Individual Returns, 65 Rev. Econ. & Stat. 363 (1983); Ann Witte, Estimating the Economic Model of Crime with Individual Data, 94 Q. J. Econ. 57 (1980); Robert McCormick & Robert Tollison, Crime on the Court, 92 J. Pol. Econ. 223 (1984).

[*Journal of Law & Economics*, vol. XXXIII (April 1990)]
 0022-2186/90/3301-0003$01.50

incorporating the detection process into the statistical analysis of violations data; to reflect this principle, the method is referred to as detection controlled estimation (DCE). I develop the simplest version of DCE, extend it to more complex settings, and discuss its statistical properties. As an example of its application, I also present a case study of Occupational Safety and Health Administration (OSHA) safety regulation.

The logic of detection controlled estimation is simple. Data are collected that pair each potential offender to the individual or agency responsible for monitoring his or her behavior. Two equations are then specified, one referring to the probability of a violation, the other to the likelihood of detection. These equations in turn lead to statistical procedures that recognize that, whenever a violation is detected, two events have occurred in succession: the potential offender has committed a violation, and, subsequently, a monitor assigned to the potential offender has detected the violation. Conversely, the procedures recognize that when a potential offender is not detected in violation, there are two possible explanations: either the potential offender really is compliant, or the potential offender did commit a violation but it was never detected. Put slightly differently, the procedures take into account that a violation may be committed but not detected, in which case it will not be observed; only violations that are subsequently detected are observed. The data on violations themselves are "missing," and the maximum likelihood techniques that I propose belong to a class of missing information algorithms that have been discussed by Dempster, Laird, and Rubin, and Cox and Oakes.[3] Previous econometric studies most closely related to this article are by Poirier, who presents a similar model of partial observability, and McFadden, who discusses nested-choice models.[4] As Poirier points out, the most serious statistical issue that arises in these models is identification; therefore I investigate this topic in some detail, both in the text and in Appendix A. Application of missing information models to law and economics is, to my knowledge, new; in fact, the only article I am aware of that incorporates detection into a formal statistical analysis is by Epple and Visscher who collect data only on detected violations (off-shore oil spills).[5]

[3] A. P. Dempster, N. M. Laird, & D. B. Rubin, Maximum Likelihood from Incomplete Data via the EM Algorithm, 39 Royal Stat. Soc. J., Series B, 1 (1977); D. R. Cox & D. Oakes, Analysis of Survival Data (1984).

[4] Dale Poirier, Partial Observability in Bivariate Probit Models, 12 J. Econometrics 209 (1980); Daniel McFadden, Qualitative Response Models, in 2 Handbook of Econometrics, ch. 24 (Zvi Griliches & Michael Intriligator eds. 1984).

[5] Dennis Epple & Michael Visscher, Environmental Pollution: Modeling Occurrence, Detection, and Deterrence, 27 J. Law & Econ. 29 (1984).

From a different perspective, the detection controlled methodology parallels recent theoretical developments in the study of regulatory audits and income tax evasion.[6] These articles model noncompliance and detection as the two halves of a larger system that includes them both. As one would expect, they emphasize that the violation decision and detection effort are interdependent, a point to which I devote a good deal of attention in an appendix, but which requires further exploration in future work.

The detection controlled methodology is an improvement over earlier models in several respects. First, and most obviously, it produces consistent parameter estimates even when detection is incomplete and the probability of detection varies from case to case. In contrast, procedures that do not control for incomplete detection are misspecified and lead to biased parameter estimates and hypothesis tests. When the proportion of violations remaining undiscovered is substantial, these biases can also be expected to be substantial. In a number of important cases, theoretical arguments suggest that these biases tend systematically in one direction, so that ignoring them seriously jeopardizes the interpretation of results. Examples of such systematic biases include tests of the deterrence hypothesis, tests of the relationship between crime and unemployment, and assessments of the relative frequency with which different subgroups of the population commit violations. All of these biases are discussed more fully in the next section.

Second, explicitly modeling the detection process provides a means of analyzing many public-policy and public-management issues. To illustrate this point, consider regulation. In its simplest form, DCE computes a different detection rate for each regulatory inspector, providing a means of evaluating personnel. Because DCE simultaneously controls for variations in the attributes of regulated firms, its estimates of inspectors' relative abilities properly take into account the fact that different inspectors have been assigned to monitor different firms. Further, the estimator's use as a management tool extends beyond the evaluation of personnel to other aspects of the regulatory environment. Thus it might be used to examine the relationship between an agency's budget and its success in detecting violations. Here, too, it is an improvement over methods that use the detected violation rate as a proxy for the true rate because increased enforcement can be expected not only to raise detection rates but

[6] See, especially, David Baron & David Besanko, Regulation, Asymmetric Information, and Auditing, 15 Rand J. Econ. 447 (1984); Jean-Jacques Laffont & Jean Tirole, Using Cost Observation to Regulate Firms, 94 J. Pol. Econ. 614 (1986); and Michael Graetz, Jennifer Reinganum, & Louis Wilde, The Tax Compliance Game: Toward an Interactive Theory of Law Enforcement, 2 J. L. Econ., and Org. 1 (1986).

also to lower violation rates—hence estimates based on the detected violation rate understate the effect of improved enforcement.

Finally, DCE provides a means of estimating the proportion of violations remaining undetected, which is not directly observable. These estimates are of considerable interest but must be treated with caution, largely because of issues of identification discussed above and to which I return several times in the main text and appendices.

An example of an application of detection controlled methods is

1. *Income Tax Evasion.* The Internal Revenue Service (IRS) Taxpayer Compliance Measurement Program chooses a stratified random sample of approximately 50,000 tax returns every three years and examines each return. In a separate analysis, Alexander and Feinstein use this data to analyze evasion, paying special attention to detection controlled models that control for heterogeneity across IRS examiners.[7] Similar data exist on corporations.

Confining myself to broad categories, I include other potential applications of detection controlled methods:

2. *Accounting Audits.* In this category, I include audits of private-sector firms, government accounting audits, and Federal Reserve bank audits.

3. *Regulation.* Many agencies routinely inspect the firms they regulate; examples include OSHA, The Nuclear Regulatory Commission (NRC), Environmental Protection Agency (EPA), Federal Aviation Administration (FAA), Department of Agriculture, municipal building inspector departments, and fire and mine safety regulators. Data from these inspections typically list the firm inspected, the inspectors performing the inspection, any violations detected, and any sanctions imposed. In a companion article, I use such data to analyze the safety regulation of nuclear power plants.[8] Detection controlled methods are clearly relevant to improved management and policy formulation in this area.

4. *Street Crimes, Drug Trafficking, Illegal Immigration, White-Collar Crimes, and a Host of Lesser Violations Related, for Example, to Automobiles and Housing.* Applications to this area are less transparent than to the others, particularly applications to violent crimes for which the fact that a crime has been committed is often reported by the victim or the victim's family.

[7] Craig Alexander & Jonathan Feinstein, A Microeconometric Analysis of Income Tax Evasion (unpublished manuscript, Massachusetts Institute of Technology, Dep't Economics, 1986).

[8] Jonathan Feinstein, The Safety Regulation of U.S. Nuclear Power Plants: Violation, Inspections, and Abnormal Occurrences, 97 J. Pol. Econ. 115 (1989).

My case study of OSHA regulation is based on inspections of 755 firms during 1985. Excellent previous studies of OSHA by Viscusi, and Bartel and Thomas have highlighted OSHA's problems in insuring workplace safety.[9] Bartel and Thomas are particularly clear in drawing a distinction between two competing explanations of OSHA's difficulties: the noncompliance hypothesis, which asserts that firms routinely violate OSHA standards; and the inefficacy hypothesis, which asserts that the standards themselves would be of little benefit even if fully complied with. I use detection controlled methods to help sort out these explanations, developing improved estimates of noncompliance and estimating the variation in detection rates among inspectors. The analysis pays particular attention to the effect of unemployment rates on noncompliance and to OSHA's treatment of union plants. Finally, the correlation between noncompliance, both detected and undetected (as estimated), and injury rates is explored.

One finding of this study parallels my findings in related studies of income tax evasion and nuclear power (cited above): there is evidence of substantial heterogeneity among OSHA inspectors in their detection rates. In fact, in all three studies the variation in detection rates among inspectors is comparable in magnitude to the variation in violation rates among potential offenders.

The remainder of the article is organized as follows. The next section presents the simplest detection controlled model, explores its statistical form, and characterizes the biases that can emerge when incomplete detection is ignored. Section III presents the case study of OSHA regulation. Section IV provides a brief conclusion. Appendix A extends the detection controlled methodology to more advanced settings and investigates identification; this section is technically more difficult than the main body of the article. Finally, Appendix B provides technical proofs, and Appendix C, an example of how to derive models of noncompliance from economic theories of choice under uncertainty.

II. Concepts

Consider the analysis of a particular illegitimate activity, such as income tax evasion or regulatory noncompliance. Associated with the activity is a population of individuals or firms, each of whom decides whether to participate in the activity or remain fully compliant; call this

[9] Kip Viscusi, The Impact of Occupational Safety and Health Regulation, 10 Bell J. Econ. 117 (1979); Kip Viscusi, Risk By Choice (1983); Ann Bartel & Lacy Glenn Thomas, Direct and Indirect Effects of Regulation: A New Look at OSHA's Impact, 28 J. Law & Econ. 1 (1985).

group the *potential offenders*. Assigned to monitor the population of potential offenders and detect violations is a second group, called *monitors*. Assuming, as I will throughout, that each potential offender is assigned either one monitor, or a small team of monitors, it is natural to organize the data into the potential offender–monitor pairings that this assignment creates; each such pair is a *case*.[10]

The interpretation of these definitions depends on the context. Thus, in applications to regulation, each case refers to an inspection, the term "monitor" refers to the regulatory personnel who inspects, and the inspected firm is the potential offender. An analysis of income tax evasion views each case as a specific tax examination; extending the other terms in the obvious way, the tax examiner is the monitor; the individual or firm being examined, the potential offender. Last, an application to accounting defines each audit to be a case; the accountant who performs the audit is the monitor, and the client is the potential offender.

If data are collected on a cross section of potential offenders, as in the empirical example presented later in this article (Section III), each potential offender is associated with only one case. However, a particular monitor will frequently be assigned more than one potential offender—that is, he will participate in many cases. In time-series extensions of the basic method, potential offenders will themselves be included in many cases; this is as one would expect in ongoing regulatory or auditing relationships.

As the terminology suggests, I focus on applications in which the identity of individual monitors is known. Nonetheless, the methodology remains useful even when such detailed information is unavailable. For example, in the analysis of street crimes or auto-emissions standards, it may be possible only to associate each potential offender with his geographic residence: city in the first instance, state in the second. The term "monitor" then refers to the municipal or state agency responsible for enforcement, and controlling for variability in detection rates makes sense when these agencies differ (across geographic locale) in their per capita budgets, personnel quality, and the like.[11]

[10] We assume the sample is random (or stratified random). If sampling is not random, estimation is still possible but is more complex; related models are in James J. Heckman, Sample Bias as a Specification Error, 47 Econometrica 153 (1979); and Charles F. Manski and Daniel McFadden, Statistical Analysis of Discrete Data with Econometric Applications (1981).

[11] It may even be possible to extend the methodology to situations in which only aggregate data are available on both potential offenders and monitors; data are similar to those used by Isaac Ehrlich, Participation in Illegitimate Activities: A Theoretical and Empirical Investigation 81 J. Pol. Econ. 521 (1973), and others. I have not explored this possibility in any great detail, however.

Now consider a representative case i, and suppose that potential offender i makes a simple binary choice whether or not to violate; extending the approach to more complex decision contexts, such as the additional choice of how much to violate (for example, in income tax evasion), is straightforward and is discussed in Appendix A. We can represent i's decision-making process in terms of a conventional latent variables formulation:

$$Y_{1i} = x_{1i}\beta_1 + \epsilon_{1i}, \tag{2.1}$$

where

$$L_{1i} = 1 \text{ (violation)} \quad \text{if } Y_{1i} > 0,$$
$$L_{1i} = 0 \text{ (compliance)} \quad \text{if } Y_{1i} \leq 0,$$

where x_{1i} is a vector of characteristics of potential offender i, and ϵ_{1i} is a mean zero random disturbance that is drawn from the distribution $F(\ )$.

In most situations, equation (2.1) can be derived from underlying economic theory. Thus, when the potential offender is an individual, the individual's choice falls within the broad category of decision-making under conditions of uncertainty and the theory of criminal behavior developed by Becker and others;[12] Appendix B presents an example of how this theory leads to equations resembling (2.1). In applications to regulation, or, more generally, when the potential offender is a firm or other large organization, an alternative derivation focuses on the resources that potential offenders devote to achieving compliance. To cite an example, in related work on nuclear power safety regulation, a model has been proposed in which power plant management chooses the level of resources to devote to compliance each month; violations then occur stochastically according to a Poisson process characterized by the parameter $x_{1i}\beta_1$—a larger value of this parameter indicating a greater likelihood of violation, just as in (2.1).[13] In general, derivation of the potential offender's decision from a structural model leads to a more complicated expression than (2.1), in that a nonlinear term $r\,(x_{1i}\beta_1)$ replaces the linear $x_{1i}\beta_1$. All of the arguments of this article continue to apply, however, to this more general case.[14]

[12] Gary Becker, Crime and Punishment: An Economic Approach, 76 J. Pol. Econ. 169 (1968); Michael Block & J. M. Heineke, A Labor Theoretic Analysis of the Criminal Choice, 65 Am. Econ. Rev. 314 (1975).

