# The Safety Regulation of U.S. Nuclear Power Plants: Violations, Inspections, and Abnormal Occurrences

Jonathan S. Feinstein

*Stanford University*

Data from more than 1,000 inspections by the Nuclear Regulatory Commission form the basis for an investigation into the nature of safety regulation at U.S. commercial nuclear reactors. Poisson (and binary choice) models of the rate of occurrence of violations during each inspection period are specified and are extended to control for nondetection and for the possibility that violations persist from one inspection to the next. These models are used to study the factors associated with noncompliance, relative rankings of plants according to propensity to violate, the variation in detection rates among NRC inspectors, and the relationship between undetected violations and abnormal occurrences.

## I. Introduction

The safety of nuclear power is a matter of considerable debate. Public concern about safety has forced a number of power plants to shut down and has led to dramatic increases in construction and operating costs at others, creating financial distress at several public utilities. Because safety has the characteristics of a public good, in the sense that power plant operators' incentives to ensure safety may be less than is socially desirable, regulation is central to controlling risks: the

I acknowledge graduate student support from the National Science Foundation, research assistance from Edward Keating, and helpful comments from the editor, the referee, Alessandra Casella, Franklin Fisher, Paul Joskow, Daniel McFadden, Roger Noll, and Barbara Sands, as well as seminar participants at the University of California at Berkeley, the University of Chicago, Harvard University, and Stanford University.

[*Journal of Political Economy*, 1989, vol. 97, no. 1]

Nuclear Regulatory Commission (NRC) is charged with formulating and enforcing safety standards at nuclear reactors.

This paper presents a statistical analysis of NRC safety regulation of commercial operating reactors in the United States based on data from more than 1,000 NRC inspections at 17 power plants over 3 years. The study focuses on the factors associated with noncompliance, the ability of NRC inspectors to detect violations, and the extent to which previously undetected violations are associated with the incidence of abnormal occurrences. To address these issues, I specify models in which the dependent variable is the number of violations cited during the inspection, assumed to be the realization of a Poisson process whose parameter depends on a number of plant characteristics, including financial status, technology, plant fixed effects, and in certain cases a random unobservable effect. Lying behind this specification is the view that the Poisson parameter summarizes plant management's choice of the level of resources to devote to achieving compliance: fewer resources lead to a larger parameter and a greater likelihood of violations. Since many inspections (approximately two-thirds) lead to no citations, binary choice models in which the dependent variable simply records whether or not at least