[13] The model is presented in Feinstein, *supra* note 8; a number of other models are also developed there.

[14] The derivation of Appendix C is an example of this phenomenon.

Despite appearances, equation (2.1) is not a conventional binary choice model because its dependent variable L_{1i} is not directly observable. Potential offender i will be observed violating the law only if he is detected; otherwise, there will be no record of his violation, and he will appear to have been compliant. Nonetheless, previous studies of crime and regulatory noncompliance, a number of which were mentioned in the introduction, have estimated β_1 from (2.1), or aggregated versions of (2.1), by substituting a measure of detected violations into the analysis as a proxy for the true violation measure L_{1i}. When many violations go undetected, such a substitution introduces serious biases into the analysis, a point I will return to shortly.

To estimate β_1 correctly requires recognizing that a complete model of illegitimate activities must include an analysis of the detection process. After all, the fact that L_{1i} is not always observable does not arise by chance but from the underlying economics of the violation decision: potential offender i chooses to commit a violation under the assumption that he has a reasonable probability of escaping detection—otherwise, he would surely remain law-abiding since detected violators face penalties that almost always exceed the returns to crime.

To incorporate the detection process into the analysis, we can supplement the violation equation (2.1) with a detection equation. Conditional on L_{1i} being equal to one, set

$$Y_{2i} = x_{2i}\beta_2 + \epsilon_{2i}, \tag{2.2}$$

where

$$L_{2i} = 1 \text{ (detection)} \quad \text{if } Y_{2i} > 0,$$

$$L_{2i} = 0 \text{ (no detection)} \quad \text{if } Y_{2i} \leq 0,$$

where x_{2i} is a vector of explanatory variables for case i's detecton process, and ϵ_{2i} is a mean-zero random disturbance drawn from the distribution $G(\ )$.

Adding a detection equation into the analysis presupposes collecting data on monitors and the detection technology. When each monitor is assigned a reasonably large number of cases, one useful collection of variables to include in x_2 are examiner fixed effects; these provide evidence on the distribution of detection rates among monitors, which in turn is useful in assessing the heterogeneity in performance among monitors (the variance of the distribution).[15] In general, x_2 may also include a variety of other variables, such as the time spent monitoring each case, or

[15] For consistency, the number of cases per monitor must be large enough for asymptotics to apply (because the models are nonlinear).

the conditions under which audits are performed. Evaluating different aspects of the detection process through the estimation of β_2 is ultimately an important public-management tool.

In spirit, equation (2.2) is the statistical analogue of a number of theoretical models of detection and auditing that have emerged in the literature in recent years, including those of Townsend, Baron and Besanko, Laffont and Tirole, and Reinganum and Wilde.[16] In form, however, it differs from this prior theoretical work in several respects. First, equation (2.2) differs from the theoretical work in its interpretation of how uncertainty affects the detection process. We can most simply interpret equation (2.2) as representing a linear detection technology in which detection depends partly on a stochastic component, (ϵ_{2i}), and partly on a systematic component that varies from case to case, ($x_{2i}\beta_2$); thus detection itself is uncertain. In contrast, the theoretical work has typically assumed that, if the monitor conducts an audit, he always collects the information he seeks, though usually at a cost. While the audit itself has a certain outcome, uncertainty derives from two other sources: the random probability of audit, and the difficulty in relating the audit's information to prior behavior by the potential offender.[17] It is likely that equation (2.2) could be reinterpreted to include either of these effects, but it is not clear which specification of the detection process is most appropriate.

Second, the theoretical work has emphasized the interdependence between the potential offender's decision and the monitor's detection effort, typically by viewing both decisions as part of a sequential move game. Appendix A discusses ways of introducing interdependency into equations (2.1) and (2.2), but a more formal game-theoretic development is left to future work.

A limitation of (2.2) is that false detection of a violation is not allowed for.[18] Extension of the basic model to allow for false detection is possible, but it tends to complicate statistical estimation.[19] It should be noted,

[16] Robert Townsend, Optimal Contracts and Competitive Markets with Costly State Verification, 21 J. Econ. Theory 265 (1979); Baron & Besanko, *supra* note 6; Laffont & Tirole, *supra* note 6; and Jennifer Reinganum & Louis Wilde, Income Tax Compliance in a Principal-Agent Framework, 26 J. Public Econ. 1 (1985).

[17] For example, a regulated firm may have claimed some ex ante marginal cost of production, and when the ex post value (which the audit uncovers) is lower, it may be difficult to determine whether the firm lied or simply misestimated its costs.

[18] This is why the term "conditional on $L_{1i} = 1$" is used.

[19] One must modify the term FG by adding a term referring to the possibility of $(1 - F)G'$, where G' is the probability of false detection (typically less than G). Similarly, $(1 - FG)$ must be modified by subtracting a term. For a further discussion, see Jonathan Feinstein, Detection Controlled Estimation: Theory and Applications, ch. 1 (unpublished Ph.D. dissertation, Massachusetts Institute of Technology, Dep't Economics, June 1987).

however, that the identity of the potential offender is not in question here, and by false detection I do not mean a situation where a crime has been committed and the accused may not be guilty. Instead, false detection refers only to the question of whether the detector cites a violation when none has occurred. False detection may well be less empirically relevant than the related problem of "grey areas" of the law, which can affect empirical analysis whenever monitors misinterpret compliant behavior as illegitimate or differ among themselves in the appropriate definition of a violation.[20]

Equations (2.1) and (2.2) together form a complete model of the compliance-detection system. It remains true, however, that L_{1i} and L_{2i} are separately unobservable. Nonetheless, a maximum-likelihood estimator can be derived that explicitly takes into account this unobservability. In this section, I derive this estimator for the system outlined in (2.1) and (2.2); in Appendix A, extensions to more complex settings are discussed.

Consider the *i*th case. Assume that ϵ_{1i} and ϵ_{2i} are independent of one another.[21] Estimation in the more general case when the two errors are correlated requires more numerical calculation, but is conceptually similar, and is illustrated for the case of normally distributed errors in Appendix A.

The probability of observing a detected violation is then simply the product $F(x_{1i}\beta_1)G(x_{2i}\beta_2)$, which represents the probability of a violation multiplied by the probability of detecting the violation (conditional on its having been committed). The term $F(x_{1i}\beta_1)G(x_{2i}\beta_2)$ is a straightforward extension of the usual binary choice calculation, in which the probability of observing an action is $F(x_{1i}\beta_1)$; what is noteworthy about its form is that it treats the violation and detection probabilities symmetrically. Similarly, the probability of not observing a detected violation can be calculated to be $[1 - F(x_{1i}\beta_1)G(x_{2i}\beta_2)]$. It is this second term, $[1 - F(x_{1i}\beta_1)G(x_{2i}\beta_2)]$, which is at the heart of the estimation procedure because it reflects the sum of two terms that cannot be separated by the econometrician, who has data only on detected crimes: (1) the probability that no violation has been committed, $1 - F(x_{1i}\beta_1)$; and (2) the probability that a violation has been committed but not detected, $F(x_{1i}\beta_1)[1 - G(x_{2i}\beta_2)]$. Hence, this second term explicitly recognizes the possibility of a violation remaining undetected.

The cases fall into two disjoint sets: one set, called set A, consists of those cases for which a detected violation is recorded; the other set, A^c,

[20] Hence, the term "detection" may require modification to "perceived wrongdoing."

[21] We also assume errors across cases are independent of one another.

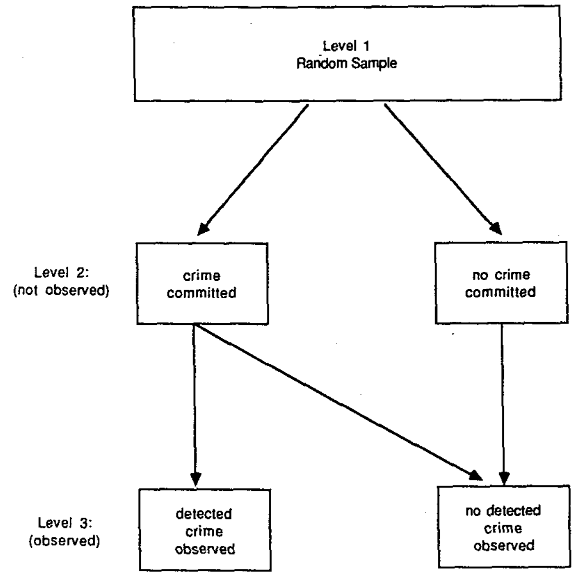


FIGURE 1

consists of the remaining cases, for which no detected violation is recorded. The log likelihood of the observations (cases) is then

$$L = \sum_{i \in A} \log[F(x_{1i};\beta_1)G(x_{2i};\beta_2)] + \sum_{i \in A^c} \log[1 - F(x_{1i};\beta_1)G(x_{2i};\beta_2)], \quad (2.3)$$

and the technique of maximum likelihood estimation (MLE) can be used to consistently estimate β_1 and β_2. Alternatively, equation (2.3) can be estimated via a nonlinear least squares regression in which the dependent variable is the product $L_{1i}L_{2i}$ (a zero-one indicator variable that takes the value one for a "detected violation" and zero for "no detected violation"), and the regression equation is $L_{1i}L_{2i} = F(x_{1i};\beta_1)G(x_{2i};\beta_2) + \nu_i$, where ν_i is a mean zero disturbance. Equation (2.3) is the basic structure from which all of the additional likelihoods of this article derive, for each of which an analogous nonlinear least squares interpretation can be developed; because it explicitly includes a model of the detection process and allows for the possibility of less-than-perfect detection, an appropriate name for the MLE based on it is *detection controlled estimation.*

Discussion

Equation (2.3) belongs to a family of MLE techniques that are often called "missing information algorithms" and that have been discussed by

Rao; Dempster, Laird, and Rubin; Rubin; and Cox and Oakes in the biometrics literature and by Poirier in the econometrics literature.[22] Figure 1 illustrates why this is an appropriate name. It displays three "levels" of data. Level 1 refers to the underlying random sample of cases. Level 2 splits these cases into two disjoint and exhaustive subsets: the set of cases for which the potential offender has in fact committed a violation, and the set for which he has not. Level 2 is not directly observable to the econometrician. Level 3 again splits the cases into two disjoint and exhaustive subsets; one subset refers to the collection of cases for which a violation was committed and later detected—this subset is itself a subset of the level 2 set consisting of the cases for which a violation was committed. The other level 3 subset refers to cases for which no detected violation is observed—it consists of both the level 2 subset for which no violation was committed and that part of the level 2 subset of committed violations for which the violation was never detected.

Level 2 represents "missing information" because it is unobservable. We would like to infer what happened at level 2—in particular, for those observations with no detected violation (the set A^c), we would like to determine whether a violation was committed or not. While this question cannot be answered absolutely, once the parameters β_1 and β_2 have been consistently estimated, we can "fill in" level 2 and compute the probability that a violation was committed for each case in A^c. Letting β_1^* and β_2^* denote the DCE estimates, a simple application of Bayes's Law demonstrates that for a case in A^c the probability of an undetected violation is $F(x_{1i}\beta_1^*)[1 - G(x_{2i}\beta_2^*)]/[1 - F(x_{1i}\beta_1^*)G(x_{2i}\beta_2^*)]$. This estimate of the probability that an individual has committed a violation and escaped detection is a useful screening device when further resources are available to be allocated to additional monitoring; relevant examples include the possibility of a second round of IRS tax audits or a reinvestigation of ostensibly compliant regulated firms. The estimate can also be used to develop statistical tests of the effect of undetected violations on related performance measures, such as in related work on nuclear power, which explores the relationship between undetected safety violations at nuclear power plants and future unsafe events.[23]

The rate of undetected violations over the population as a whole can be consistently estimated as

[22] C. R. Rao, Linear Statistical Inference and Its Applications (1965); Dempster, Laird, & Rubin, *supra* note 3; Donald Rubin, Inference and Missing Data, 63 Biometrika 581 (1976); Cox & Oakes, *supra* note 3; and Poirier, *supra* note 4.

[23] Feinstein, *supra* note 8.

$$(1/T) \sum_{i \in A^c} \frac{F(x_{1i}\beta_1^*)[1 - G(x_{2i}\beta_2^*)]}{1 - F(x_{1i}\beta_1^*)G(x_{2i}\beta_2^*)}. \tag{2.4}$$

Such an estimate is of considerable policy interest, particularly as compared with the rate of detected violations, which is directly observable and is measured as the number of observations in A divided by T.[24]

Since equation (2.3) is a missing information system, it is natural to wonder what its complete information counterpart is. In fact, if complete information about violations were available, the cases would fall into three sets rather than two, consisting of (1) potential offenders who have not committed a violation, (2) potential offenders who have committed a violation and escaped detection, and (3) potential offenders who have been detected committing a violation. Letting B^1, B^2, and B^3 denote these three sets, the complete-information likelihood analogous to (2.3) would be

$$\begin{aligned} \mathrm{L} = & \sum_{i \in B^1} \log[1 - F(x_{1i}\beta_1)] + \sum_{i \in B^2} \log \{F(x_{1i}\beta_1)[1 - G(x_{2i}\beta_2)]\} \\ & + \sum_{i \in B^3} \log [F(x_{1i}\beta_1)G(x_{2i}\beta_2)], \end{aligned} \tag{2.5}$$

which is a nested binary choice model of the type discussed by McFadden.[25] The missing information likelihood (2.3) differs from (2.5) in that it combines the sets B^1 and B^2 into the set A^c, while identifying B^3 with A. In turn, this inability to separate compliant potential offenders from undetected violators reduces the efficiency with which the parameters β_1 and β_2 can be estimated. Controlled experiments (such as the Monte Carlo experiments discussed at the end of the next section) provide a setting in which this efficiency loss can be quantified.[26]

The most serious statistical issue that arises in estimating detection controlled models like (2.3) is identification. Intuitively, identification arises because the DCE decomposes a single datum, detected violations, into two disjoint behavioral categories, violation and detection, and it is not initially clear whether this decomposition can be performed uniquely.

[24] However, this estimate is also more sensitive to assumptions about the form of F and G than the parameter estimates themselves—this is discussed further below and in Appendix A.

[25] McFadden, *supra* note 4.

[26] Here the "behavior" of potential offenders is simulated, and so a classification into the three B sets is possible, allowing a direct comparison of estimates based on both (2.3) and (2.5).

As an illustration of the problem, consider the following example. Suppose that the probability of potential offender i choosing to violate is $p_0 e^{x_{1i}\beta_1}$ (which refers to $F(x_{1i}\beta_1)$ in the earlier notation), where p_0 is the average level of noncompliance in the population and $e^{x_{1i}\beta_1}$ fluctuates around one. Similarly, let the probability of monitor i detecting the violation be $q_0 e^{x_{2i}\beta_2}$ ($G(x_{2i}\beta_2)$). Data are available only on detected violations, the product FG on which likelihood (2.3) depends. In this example, that product is $e^{x_{1i}\beta_1} p_0 q_0 e^{x_{2i}\beta_2}$, and it is seen that p_0 and q_0 cannot be separately identified, only their product $p_0 q_0$. Put differently, a given average level of detected violations (the product $p_0 q_0$) might refer to a high average level of violation and low average detection rate (high p_0 and low q_0) or to the reverse (low p_0 and high q_0); we cannot distinguish between these two possibilities when we possess data only on detected violations.

Continuing to refer to the example, observe that, while variables included only in x_{1i} or x_{2i} have fully identified β's, variables in common have parameters identified only as reduced forms. In response to this point, however, it should be noted that a variable that is expected to lower noncompliance when increased (such as sanctions), and is included in the detection equation, is likely to raise detection (sanctions promise higher rewards); and, conversely, variables likely to raise noncompliance will lower detection. When this is the case, the variable's predominant effect is identifiable by the sign of its reduced-form coefficient.

This double exponential form is extreme. Appendix A, in the context of a much more complete (and more technical) exploration of the identification issue, demonstrates that the double exponential is the only functional form in which identification of absolute levels formally fails.[27] Nonetheless, the example serves as a warning that relative noncompliance and detection rates (those that vary with x_i across cases) will typically be better estimated than the average levels of violation and detection. In turn, this indicates that estimates of the undetected violation rate, such as (2.4), must be treated with caution since they rely on knowledge of the absolute levels of noncompliance and detection.

Biases

Model 1 (consisting of eqq. [2.1], [2.2], and [2.3]) is a convenient medium through which to explore the biases that arise when analysts do not control for nondetection. Statistical methods that do not model the detection process will typically fit a binary choice model based on equation

[27] Appendix A also discusses identification when F and G are not restricted to parametric families, such as the normal, but are left free.

(2.1) to the data (which is restricted to the variables x_1), assuming that observations in the set A refer to individuals who have committed a violation, and those in A^c refer to individuals who have not.[28] The likelihood function then reduces to the conventional form

$$L = \sum_{i \in A} \log[F(x_{1i}\beta_1)] + \sum_{i \in A^c} \log[1 - F(x_{1i}\beta_1)]. \qquad (2.6)$$

Comparing equation (2.6) to (2.3), it is apparent that studies of illegitimate activities that fail to incorporate detection into the analysis are implicitly assuming that G is always one, that is, that detection is perfect.[29] A comparison of the fits of (2.6) and (2.3) (via a likelihood ratio test) allows a direct test of the assumption that detection is complete and homogeneous across cases. In most instances, we cannot expect this hypothesis to be coriect, and equation (2.6) is then misspecified; parameter estimates based on it will be inconsistent.

What sort of biases will emerge if (2.6) is estimated instead of (2.3)? Since the detection probability G is always less than or equal to one, (2.6) tends to systematically understate the true extent of violations. Hence, the estimate of positive elements of β_1, which are associated with increases in the probability of an offense, are likely to be biased downward. The following theorem formalizes this intuition, drawing on the analysis presented by White, who discusses the general structure of misspecified maximum likelihood models.[30]

THEOREM 2.1. (i) Let F be a distribution function such that the likelihood (2.6) is concave (such as probit, logit), and let x_{1k} be nonnegative. Let $d(x_1) = E_{x2}[G(x_2\beta_2|x_1)]$ be the average detection rate for each x_1. Then when (2.6) is estimated and detection is incomplete ($d(x_1) < 1$ on some x_1 set of positive measure), the estimate of β_{1k} will be biased downward.

(ii) As detection rates fall ($d(x_1)$ decreases on at least one x_1 set of positive measure and does not increase for any x_1), the bias becomes larger.[31]

All proofs are in Appendix B.

Some of the bias described in theorem 2.1 can be removed by express-

[28] More generally, they will use whatever statistical form their compliance equation indicates, such as Tobit, etc.

[29] Interestingly, even if detection is the same on every case, but equal to $g < 1$, equation (2.6) remains misspecified.

[30] Halbert White, Maximum Likelihood Estimation of Misspecified Models, 50 Econometrica 1 (1982).

[31] While theorem 2.1 does require likelihood (2.6) to be concave, it places no restrictions other than nonnegativity on x_{1k}—in particular, x_{1k} may be correlated with the other components of x_1 or with x_2.

ing variables as deviations from their means; nonetheless, some bias will remain whenever the relevant element of x_1 possesses an asymmetrical distribution or is correlated with some other element of x_1 that possesses an asymmetrical distribution.

Potentially more serious biases arise when elements of x_1 are correlated with some aspect of the detection process; in this case, the biases may be systematic and of direct relevance to hypotheses of economic behavior. One example of such a bias arises in tests of the deterrence hypothesis emphasized by Becker; Ehrlich; and Block, Nold, and Sidak, among others.[32] The deterrence hypothesis claims that increases in sanctions deter potential offenders from committing offenses.

Tests of the deterrence hypothesis that use detected violation rates may be systematically biased toward finding no deterrent effect for the following reason.[33] When sanctions rise, detection rates are likely to increase on two counts: (1) individuals have stronger incentives to monitor offenders and detect violations; and (2) an increase in sanctions signals increased public concern about crime, which may well be accompanied by an increase in the resources devoted to law enforcement. (Against these two effects must be weighed the countervailing tendency for offenders to expend more effort concealing their crimes, relevant only when concealment is possible.) Since detection rates increase, the detected violation rate falls proportionately less than the true violation rate. This reduces the observed deterrent effect and biases tests of the null hypothesis of no deterrent effect toward being accepted.

A second instance where biases may arise is in examining the relationship between crime and unemployment. Typically, crime rates are expected to rise when unemployment increases because individuals face restricted legal opportunities, or, in the regulatory context, firms have fewer resources available to meet regulatory requirements. However, when economic activity is lower, governments collect fewer revenues

[32] Becker, *supra* note 12; Ehrlich, *supra* note 11; Isaac Ehrlich, The Deterrent Effect of Capital Punishment: A Question of Life and Death, 65 Am. Econ. Rev. 397 (1975); Michael Block, Frederick Nold, & Joseph Sidak, The Deterrent Effect of Antitrust Enforcement, 89 J. Pol. Econ. 429 (1981).

[33] Ehrlich estimates deterrent effects for a variety of street crimes. He uses aggregate data on reported crimes, so that the detection issues of this article are not directly applicable to his results. To the extent that many crimes go unreported, however, a "nonreporting bias" similar to the nondetection bias will emerge, and comparable techniques can be used to correct for it. Block, Nold, and Sidak, *supra* note 33, investigate the deterrent effect of antitrust enforcement, using less aggregated data. These authors implicitly assume that all unexplained price changes are due to collusive behavior, an extreme form of the complete detection assumption; their analysis could be improved on by estimating a joint violation-detection system.

(and frequently spend more of what they do collect on expansionary fiscal policies); hence, less is spent on detection, and detection rates can be expected to fall. As a result, the detected crime rate rises proportionately less than the true crime rate during periods of higher unemployment, so that tests of the null hypothesis of no relationship between unemployment and crime will tend to accept the null too frequently.

Finally, consider estimates of the relative frequencies with which different subgroups of the population commit violations. If two different subgroups, represented by x_{1ij} ($= 1$ if potential offender i is a member of subgroup j, 0 otherwise) and x_{1ik}, are monitored with equal intensity, then estimates of the relative frequency with which these two groups commit violations (represented by the ratio of β_{1j} to β_{1k}) based on (2.6) are likely to be approximately valid. However, if one subgroup, say j, is assigned better monitors, its detected violation rate will be higher relative to its true rate of violation than will be the case for subgroup k; hence, the ratio of β_{1j} to β_{1k} estimated by (2.6) will be biased upward. The following theorem formalizes this intuition, albeit under additional restrictions.

THEOREM 2.2. Let x_{1j} and x_{1k} be two components of x_1. Restrict x_{1k} and x_{1j} to be zero-one indicator variables that refer to population subgroups, so that x_{1j} and x_{1k} are orthogonal. Also assume that x_{1j} and x_{1k} are uncorrelated with the other components of x_1, and that $\beta_{1j} = \beta_{1k}$. Then, if population group j faces higher detection rates than group k, the estimate of β_{1j}/β_{1k} will be biased upward.

The bias discussed in theorem 2.2 and above extends naturally to the interesting situation where one subgroup, say j, is more likely to commit violations, so that β_{1j} exceeds β_{1k}.[34] Economic theory suggests that subgroup j will be assigned better monitors, so that group j's detected violation rate will exceed its true violation rate by relatively more than is the case for subgroup k. Estimates based on (2.6) then possess an "overdispersion" bias in which subgroup j appears to commit an even higher proportion of violations than is actually the case.

III. EMPIRICAL EVIDENCE

This section illustrates the detection controlled method with a case study of OSHA safety regulation.[35] I focus on topics discussed in the previous section, specifically the variation in detection rates among OSHA inspectors and the biases that arise in nondetection controlled

[34] Unfortunately, I have not been able to extend the theorem to this situation.

[35] More comprehensive applications to nuclear power and income tax evasion were cited in Section I.

models; I also explore the relationship between violations, both detected and undetected, and injury rates.

In addressing these issues, the detection controlled analysis extends previous studies of OSHA by Viscusi, and Bartel and Thomas.[36] Viscusi follows the traditional approach to evaluating regulatory effectiveness exemplified in a series of essays by Stigler.[37] He constructs a measure of welfare that he believes to be "OSHA's target," injury rates aggregated (by Standard Industrial Classification [SIC] code) to the industry level. He then regresses this variable against the size of OSHA's budget, the number of OSHA actions against the industry in the preceding year, and the number of OSHA employees. He also reverses this logic and regresses measures of enforcement against injury rates. He finds nearly all these relationships to have a statistically insignificant effect and concludes that OSHA regulation is relatively ineffective at controlling workplace hazards as measured by injuries suffered on the job.

Though Viscusi's analysis is valuable, it does not adequately explore possible reasons *why* OSHA is ineffective. Bartel and Thomas argue that there are two principal hypotheses: (1) the noncompliance hypothesis—OSHA is ineffective because firms do not comply with the regulations, and if compliance rates could be improved OSHA would be more successful; and (2) the inefficacy hypothesis—compliance with OSHA regulations is high, but the regulations themselves are poorly designed and need to be improved. They note that the traditional approach to regulation, of which Viscusi's work is an example, "excludes specific consideration of corporate noncompliance with OSHA standards. The relationship (if any) between accidents and violations (the inefficacy hypothesis) and between violations and enforcement (the noncompliance hypothesis) thus cannot be isolated and examined."[38] As Bartel and Thomas go on to note, "the inefficacy and noncompliance hypotheses have profoundly different implications for public policy, so that formulation of an appropriate policy response requires not merely a demonstration that OSHA fails but an understanding of why that failure occurs."[39]

Bartel and Thomas do specify a simultaneous system with separate equations to explain violations and injury rates and the relationship between them. However, they do not model the inspection process and proxy the violations variable with the rate of detected violations, thereby

[36] *Supra* note 9.

[37] George Stigler, The Citizen and the State (1982).

[38] Bartel & Thomas, *supra* note 9, at 2–3.

[39] *Id*. at 2–3.

implicitly assuming complete detection. In fact, the detection process is a central part of OSHA; more than half of its budget is devoted to enforcement, and the inspections themselves are the primary channel for day-to-day interaction between the administration and industry.[40] From a management perspective, improving OSHA's performance is likely to require improving the inspection and enforcement process.

To apply the detection controlled method, I contacted OSHA and received data on all programmed safety inspections of industrial (SIC codes 2000–3999) plants in region 1 (New England) during April–October 1985. Programmed inspections refer to a random sampling of firms for inspection within each broad SIC class and therefore avoid sample selection problems.[41]

The data consist of 755 inspections, and each inspection lists the company name, its address, its SIC code, whether its employees belong to a union, the number of employees on site, the total number of employees on the company's payroll, the inspection date, and the company's average workdays-lost-to-injury rate during 1985. Importantly, the data also list the code number of the inspector who performed the inspection, the number of violations detected, and the dollar value of any penalties assessed the company.

The data are used to construct the following variables: SIC, which lists the two-digit SIC code of the company; UNION, a dummy variable set to one if the company's employees belong to a union and zero otherwise; EMPS, the total number of employees on site; and EMPT, the total number of employees on the company's payroll, which can substantially exceed EMPS when the site is one plant of a large company. I summarize the data on violations by the indicator variable VIOL, which takes the value one if any violations were cited, and the value zero otherwise.[42] I have supplemented this data set with the variable U, the unemployment rate for the month of the inspection in the state in which the company branch is located.[43]

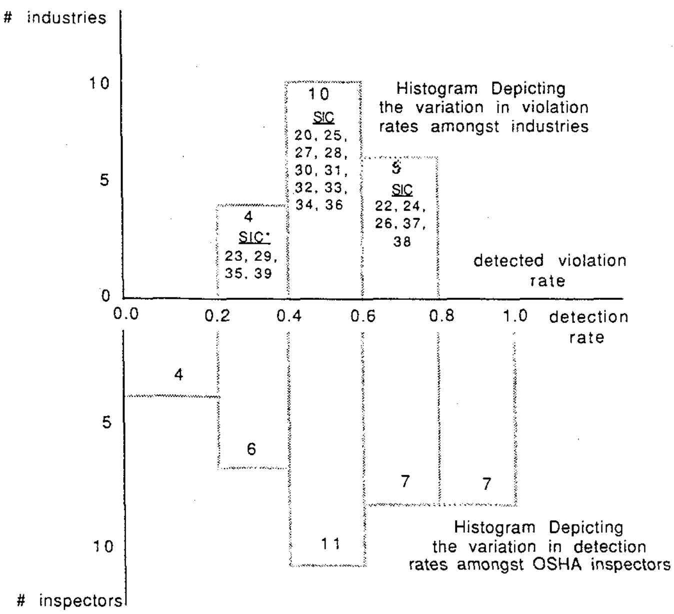
Figure 2 describes the raw data on noncompliance and detection. The upward pointing histogram depicts the variation in detected noncompliance rates (as measured by VIOL) across the nineteen two-digit SIC

[40] Viscusi, Risk, *supra* note 9.

[41] The other type of inspection is "complaint" based and originates in employees coming to OSHA requesting a review of possible safety violations; these inspections represent a selected sample.

[42] For an analysis in which the number of violations is used directly (as the realization of a Poisson process), see Feinstein, *supra* note 8.

[43] Source: Department of Commerce, State and County Data (1985).



* List of two digit SIC categories in this bin.

FIGURE 2.—Raw data histograms

industry groups represented in the sample—thus four industry groups (apparel; petroleum refining; machinery; and miscellaneous) have detected rates (fraction of inspections for which at least one violation was cited) between 20 percent and 40 percent; ten (food; furniture; printing and publishing; chemicals; rubber and plastics; leather; stone, clay, and glass; primary metal; fabricated metal; and electrical machinery) have rates between 40 percent and 60 percent; and five (textiles; lumber and wood; paper; transportation excluding motor vehicles; and instruments) between 60 percent and 80 percent. The downward pointing histogram depicts the variation in detection rates among the thirty-five inspectors listed in the data. Notice that the variation in detection rates is comparable in magnitude to the variation in violation rates, which suggests that DCE will be a considerable improvement over nondetection controlled methods that ignore the variation in detection among inspectors. To explore the variation in detection rates further, individual inspectors' rates were compared to the average citation rate in the sample, which is 52 percent (at least one violation was cited 52 percent of the time). As the histogram indicates, many inspectors have rates far above the average (seven have rates above 80 percent), and many others rates far below. By computing the difference between each inspector's rate and the average, squaring this difference, and summing over all thirty-five inspectors, one quantifies the variation in citation rates and obtains a formal test of the null hypothesis that all inspectors possess the same rate (in which case the variation in rates would arise entirely from noise in the data); this test strongly rejects the null.[44]

Figure 2 and the test for homogeneous detection based on it are too simple because the two dimensions of noncompliance and detection cannot be analyzed in isolation from one another. That is, some inspectors may have low detection rates simply because they are more frequently assigned compliant firms. Similarly, different quality inspectors may be assigned to different SIC classifications, leading to biases in the raw data assessment of violation rates across industry groups. A correct analysis must analyze the data simultaneously along both dimensions, controlling both for variations in firm characteristics and inspectors; this is the point of the detection controlled technique.

The model of noncompliance to be estimated presumes that the likelihood of a plant committing at least one violation depends on its union status (UNION), the unemployment rate in its state at the time of the inspection (U), the number of employees on site (EMPS), the total number of employees employed by the firm (EMPT), and the plant's main industrial operation (SIC code).

There are a number of theoretical and policy arguments of interest in assessing these relationships. First, consider the effect of union status. Viscusi summarizes and extends a large literature that describes the effects of unionization on fringe benefits in general and workplace safety in particular.[45] Theoretical arguments suggest that the presence of unions steepens the trade-off between risk and compensating wages (that is, unions will demand a larger dollar compensation for each increment of added risk), leading unionized firms to provide safer working conditions than nonunion firms. However, as Viscusi notes, unions are more likely to organize in unsafe firms (since the potential benefits are greater), and therefore it is not clear whether unionization will be associated with safer conditions empirically. In fact, Bartel and Thomas find union status to have a positive, but statistically insignificant, relationship to the propen-

[44] Formally, this test statistic (when divided by 34) is distributed chi-squared with 34 degrees of freedom under the null hypothesis of identical detection rates.

[45] Kip Viscusi, Union Labor Market Structure and the Welfare Implications of the Quality of Work, 1 J. Labor Res. 175 (1980).

sity to violate OSHA standards, and a positive significant relationship to injury rates.

Second, consider the effect of unemployment. The relationship of firm activity to compliance and safety has not been widely addressed in the literature. Nonetheless, there are a number of theoretical reasons for believing the plant's level of activity might affect its ability and willingness to comply. On the one hand, low activity may increase compliance for three reasons: increased idle machine time available for maintenance; lessened workplace congestion; and, if plant management exhibits increasing risk aversion at lower incomes, a desire to reduce accident risks. On the other hand, low activity reduces working capital available for maintenance, capital improvements, and safety training, all of which may lead to decreased compliance. Overall, then, the relationship between economic activity and compliance is ambiguous. It is not clear what measure of activity would be most relevant to these issues. Thus, in addition to U, I also experimented with two other variables designed to measure the level of economic activity: the percent change in prices from June 1984 through June 1985 in the company's SIC group listed as its principal line of business, and the level of employment in the company's SIC group by the month of inspection.[46] Results with these two were similar to those reported in the text and are not discussed further. Direct information on plant orders and sales might be most appropriate, but it was not available.

The variables EMPS and EMPT may be expected to affect compliance whenever there are returns to scale in safety technologies. Thus, larger plants may comply more readily if they experience scale economies in safety equipment or training programs. Similarly, if scale economies exist at the firm level—for example, if larger firms more efficiently disseminate information about new regulations, steps needed to achieve compliance, or training—plants that are members of larger firms may comply more readily.

Model 1 in Table 1 reports the results of fitting this model of noncompliance as a standard probit specification; thus, model 1 ignores the detection process and serves as a point of comparison with the DCE models presented next. In model 1, and all remaining models in Table 1, dummies are included for the seven SIC categories (22, 24, 26, 34, 35, 37, and 39) that have either many cases or unusually high or low detected compliance rates; the dummies are reported in Table 2.

According to model 1, unionization is positively associated with noncompliance, and increased unemployment increases compliance; both of

[46] Both of these variables are constructed using data from the Department of Commerce, *supra* note 43.

TABLE 1

OSHA Results, Dependent Variable: VIOL

Independent Variables	Probit Model 1 (Base)	DCE Model 2 (Base)	DCE Model 3 (Correlated Errors)	DCE Model 4 (Expectations Simultaneous)
Compliance equation:				
CONSTANT1	.648	1.25	.683	.287
	(.223)	(.773)	(.438)	(.467)
UNION	.606*	1.01**	.984*	.520*
	(.127)	(.538)	(.350)	(.127)
EMPS	2.57×10^{-4}	-1.27×10^{-5}	1.25×10^{-4}	1.65×10^{-4}
	(3.21×10^{-4})	(3.24×10^{-4})	(4.01×10^{-4})	(3.40×10^{-4})
EMPT	-3.99×10^{-4}	-1.78×110^{-6}	1.57×10^{-6}	-3.19×10^{-6}
	(8.91×10^{-6})	(2.85×10^{-5})	(3.10×10^{-5})	(1.97×10^{-5})
U	−.191*	−.237	−.0869	.0543
	(.060)	(.223)	(.119)	(.121)
Detection equation:				
CONSTANT 2	. . .	.704	.663	.240
		(.475)	(.597)	(1.31)
$\Phi(\times \beta_1)$	. . .	. . .	. . .	1.39
				(1.14)
ρ	. . .	. . .	.20	. . .
Log likelihood	−486	−451	−451	−453

Note.—The number in parentheses is the standard error.
* Significant at the 5 percent level.
** Significant at the 10 percent level.

TABLE 2

INSPECTOR AND INDUSTRY FIXED EFFECTS

Inspector Number	Model 2	Model 3	Model 4	Model 5
1	.358	.362	−.384	−.246
2	*	1.24	*	.656
3	.358	.916	1.49	−.426
4	−.480	−.312	−.875	−.723
5	−.197	.144	−.214	−.0737
6	.110	.586	*	.2001
7	−.104	−.928	−.966	−.591
8	−1.04	−.945	−1.41	−1.06
9	.354	.681	*	.605
10	*	*	*	.894
11	−.212	.106	−.447	−.715
12	.711	1.47	*	.326
13	−.120	−.995	−1.27	−1.64
14	−.295	−.193	−.433	.0798
15	−.665	−.503	−.945	−1.01
16	−.0857	.0430	.0805	−.869

17	−.416	−.115	−.470	−.0934
18	−.0579	−.0521	−.678	−.822
19	*	*	*	.301
20	*	*	*	*
21	−.107	−.914	−1.38	−1.33
22	−.713	−.572	−1.03	−1.26
23	−.819	−.528	−.830	−1.48
24	−2.16	−2.15	−2.36	−2.12

Industry Number	Model 1	Model 2	Model 3	Model 4	Model 5
22	.206	.462	.270	.116	†
24	.495	†	†	.615	.670
26	.333	.627	.728	.580	†
34	.0735	.105	.161	.117	.0661
35	−.581	−.740	−.713	−.607	−1.05
37	.404	.396	.517	.414	−.362
39	−.400	−.537	−.545	−.527	−.580

NOTE.—The inspector's actual detection rates can be calculated as F applied to the fixed effect plus the constant of the relevant model plus the other terms in the detection equation evaluated at the variables' means reported in Table 1.

* Likelihood is flat at large positive values. Data are consistent with the hypothesis that this inspector is "perfect."

† Likelihood is flat at large positive values. Data are consistent with the hypothesis that essentially all firms in this industry violate.

these relationships are statistically significant. The number of employees on site is postively related to compliance, while total number of firm employees is negatively related to compliance; neither of these effects is significant. Finally, the industry dummies are jointly significant.

The remaining three models reported are detection controlled, in which a detection equation supplements the violation equation. The Occupational Safety and Health Administration has not provided me with information on inspector socioeconomic traits (such as education, age, and race) or experience. Therefore, to control for variations in detection, I have included a series of inspector-fixed effects. Consistency of the maximum-likelihood procedure requires that such effects be specified only for inspectors with a sufficiently large number of cases.[47] In fact, of the thirty-five inspectors listed in the data, twenty-four have performed ten or more inspections; I specify effects for these twenty-four, who are collectively responsible for more than 70 percent of the inspection man-hours in the sample. The detection equation then takes the form

$$Y_{2i} = \beta_{20} + \gamma_j + \epsilon_{2i}, \tag{3.1}$$

where inspector j has been assigned to the ith case and possesses fixed effect γ_j. We are interested, among other things, in the distribution of inspector detection rates implied by these γ's, specifically, (1) how widely spread this distribution is (the degree of heterogeneity), and (2) the shape of the distribution—is it peaked in its center and approximately symmetric, peaked on its two ends (two distinct classes of inspectors), or asymmetric, indicating most inspectors to be of good quality but revealing a tail of underperformers?

Model 2 is the basic DCE model. It assumes ϵ_{1i} and ϵ_{2i} to be independent standard normals, leading to a likelihood of the form (2.3). Model 3 differs from model 2 in allowing arbitrary correlation ρ between the errors ϵ_{1i} and ϵ_{2i} of the violation and detection equations (equations across observations are still assumed to be independent). Interpreting the errors as left-out or unobserved explanatory variables, we might expect them to be correlated if the inspector and/or the inspected plant possess information about one another that is unavailable to the econometrician—thus a positive ϵ_{1i} may lead to increased inspection effort (a positive ϵ_{2i}), inducing a positive correlation. In contrast, model 4 assumes the epsilons to be independent but includes in the detection process a term referring to the inspector's expectation of a violation; we might expect a larger value of this expectation to generate increased inspection effort, suggesting a posi-

[47] More formally, the number of cases for each inspector must become large as the sample becomes large and must be large enough for asymptotic arguments to apply.

TABLE 3

MODEL 5: THE EFFECT OF INJURY RATES ON DETECTION, DEPENDENT VARIABLE: VIOL

Independent Variables	DCE: Model 5 (Injury Rates)
Compliance Equation:	
CONSTANT 1	3.12
	(.773)
UNION	.624
	(.538)
EMPS	-4.85×10^{-4}
	(3.24×10^{-4})
EMPT	-3.35×10^{-6}
	(2.85×10^{-5})
U	−.442*
	(.223)
Detection equation:	
CONSTANT 2	−.504
	(.500)
LDIR	.269*
	(.0277)
Log likelihood	−325

* Significant at 95 percent level.

tive coefficient. Both models 3 and 4 are discussed more fully in Appendix A.

The statistical results suggest a number of conclusions about noncompliance and detection. First, the coefficient on UNION remains positive and statistically significant in all three DCE models, and its magnitude has increased sharply (more than 50 percent) in both models 2 and 3. Thus unionized firms are actually more likely to violate, a finding similar to that of Bartel and Thomas. The fact that UNION's coefficient increases in the DCE models indicates that OSHA assigns less able inspectors to these firms, perferring to send better inspectors to nonunion plants, where OSHA's effect might be larger. The average number of employees on site (EMPS) is 89 in the sample, the average number of employees in the firm (EMPT) is 1,047, and the average unemployment rate (U) is 3.6 percent. Evaluated at these averages, a nonunionized firm has a 65 percent likelihood of at least one violation; a unionized firm, a 92 percent likelihood—both numbers are based on model 2 estimates.

The coefficient on U was statistically significant in model 1, but it is uniformly insignificant (but see model 5 in Table 3) in the DCE models, though it retains a negative coefficient. Thus it would be incorrect to infer that economic activity significantly affects noncompliance; instead, it appears that detection is worse at plants in depressed areas—either because OSHA assigns less able inspectors to these plants or because the inspectors at these plants monitor less closely.[48]

Neither EMPS nor EMPT achieves statistical significance in any of the models. However, in three of the four models, EMPS has a positive coefficient, and in three of the four, EMPT has a negative coefficient. This suggests the following possibility: violations are most frequent at large plants that are part of a relatively small firm; in a slightly different interpretation, violations are highest at large single-site firms. The fact that the magnitude of this effect is reduced in the DCE models indicates that, to some extent, this may be because detection is better at such "central" plants. The quantitative effect of EMPS and EMPT is small: increasing the number of employees on site by 100 above its average level of 89 (and thus also increasing EMPT by 100) does not noticeably alter the probability of a violation.

The SIC dummies across the four models (listed in Table 2) are similar, though their magnitude increases in the DCE specifications. The results indicate that the propensity to violate is higher than average among plants in categories 22 (textiles), 24 (lumber and wood), 26 (paper), and 37 (transportation) and lower than average among plants in categories 35 (machinery) and 39 (miscellaneous), results that mirror the raw data histogram in Figure 2.

Finally, the correlation ρ in model 3 is positive but insignificant—there is weak evidence that unobserved factors tending to increase noncompliance also increase detection; and the expectations term of model 4 is also positive but insignificant.

A comparison of the fits of models 2 and 1 allows a direct test of the hypothesis that detection is complete and homogeneous among inspectors. Twice the difference in the log likelihoods is 70, which far exceeds the critical value of 38 for a chi-squared test with 25 degrees of freedom (representing the overall CONSTANT2 and the twenty-four inspector effects). Thus, the hypothesis is decisively rejected: variation in detection is an important aspect of the data, even controlling for firm characteristics, as these models do. Complete detection is not a sensible working hypothesis for the analysis of OSHA regulation.

Figure 3 presents detection rates for the twenty-four inspectors with ten or more cases. The upward-pointing histogram illustrates the distribution in detection rates computed from the estimates in model 2, while the

[48] In fact, examination of the data indicates that the second explanation is more likely; this conclusion is supported by the fact that, in models 3 and 4, where the detection process depends on the violation process, U's effect is further reduced.

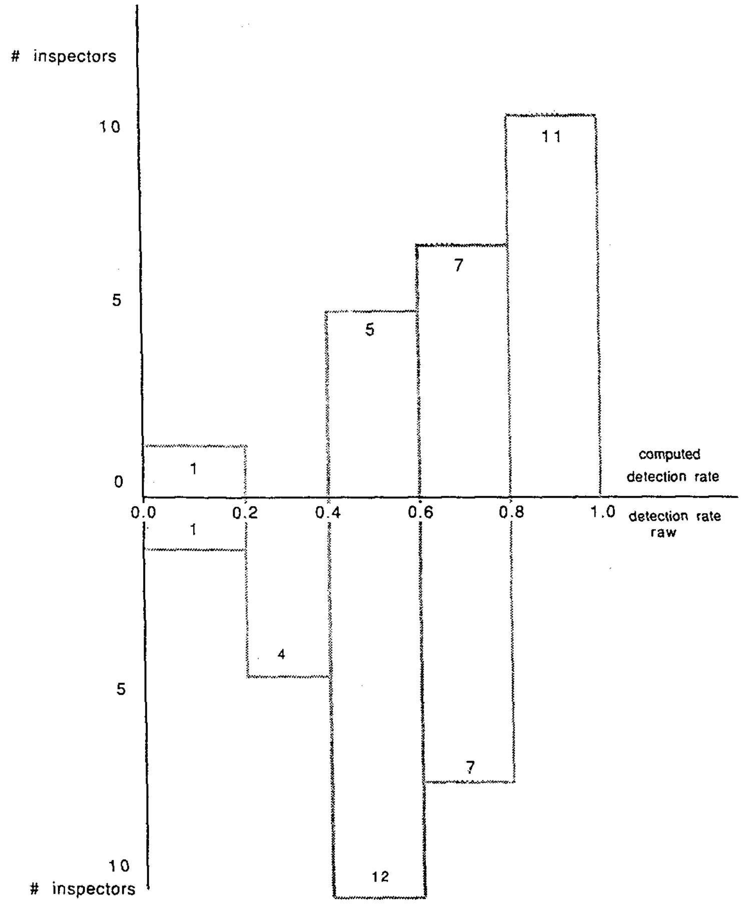


FIGURE 3.—Inspector detection rates calculated via DCE results

downward-pointing histogram depicts the raw data rates for these twenty-four (and is similar to Figure 2 which was based on all thirty-five inspectors). Notice that the upward histogram has shifted right (toward higher detection rates), as compared with the raw data; this phenomenon will occur generally in DCE because the raw data histogram implicitly assumes that all firms are noncompliant in calculating detection rates while the DCE computation allows for the possibility of compliance. The computed histogram in Figure 3 is of considerable policy interest, pointing to several OSHA inspectors with abnormally low detection rates; since the analysis has controlled for the different types of firms the inspectors are assigned, this histogram is the appropriate tool for evaluating inspector performance. The inspector fixed effects themselves for models 2, 3, and 4 are listed in Table 2, along with the industry SIC fixed effects.

The DCE estimates can be used to calculate an estimate of the rate of undetected violations. Based on the estimates of model 2, this rate is 37 percent. Thus noncompliance, both detected and undetected, does seem to be a serious problem for OSHA. The DCE method can also help illuminate Bartel and Thomas's inefficacy hypothesis. To explore the relationship between violations and injury rates, the cases were subdivided by SIC group (with only those cases in the seven industries 22, 24, 26, 34, 35, 37, and 39—corresponding to the SIC fixed effects—included) and union status into fourteen categories.[49] Within each category, two regressions were run, both using the variable LDIR, which measures each company's average injury rate during 1985, as dependent variable. The first regression included as independent variables a constant and the company's detected violation rate, which is the zero-one indicator VIOL. The second regression included as dependent variables a constant and an estimate of the company's overall violation rate, which is the sum of VIOL and the probability of an undetected violation, which is computed based on the arguments leading to equation (2.4).[50] Since the rate of violation may well be correlated with the error term (if different injury rates lead to different rates of violations, we would expect the violation rate to be an endogenous variable—see below), the model was estimated using a standard instrumental variables procedure, with the dollar value of any penalties assessed and the estimated detection rate from model 2 used as instruments.

Table 4 reports the results of this exercise.[51] As expected, the rate of violation is positively associated with injury rates in nearly all categories. However, the detected rate performs, on average, as well as the total rate, which would not be expected to be the case if the direction of causality were from violations to injuries, as these regressions presuppose. Instead, the results suggest that the opposite effect may be at work: inspectors increase their detection rates and cite more violations when they encounter a firm with a higher than average injury rate.

To explore this possibility, a further detection controlled model, model 5, is reported in Table 3, which includes LDIR in the detection equation.

[49] These are the two sets of variables that are systematically significant in the models of Table 1.

[50] In this second regression, the regression-error term is heteroscedastic because the undetected violation can only be estimated and therefore is measured with error; hence, a two-stage procedure was used to compute efficient estimates.

[51] Only those categories with nine or more cases were used.

TABLE 4

THE RELATIONSHIP BETWEEN VIOLATIONS AND INJURY RATES, DEPENDENT VARIABLE: LDIR

Industry SIC	Numbered Observations	Constant	Detected Violation Rate Used as Explanatory Variable: Violation Rate	Detected Violation Rate Used as Explanatory Variable: SSR	Total Violation Rate Used as Explanatory Variable: Constant	Total Violation Rate Used as Explanatory Variable: Violation Rate	Total Violation Rate Used as Explanatory Variable: SSR
22:							
Union	32	2.09 (2.41)	5.79 (3.96)	811	−39.7 (64.6)	48.2 (69.1)	4.43×10^6
Nonunion	9	−1.38 (6.92)	15.21 (8.17)*	470	−14.47 (27.6)	19.47 (21.3)	1.54×10^5
24:							
Union	48	3.24 (3.24)	8.74 (4.60)**	2,409	−5.50 (7.74)	8.74 (4.60)**	1.26×10^5
Nonunion	8						
26:							
Union	23	3.53 (1.98)	2.06 (3.15)	359	1.46 (5.01)	2.06 (3.16)	6,135
Nonunion	12	14.93 (14.84)	−5.74 (16.0)	655	20.7 (30.7)	−5.74 (16.0)	42,970
34:							
Union	104	2.12 (1.21)	6.96 (2.10)*	2,446	−2.23 (2.12)	8.61 (2.21)*	58,660
Nonunion	30	−11.1 (11.3)	29.0 (15.8)	4,066	−18.8 (14.9)	20.4 (10.7)**	2.03×10^5
35:							
Union	100	2.45 (.663)	2.77 (1.85)	1,479	1.44 (1.03)	4.08 (2.16)**	19,770
Nonunion	9	1.66 (2.58)	6.76 (4.94)	116	.966 (3.45)	5.52 (4.75)	1,881
37:							
Union	19	1.13 (1.29)	3.86 (3.01)	285	−.809 (2.71)	6.48 (5.09)	4,550
Nonunion	6						
39:							
Union	25	3.36 (1.92)	4.19 (2.75)	365	−.825 (4.47)	4.19 (2.75)	7,852
Nonunion	2						

NOTE.—Numbers in parentheses are standard errors.
* Statistically significant at 95 percent level.
** Statistically significant at 90 percent level.

Interestingly, the coefficient on LDIR is positive and highly significant, and the log-likelihood value is dramatically reduced below its value in any of Table 1's models. The average rate of lost workdays due to injury on the job (LDIR) is six days in the sample; according to this model, an increase of one day in this rate from its average value increases the likelihood of detection from 86 percent to 92 percent. In addition, inclusion of LDIR makes UNION's coefficient insignificant and *U*'s marginally significant in the violation equation; the inspector effects and industry dummies each remain significant. Overall, it is clear that injury rates do play an important role in the inspection process, even for these "programmed" inspections. This serves to highlight the need for further investigation of the inspection process itself, which must be left to future work.

Monte Carlo

As a second measure of DCE's practical usefulness, I have performed a series of Monte Carlo experiments.[52] From the experiments, I draw the following conclusions. For the base DCE model, estimation is accurate and reasonably efficient, even when compared with the complete information system (eq. [2.5]); this is true when the errors are correctly specified as normal, and even when they depart somewhat from normality but are assumed normal. Hence, the technique appears to be as robust to parametric specification as conventional binary choice techniques (but see the discussion of identification in Appendix A). More advanced DCE models, however, of the type discussed in the appendices, do not perform as well. In particular, models with expectations simultaneity (defined in Appendix A) can lead to biased estimates. Hence, the ability of DCE to handle more complex detection processes needs to be studied further.

IV. Conclusion

The problem of detection arises in many contexts, including regulation, income tax evasion, street crimes, fraud, and auditing. In fact, the difficulties a principal encounters in detecting violations is arguably the identifying characteristic of many of these activities. Commensurately, the problem of nondetection is the distinguishing feature of data on these activities. In this article, I have presented a general econometric model that accounts for the nondetection in the statistical analysis of noncompliance. The method incorporates the detection process directly into the

[52] The details of these experiments are reported in Feinstein, *supra* note 19, ch. 1, app. 3.

statistical model, controlling simultaneously for variations in noncompliance among potential offenders and variation in detection among monitors. As an example of how these detection controlled estimators work, I have presented a case study of OSHA safety regulation, finding evidence of substantial heterogeneity among inspectors in detection; I suspect this finding may generalize to many other contexts. Finally, in Appendix A, more complex models are developed, and the statistical issue of identification is discussed in some detail.

APPENDIX A

Statistical Models and Issues

This appendix discusses two technically more complex extensions of the basic detection controlled methodology: first, additional models; second, an examination of conditions for identification.

Section II introduced the methodology of detection controlled estimation in the form of a simple model consisting of equations (2.1) and (2.2) and the likelihood (2.3). This model, which I have labeled model 1, serves as a point of departure for the development of more complex detection controlled methods. Some of these extensions are discussed in this appendix; others must await further work.

Additional Models

The restriction of model 1 to situations in which the potential offender's decision is a simple binary choice between compliance and noncompliance (that is, that violating be "all or nothing") is not essential.

Thus, for example, if the potential offender chooses not only whether or not to commit a violation, but also how much to violate, the probit formulation of equation (2.1) becomes Tobit; Clotfelter has proposed this model for the study of income tax evasion.[53] Continuing to assume that detection is all or nothing and that the likelihood of detection is independent of the quantity of crime, it is easy to derive the appropriate likelihood:

$$\begin{aligned} L_T = & \sum_{i \in A} \log[(1/\sigma_2) f\left(\frac{Y_{1i} - x_{1i}\beta_1}{\sigma_1}\right) G(x_{2i}\beta_2)] \\ & + \sum_{i \in A^c} \log[1 - F\left(\frac{x_{1i}\beta_1}{\sigma_1}\right) G(x_{2i}\beta_2)], \end{aligned} \tag{A1}$$

where f is the density of F and possesses an identifiable variance σ_1.[54] [55] Similarly,

[53] Clotfelter, *supra* note 2.

[54] Equation (A1) is derived by integrating the probability of detection over all realizations of ϵ_{1i} for which the quantity of violation is positive. As long as detection is independent of the quantity of violation (namely of ϵ_{1i}), the integral simplifies to the displayed equation; if detection were to depend on the quantity of violation, a one-dimensional integration would have to be performed.

[55] Allowing for fractional detection would complicate this equation further.

model 1 can be extended to the case in which violations follow a Poisson process and to multiequation models.[56]

Finally, it should be noted that ϵ_{1i} and ϵ_{2i} need not be assumed independent, either in the original model 1 or in the extensions discussed above. Correlation between the errors may arise for a number of reasons, particularly when the potential offender and monitor possess information about one another (see the next subsection) or when they voluntarily choose one another as partners in execution of a performance contract. As an illustration of how to incorporate correlation into model 1, assume $(\epsilon_{1i}, \epsilon_{2i})$ to be drawn from a bivariate normal distribution with correlation ρ (the variances remained normalized to one for model 1, though not for the extensions). Certain well-known properties of the normal[57] allow the probability of observing a detected violation on case i to be computed as a one-dimensional integral of the form

$$\frac{1}{\Phi(x_{1i}\beta_1)} \times \int_{-x_{1i}\beta_1}^{+\infty} \phi(\epsilon_{1i})\Phi\left(\frac{x_{2i}\beta_2 + \rho\phi(\epsilon_{1i})/\Phi(x_{1i}\beta_1)}{1 - \rho^2\left(\epsilon_{1i}\phi\{\epsilon_{1i}/\Phi(x_{1i}\beta_1) + [\phi(-x_{1i}\beta_1)/\Phi(x_{1i}\beta_1)]^2\}\right)^{1/2}}\right)d\epsilon_{1i}, \tag{A2}$$

in which the term inside the integral refers to $\text{prob}(\epsilon_{1i} \geq -x_{1i}\beta_1)\text{prob}(\epsilon_{2i} \geq -x_{2i}\beta_2 | \epsilon_{1i})$. Calling the above expression K_i, the likelihood of the observations is

$$L_T = \sum_{i \in A} \log(K_i) + \sum_{i \in A^c} \log(1 - K_i). \tag{A3}$$

Expectations Simultaneity

Model 1 fails to recognize the interdependence between the violation and detection processes that derives from the fact that a potential offender and his monitor make decisions based in part on their expectations of the other's behavior. This sort of interdependence is denoted *expectations simultaneity;* it emerges most clearly in the game-theoretic structure underlying the compliance-detection system proposed by Graetz, Reinganum, and Wilde (see also Reinganum and Wilde), and Baron and Besanko.[58] Suppose first that each monitor forms an expectation of his potential offender's likelihood of committing a violation but that the potential offender does not form comparable expectations of monitor behavior; this is *one-sided* expectations simultaneity.[59]

Economic theory suggests that monitor i's expectation should depend on his assessment of the probability of violation, $F(x_{1i}\beta_1)$. If we pursue this tack, it seems sensible to revise (2.1) to

[56] Feinstein, *supra* note 8; Alexander & Feinstein, *supra* note 7.

[57] Heckman, *supra* note 10; Arnold Zellner, 1 Handbook of Econometrics, ch. 2 (1984).

[58] Graetz, Reinganum, & Wilde, *supra* note 6; Reinganum & Wilde, *supra* note 16; and Baron & Besanko, *supra* note 6.

[59] A symmetric model arises when the potential offender forms an expectation of monitor behavior, but not vice versa.

$$\left.\begin{aligned}
&Y_{2i} = x_{2i}\beta_2 + F(x_{1i}\beta_1)\delta_2 + \epsilon_{2i}, \quad \text{conditional on } L_{1i} = 1,\\
&L_{2i} = 1 \text{ (detection)}, \quad \text{if } Y_{2i} \geq 0,\\
&L_{2i} = 0 \text{ (no detection)}, \quad \text{if } Y_{2i} < 0,\\
&\text{maintaining}\\
&Y_{1i} = x_{1i}\beta_1 + \epsilon_{1i},\\
&L_{1i} = 1 \text{ (offense committed)}, \quad \text{if } Y_{1i} \geq 0,\\
&\text{and}\\
&L_{1i} = 0 \text{ (no offense)}, \quad \text{if } Y_{1i} < 0,
\end{aligned}\right\} \quad \text{(A4)}$$

where δ_2 is a parameter to be estimated together with β_2. System (A4) is called model 2. The parameter δ_2 is expected to be greater than or equal to zero because an increased likelihood of violation will normally call forth increased enforcement. This system is not truly simultaneous because L_{1i} itself does not enter the right-hand side of the equation for Y_{2i}, but only the expectation of L_{1i}, which in this case is also its probability of equaling one since it is an indicator (zero-one) variable—this is in contrast to dummy simultaneous equation systems.[60] The pair of equations in (A4) determining L_{1i} and L_{2i} are best described as recursive.

Adding the variable $F(x_{1i}\beta_1)$ to the first equation attributes rational expectations to the monitor because it assumes that she knows the true probability of potential offender i committing a violation. More to the point, adding $F(x_{1i}\beta_1)$ assumes that each monitor knows the particular characteristics x_{1i} of the potential offender she monitors, as well as the true parameter vector β_1. To emphasize this viewpoint let us introduce a new variable, H_{2i}, which represents monitor i's assessment of the probability of violation:

$$H_{2i} = E[L_{1i} | \text{information known to monitor } i].$$

The more general form for Y_{2i} in (A4) is then

$$Y_{2i} = x_{2i}\beta_2 + H_{2i}\delta_2 + \epsilon_{2i}, \qquad \text{(A5)}$$

within which the particular case of (A4) is $H_{2i} = E[L_{1i}|x_{1i}, \beta_1] = F(x_{1i}\beta_1)$.

Now suppose that monitors do not know the specific characteristics of the potential offenders they monitor. In this case, H_{2i} will be constant across i, meaning that all monitors assess the probability of violation identically. The term $H_{2i}\delta_2$ is then constant across i and is absorbed into the constant term of x_{2i}. Equation (A5) reverts exactly to the original form of model 1. We conclude that model 1 is the appropriate specification when monitors do not know the specific characteristics of the potential offenders they monitor, while model 2 is appropriate when they do.

The log likelihood based on (A4) (allowing for the more general form of [A5]) is

$$\begin{aligned} L_T(\beta_1, \beta_2, \delta_2) = &\sum_{i\in A} \log[F(x_{1i}\beta_1)G(x_{2i}\beta_2 + H_{2i}\delta_2)]\\ &+ \sum_{i\in A^c} \log[1 - F(x_{1i}\beta_1)G(x_{2i}\beta_2 + H_{2i}\delta_2)]. \end{aligned} \qquad \text{(A6)}$$

[60] See James Heckman, Dummy Endogenous Variables in a Simultaneous Equation System, 46 Econometrica 931 (1978).

Monitors' information about their potential offenders may not fit into either model 1 (no information) or model 2 (information identical to the econometrician's). If monitors possess less information than the econometrician but still some, H_{2i} is defined as

$$H_{2i} = \int_{x_{1i(-k)}} F(x_{1i}\beta_1)dR(x_{1i(-k)}), \tag{A7}$$

where x_{1ik} constitutes the information known to the monitors, and R is the marginal distribution function of $x_{1i(-k)}$. If monitors possess more information than the econometrician, they are likely to have some knowledge of ϵ_1, which leads to the problem of correlated errors discussed above.

A natural extension of model 2 arises when both the potential offender and the monitor possess information about one another. We can generalize model 2 to a new model, labeled model 3, which includes terms reflecting each individual's assessment of the other's L_i function. Define

$$\left.\begin{aligned} Y_{1i} &= x_{1i}\beta_1 + H_{1i}\delta_1 + \epsilon_{1i}, \\ Y_{2i} &= x_{2i}\beta_2 + H_{2i}\delta_2 + \epsilon_{2i}, \\ L_{1i} &= 1 \quad \text{if } Y_{1i} \geq 0, \\ L_{2i} &= 1 \quad \text{if } Y_{2i} \geq 0, \quad \text{conditional on } L_{1i} = 1, \\ L_{1i} &= 0 \quad \text{if } Y_{1i} < 0, \\ L_{2i} &= 0 \quad \text{if } Y_{2i} < 0, \end{aligned}\right\} \tag{A8}$$

where

$$\left.\begin{aligned} H_{1i} &= E[L_{2i}|x_{1i}, x_{2i}, \beta_1, \beta_2, \delta_1, \delta_2] = G(x_{2i}\beta_2 + H_{2i}\delta_2), \\ H_{2i} &= E[L_{1i}|x_{1i}, x_{2i}, \beta_1, \beta_2, \delta_1, \delta_2] = F(x_{1i}\beta_1 + H_{1i}\delta_1). \end{aligned}\right\} \tag{A9}$$

Equations (A9) define a pair of algebraic equations in H_{1i} and H_{2i}. Substituting yields

$$H_{1i} = G[x_{2i}\beta_2 + F(x_{1i}\beta_1 + H_{1i}\delta_1)\delta_2], \tag{A10}$$

a single equation in the single unknown H_{1i}; a comparable equation may be derived for H_{2i}. We expect δ_2 to be greater than or equal to zero since monitors will devote more resources to detection when a violation is more likely and δ_1 to be less than or equal to zero since increased enforcement deters violations. The following lemma ensures that H_{1i} and H_{2i} are properly defined by (A9).

LEMMA A1. (i) Given any set of values for x_{1i}, x_{2i}, β_1, β_2, δ_1, δ_2, a solution to (A9) (via [A10]) exists for H_{1i} (H_{2i}).

(ii) When $\delta_1 \leq 0$ and $\delta_2 \geq 0$, this solution is unique.

The log likelihood associated with model 3 can be written

$$L_T(\beta_1, \delta_1, \beta_2, \delta_2) = \sum_{i \in A} \log(H_{1i}H_{2i}) + \sum_{i \in A^c} \log(1 - H_{1i}H_{2i}). \tag{A11}$$

When potential offenders observe only the subset x_{2k} and monitors the subset x_{1j} of variables, H_{1i} and H_{2i} must be redefined accordingly, which leads to

$$H_{1i} = \int_{x_{2i(-k)}} G\left[x_{2i}\beta_2 + \delta_2 \int_{x_{1i(-j)}} F(x_{1i}\beta_1 + \delta_1 H_{1i}) dR_1(x_{1i(-j)})\, dR_2(x_{2i(-k)})\right] \quad \text{(A12)}$$

as the definition of H_{1i}.

Identification

As was discussed briefly in Section II, identification of parameters is an important statistical issue in detection controlled models. As was also discussed there, the difficulty in identifying parameters arises because the data list only the joint-category detected violations, which must be decomposed along the two analytic dimensions of violation and detection. To apply successfully DCE requires being able to ascertain whether a collection A of cases for which detected violations are low is characterized by high compliance rates or low detection. Speaking loosely, identification relies on x_1 and x_2 varying somewhat independently of one another (though of course they may be correlated), so that the collection A may be compared to two other collections: a set B on which x_1 is relatively similar to that on A (similar potential offenders) while x_2 differs; and a set C on which x_2 is similar (similar monitors) while x_1 differs.[61] This argument suggests (as was discussed in Section II) that we must be especially cautious of parameter estimates for x variables included in both x_1 and x_2; in fact, when x_1 and x_2 are identical, identification fails.

When the distributions F and G are assumed to belong to known parametric families, such as the normal, the DCE models are generally identified (by the nonlinear curvature of the normal or other parametric family) under suitable regularity conditions—except for the special case in which F and G are each exponential, a point that was discussed in Section II and to which I will return shortly.[62] Since the error distributions are typically unknown, however, it is of interest to determine whether identification derives solely from these parametric assumptions or whether the models could be identified and estimated under "semiparametric" conditions.

To explore this question, it is worthwhile first reviewing what is known about identification of the semiparametric binary choice model. In that model, the dependent variable Y has probability $F(x\beta)$ of equaling one, and $1 - F(x\beta)$ of equaling zero. The distribution function F is assumed to be strictly increasing and to possess a continuous density f.[63] Manski, Ichimura, and others have discussed identification in this model.[64] In all cases, β is only identified up to a constant (that is, the intercept is not determined uniquely) and a scalar multiple. If the x's are allowed to be both continuous and discrete, the condition for identification is strong: in addition to the requirements on F above, there must be at least one

[61] This is only speaking loosely; formal proofs of conditions for identification are presented below.

[62] Feinstein, *supra* note 19.

[63] The model is semiparametric because $x\beta$ is assumed to have a linear form with parameters β; since quadratic and higher-order terms may be included as new explanatory variables, this is not a strong restriction.

[64] Charles Manski, Semiparametric Analysis of Discrete Response, 27 J. Econometrics 313 (1985); Hide Ichimura, Estimation of Single Index Models (unpublished report, Massachusetts Institute of Technology, Dep't Economics, 1987).

continuous x that possesses unbounded support (it ranges over $(-\infty, +\infty)$). If no x possesses unbounded support, the β's are bounded but not fully identified.[65] If all x's are continuous, identification requires just the restrictions on F.

Now consider the basic DCE model 1, $F(x_1\beta_1)G(x_2\beta_2)$. Assume that F and G are everywhere strictly increasing and possess continuous densities f and g. From the review of the semiparametric binary choice model above, it is clear that β_1 and β_2 can at best be identified up to a constant and a scalar multiple.

As long as each of x_1 and x_2 possesses at least one continuous component (which differs from one another), the condition for identification is that there exists a point (x_1^*, x_2^*), a neighborhood of which possesses positive x density, for which

$$F(x_1^*\beta_1)G(x_2^*\beta_2) \neq F_0(x_1^*\beta_{10})G_0(x_2^*\beta_{20}) \tag{A13}$$

for each possible collection F, G, β_1, and β_2 of candidate values, where F_0, G_0, β_{10}, and β_{20} are the true values.

Let us first consider the simpler case in which the x's contain only continuous components (other than the intercepts). Assume first that x_1 contains no components in common with x_2 (again, other than intercepts). Since all x's are continuous, condition (A13) may be differentiated. Fix x_1. Differentiating with respect to x_{2k}, we see that

$$\frac{\beta_{2k}}{\beta_{2k0}} = c\frac{g_0(x_2^*\beta_{20})}{g(x_2^*\beta_2)}, \tag{A14}$$

where c equals F_0/F, and g and g_0 are the densities of G and G_0. Differentiating (A13) instead with respect to x_{2j} yields the analogous equation $\beta_{2j}/\beta_{2j0} = cg_0(x_2^*\beta_{20})/g(x_2^*\beta_2)$. Since c, $x_2^*\beta_2$, and $x_2^*\beta_{20}$ are constant, this shows that all components of β_2 are identified up to scale (and implicitly up to a constant). A similar argument applies to β_1. Note, however, that β_1 and β_2 are not identified relative to one another since c is arbitrary and not the same for β_1 and β_2 (c is F_0/F, whereas G_0/G is $1/c$; c need not equal one). The logic behind this proof is straightforward: since x_1 and x_2 are disjoint, we may hold detection constant and vary compliance—essentially the model reduces to a binary choice over violation—and, alternatively, hold compliance fixed and vary detection; this thought experiment is replicated in a sufficiently large data set, allowing the violation and detection processes to be separated from one another.

Adding expectations to the model (moving to models 2 and 3) does not change the basic argument. In fact, with two-sided expectations simultaneity, condition (A14) can be shown to be

$$\frac{\beta_{2k}}{\beta_{2k0}} = \frac{(1 - fg\delta_1\delta_2)(g_0F_0 + f_0g_0\delta_{10}G_0)}{(1 - f_0g_0\delta_{10}\delta_{20})(gF + fg\delta_1G)}. \tag{A15}$$

Since the right-hand side of (A15) is the same for all components k, β_2 is again identified up to a scalar multiple; similarly, β_1 is identified up to a scalar multiple. However, since F and G are not identified (only FG is identified), the expectations terms δ_1 and δ_2 are not identified.

[65] As the support grows, the bound tightens.

Next, consider the case in which all of the x's are continuous and x_1 and x_2 contain some components in common. We know from the discussion in Section II that, when the true forms of F and G are each exponential, identification of the β's associated with overlapping x components breaks down. In fact, we can prove:

THEOREM A1. Assume that x_1 and x_2 each contain only continuous components (apart from intercepts) and that they have some elements in common. If identification fails, F_0 and G_0 must each belong to the exponential family.

The conclusion is that the exponential case is unique.

In the more general case, the x's possess both discrete and continuous components. As it turns out, identification is essentially similar to the semiparametric binary choice case, as demonstrated by theorem A2.

THEOREM A2. Assume that each of x_1 and x_2 possesses at least one continuous component with unbounded support and that at least one of each of these components enters only into x_1 and only into x_2. Then β_1 and β_2 in DCE model 1 are each identified up to a constant and a scalar multiple, but the scalar multiples may differ.

Intuitively, if some monitors are close to perfect (the unbounded component, say, k, of x_2 becoming either very positive or negative so that $x_{2k}\beta_{2k}$ becomes very positive), we may use their cases to identify the compliance equation parameter β_1; the remaining monitors' detection rates can then be determined by scaling up their raw detection rates commensurately. Notice also that the distributions F and G are also fully identified, allowing estimation of the undetected violation rate.

This theorem extends with little modification to models 2 and 3:

COROLLARY A. (i) Under the same conditions as theorem A1, model 2's (with the potential offender possessing information about the monitor, but not vice versa) parameters β_1, β_2, and δ_1 are identified.

(ii) Under the same conditions as theorem A1, model 3's parameters β_1, β_2, δ_1, and δ_2 are identified.

Semiparametric estimation of the various DCE models is not discussed here. In general, it is likely that a number of semiparametric binary choice estimators, such as Manski's maximum score estimator, can be modified to suit the DCE model.[66]

APPENDIX B

PROOFS OF THEOREMS

Proof of theorem 2.1, (i) *and* (ii). The misspecified binary choice model has likelihood

$$L_T = \sum_{i \in A} \log[F(x_{1i}\beta_1)] + \sum_{i \in A^c} \log[1 - F(x_{1i}\beta_1)].$$

Assume that the density of x_1 is $h_1(x_1)$, independent of the density of x_2, which is $h_2(x_2)$; h_2 may depend on x_1 (conditionally) when x_1 and x_2 are correlated. Asymptotically (when L_T converges to the true likelihood L), the first-order condition

[66] Manski, *supra* note 64.

that determines β_{1k} is

$$\frac{\partial L}{\partial \beta_{1k}} = \int_0^{+\infty} x_{1k} \int_{-\infty}^{+\infty} f(x_1\beta_1) \frac{F(x_1\beta_{10}) \int_{-\infty}^{+\infty} G(x_2\beta_{20})h(x_2|x_1)dx_2 - F(x_1\beta_1)}{F(x_1\beta_1)[1 - F(x_1\beta_1)]} \times h_1(x_1)dx_{1k}dx_{1(-k)},$$

where f is the density of F, and the fact that x_{1k} is nonnegative has been used in defining the limits of the first integral. Since F is globally concave for every data sequence (represented by the set A), it is globally concave over the expected value of these data sequences. Hence, the sign of the bias in b_{1k} will be the same as the sign of $\partial L/\partial\beta_{1k}$ evaluated at the true β_{10}. Set $d(x_1) = E[G(x_2\beta_{20})|x_1]$. It is easy to see that, when $d(x_1)$ equals one everywhere (complete detection), the bias is zero; when $d(x_1)$ is less than one over any x_1 set of nonzero h_1 measure, the right-hand side above becomes negative; hence, the bias is negative, and b_{1k} will be biased downward.

For (ii), let $d_0(x_1)$ equal the original level of detection. When detection rates fall, $d_0(x_1)$ falls to some $d_1(x_1)$ less than or equal to $d_0(x_1)$ (and strictly less on some x_1 set of positive h_1 measure) for each x_1. Let β_1^* be the value of β_1 that maximizes the misspecified likelihood when detection is $d_0(x_1)$; thus, $\partial L/\partial\beta_{1k}(\beta_1^*)$ equals zero at detection rates $d_0(x_1)$. It then follows that, at detection rates $d_1(x_1)$, the term inside the two x_1 integrals is always more negative, from which it follows that $\partial L/\partial\beta_{1k}(\beta_1^*)$ is now negative. Since F is concave, an argument just like that used in (i) demonstrates that β_{1k} is now biased even further downwards than β_1^*.

Proof of theorem 2.2. Consider again theorem 2.1. When x_{1k} and x_{1j} are orthogonal zero-one indicator variables that are uncorrelated with the other x_1's, possess the same $d(x_1)$'s, and $\beta_{1k0} = \beta_{1j0}$, it follows that the estimates of β_{1k} and β_{1j} will each possess the same downward bias, so that the ratio $\beta^*_{1k}/\beta^*_{1j}$ will be one, as it should be. When $d_k(x_1)$ is no larger than $d_j(x_1)$ (and strictly smaller over some x_1 set of positive h_1 measure), $\partial L/\partial\beta_{1k}$ will be biased downward when evaluated at $\beta_1^* = (\beta^*_{1(-k,-j)}, \beta^*_{1j}, \beta^*_{1j})$ (that is, at $\beta^*_{1k} = \beta^*_{1j}$); hence, the ratio $\beta^*_{1k}/\beta^*_{1j}$ will now be biased downward.

Proof of lemma A1. (i) Recall that $H_{1i} = E[L_{2i}|x_{1i}, x_{2i}, \beta_1, \beta_2, \delta_1, \delta_2]$. H_{1i} solves the following recursive equation:

$$H_{1i} = G[x_2\beta_2 + \delta_2 F(x_1\beta_1 + \delta_1 H_{1i})].$$

For arbitrary values of $x_1, x_2, \beta_1, \beta_2, \delta_1, \delta_2$, the following facts hold. When $H_{1i} = 0$, the right-hand side of the above equation becomes $G[x_2\beta_2 + \delta_2 F(x_1\beta_1)]$, which is strictly positive. When $H_{1i} = 1$, the right-hand side becomes $G[x_2\beta_2 + \delta_2 F(x_1\beta_1 + \delta_1)]$, which is strictly less than one under the assumptions. Since both the left-hand side and the right-hand side are continuous functions of H_{1i}, the two must cross somewhere in the interior of the interval [0, 1], and therefore a solution (which is interior) always exists.

(ii) Assume $\delta_1 \leq 0$ and $\delta_2 \geq 0$. Then the derivative of the right-hand side in the above equation with respect to H_{1i} is $gf\delta_1\delta_2$, which is less than or equal to zero. Since an increasing function and a decreasing function can only intersect once, the result follows. Identical arguments apply to H_{2i}.

Proof of theorem A1. Identification fails if

$$F(x_1\beta_1 + \theta_1 z)G(x_2\beta_2 + \theta_2 z) = F_0(x_1\beta_{10} + \theta_{10}z)G_0(x_2\beta_{20} + \theta_{20}z)$$

over all measurable sets, where the notation refers to the overlapping component z and the remaining nonoverlapping components, x_1 and x_2. For the components β_1 and β_2 of nonoverlapping x's, differentiation yields

$$\frac{\beta_1}{\beta_{10}} = \frac{G_0 f_0}{Gf} \quad \text{and} \quad \frac{\beta_2}{\beta_{20}} = \frac{F_0 g_0}{Fg}.$$

Differentiating with respect to z,

$$fG\theta_1 + Fg\theta_2 = f_0 G_0 \theta_{10} + F_{0g}\theta_{20},$$

$$\frac{fG}{F_0 g_0}\theta_1 + \frac{Fg}{F_0 g_0}\theta_2 = \frac{G_0 f_0}{F_0 g_0}\theta_{10} + \theta_{20},$$

$$\frac{G_0 f_0}{F_0 g_0}\frac{\beta_{10}}{\beta_1}\theta_1 + \frac{\beta_{20}}{\beta_2}\theta_2 = \frac{G_0 f_0}{F_0 g_0}\theta_{10} + \theta_{20},$$

or

$$\frac{G_0 f_0}{F_0 g_0}\theta_1\left(\frac{\theta_{10}}{\theta_1} - \frac{\beta_{10}}{\beta_1}\right) = \theta_2\left(\frac{\theta_{20}}{\theta_2} - \frac{\beta_{20}}{\beta_2}\right).$$

If $\theta_{10}/\theta_1 = \beta_{10}/\beta_1$ and $\theta_{20}/\theta_2 = \beta_{20}/\beta_2$, then the θ's are identified up to the same scalar multiple as the β's. Further, the equations above guarantee that, if $\theta_{20}/\theta_2 = \beta_{20}/\beta_2$, then $\theta_{10}/\theta_1 = \beta_{10}/\beta_1$, and vice versa. Finally, suppose neither $\theta_{10}/\theta_1 = \beta_{10}/\beta_1$ nor $\theta_{20}/\theta_2 = \beta_{20}/\beta_2$. Then, since the θ's and β's are all constant, the above condition implies that $G_0 f_0/F_0 g_0$ is constant over variations in x and z. Varying just x_1, G_0/g_0 remains constant; hence, f_0/F_0 is constant. Similarly, varying x_2 shows that g_0/G_0 is constant. But then f_0 must satisfy the differential equation $d(\log F_0) =$ a constant, which implies that F_0 has the form $F(w) = h_f e^{-q_f w}$. Similarly, G_0 must have the exponential form $G(w) = h_g e^{-q_g w}$. Hence, failure of identification implies the exponential form for both F and G.

Proof of theorem A2. Fix x_1. If $F(x_1\beta_1) = F_0(x_1\beta_{10})$, then the condition (A13) becomes the usual binary choice condition of identification for $G(x_2\beta_2)$, in which case the conditions are sufficient to ensure identification. If not, then two cases are possible. Case 1:

$$F_0(x_1\beta_{10})/F(x_1\beta_1) = 1 + \alpha, \quad \alpha > 0. \tag{B1}$$

Let x_{2k} be the component of x_2 that possesses unbounded support, and suppose, without loss of generality, that $\beta_{2k0} > 0$. Then there exists a μ such that, for all $x_{2k} > \mu$, $G_0(x_2\beta_{20}) > 1/(1 + \alpha)$. Hence, for all $x_{2k} > \mu$, $G/G_0 < 1/G_0 < 1 + \alpha$. Hence, $F/F_0 = 1 + \alpha$, while $G/G_0 < 1 + \alpha$, so that condition (A13) holds. Now choose a second x_1' close enough to the original x_1 such that (B1) still holds (this is possible by continuity), and find a μ' for which G_0 is $> 1 + \alpha'$. Finally, set $\mu^* = \max(\mu, \mu')$, and over the region (x_1, x'_1) and (μ^*, ∞) (the other components of x_2 are also varied a bit), condition (A13) holds, which guarantees identification of $F(x_1\beta_1)$ up to a constant and a scalar multiple. Otherwise, case 2:

$$F_0(x_1\beta_{10})/F(x_1\beta_1) = 1 - \alpha, \quad \alpha > 0.$$

In this case, choose μ such that $G(x_2\beta_2)$ is $> 1 - \alpha$ (this must be possible unless β_{2k} is zero, in which case choose x_{2k} such that $1/G_0 > (1 - \alpha)G$, where G is fixed since β_{2k} is zero) and, therefore, $G/G_0 > 1 - \alpha$. The same argument as that used for case 1 now applies.

Proof of corollary A1. (i) Condition (A13) is now

$$F[x_1\beta_1 + \delta_1 G(x_2\beta_2)]G(x_2\beta_2) \neq F_0[x_1\beta_{10} + \delta_{10}G_0(x_2\beta_{20})]G_0(x_2\beta_{20}). \quad \text{(B2)}$$

Again there are two cases. Case 1:

$$F_0(x_1\beta_{10} + \delta_{10})/F(x_1\beta_1 + \delta_1) = 1 + \alpha, \quad \alpha > 0.$$

Choose $\alpha' > 0$ and find γ such that, for all γ' between γ and one, $F_0(x_1\beta_{10} + \delta_{10}\gamma')/F(x_1\beta_1 + \delta_1\gamma') \geq 1 + \alpha'$. Then find μ' such that G and G_0 are both greater than γ for all $x_{2k} > \mu'$ (if β_{2k} is of opposite sign to β_{2k0}, simply replace δ_1 by $-\delta_1$, and the same argument applies). Next find μ'' such that $G_0 > 1/(1 + \alpha')$ for all $x_{2k} > \mu''$, such that $G/G_0 < 1 + \alpha'$ for all $x_{2k} > \mu''$. Finally, choose $\mu^* = \max(\mu', \mu'')$. It then follows that (B2) holds for all $x_{2k} > \mu^*$. The remainder of the argument for case 1 and the argument for case 2 now follow just as for theorem A1.

(ii) Condition (B2) now becomes

$$\begin{aligned} F[x_1\beta_1 + \delta_1 G(x_2\beta_2 + \delta_1 F)]G[x_2\beta_2 + \delta_2 F(x_1\beta_1 + \delta_1 G)] \\ \neq F_0(x_1\beta_{10} + \delta_{10}G_0)G_0(x_2\beta_{20} + \delta_{20}F_0). \end{aligned} \quad \text{(B3)}$$

Case 1 is now

$$F_0(x_1\beta_{10} + \delta_{10})/F(x_1\beta_1 + \delta_1) = 1 + \alpha, \quad \alpha > 0.$$

We now follow exactly the same argument as was used in the case of part (i) to choose μ', μ'', and μ^*, thereby demonstrating that (B3) holds. The added subtlety is that the determination of F and G is joint. However, the effect of G on F is bounded by δ_1, and that of F on G by δ_2; continuity plus this boundedness ensures that the μ's exist.

APPENDIX C

Example of a Behavioral Model for Potential Offenders

The reasons for deriving noncompliance equations such as (2.1) from a behavioral model are at least fourfold. (The same reasoning applies also to detection equations such as [2.2].) First, a theoretical model helps identify the factors that should generally be included in the variables x_1. Second, such models typically produce nonlinear equations that more fully reflect the underlying decision process than the linear approximation used in (2.1); these equations may improve model fit and the interpretation of results. Third, behavioral theories draw attention to the possible interdependence of the violation decision and the detection process, which can lead to models of the type presented in Appendix A. Finally, there is the familiar point that structural models allow analysis of the potential behavioral response to alternative policies regarding, for example, penalties or the resources devoted to detection. Bearing these points in mind, we proceed to derive a structural model of the compliance decision, drawing on the extensive prior literature that includes, among others, Becker, Ehrlich, Block and Heineke, and Allingham and Sandmo (who study the particular case of income tax evasion), and fits more generally within the category of decision making under conditions of uncertainty.[67]

Suppose that the potential offender possesses a concave utility function $U(\ )$ and baseline wealth W. If the potential offender remains legal, he earns the monetary equivalent of z, which is referred to as his legal-sector opportunity. If instead he chooses to violate the law, he earns a monetary equivalent of h if he escapes detection but has to pay the monetary equivalent (in fines and/or jail sentences) of e (and does not earn h) if detected. If the potential offender's subjective assessment of the probability of detection is p, he will commit a violation whenever

$$(1 - p)U(W + h) + pU(W - e) > U(W + z). \tag{C1}$$

Equation (C1) can be extended to situations in which potential offenders choose not only whether or not to commit a violation but also how much, as in the choice of how much income tax to evade; related statistical models were discussed in Appendix A.

In order to transform (C1) into an estimable equation, we must introduce a collection of explanatory socioeconomic variables, x_1; a stochastic disturbance, ϵ_1; and a parametric form for the utility function. A particularly convenient way of introducing x_1 and ϵ_1 into (C1) is to specify the potential offender's assessment of p to depend on these variables. Since p must be bounded between zero and one, let us introduce the cumulative distribution function S and define p by $p = S(-x_1\beta_1 - \epsilon_1)$ ($-x_1$ is used so that, assuming β_{1k} is positive, p falls and the likelihood of a violation rises when x_{1k} increases). The dependence of p on x_1 and ϵ_1 arises principally from two sources: (1) the relationship of x_1 to the detection effort expended on this particular individual, for example, when an element of x_1 signals monitors of an increased likelihood of violation; and (2) differences in the prior experiences and beliefs of individuals. The computation of (C1) is particularly straightforward when S is the logistic distribution, in which case $p = 1/(1 + e^{x_1\beta_1 + \epsilon_1})$. Two convenient parametric families for U are the constant absolute risk aversion, $U(x) = -e^{-\alpha x}$, for which α is the coefficient of absolute risk, and the constant relative risk aversion, $U(x) = x^{\alpha}$, for which $(1 - \alpha)$ is the coefficient of relative risk. As an example of the transformation, suppose that U belongs to the absolute risk family. Then if S is logistic, the potential offender commits a violation whenever

$$\epsilon_1 > -x_1\beta_1 - \log\left(\frac{e^{\alpha e} - e^{-\alpha h}}{e^{\alpha z} - e^{-\alpha h}} - 1\right),$$

which leads to a nonlinear analogue of (2.1) of the form

$$Y_1 = x_1\beta_1 + \log\left(\frac{e^{\alpha e} - e^{-\alpha h}}{e^{\alpha z} - e^{-\alpha h}} - 1\right) + \epsilon_1.$$

Equation (C1) and its derivation suggest that the potential offender's decision will depend on, among other things, his returns to breaking the law, h; the penalty, e;

[67] Becker, *supra* note 12; Ehrlich, *supra* note 32; Block & Heineke, *supra* note 12; M. B. Allingham & A. Sandmo, Income Tax Evasion: A Theoretical Analysis, 1 J. Pub. Econ. 323 (1972).

his legal opportunity, z; and any variables which affect his assessment of p and are included in x_1. Of the four variables h, e, z, and W (which does not enter [C2] but does enter the corresponding relative risk equation), which can enter the nonlinear term in (C2), W, e, and z are commonly observed, while h is observed only for violators who are detected; therefore, h must itself be specified as depending on the individual's characteristics and the economic environment.

Also to be included in x_1 is any information the individual possesses about the monitor he has been assigned and the detection process he faces. These are the variables that provide the main point of dependency of the violation decision on detection and illustrate the game-theoretic basis of this interaction. Structural models that incorporate expectations about detection will not typically belong to the linear rational expectations models presented in Appendix A.

The logic underlying the derivation of (C2) must be modified when one studies regulatory noncompliance, for which it is better to think of the firm as choosing, ex ante, the level of care to devote to compliance. While the choice of care will reflect the rational calculation we have outlined above, violations themselves will arise stochastically—more violations arising when the level of care is either chosen to be low or turns out to be low after the fact.[68] Such a "take care" model would presumably link this article's statistical models to the large literature on accident prevention.

[68] See Feinstein, *supra* note 8, for such a model applied to safety violations at nuclear power plants.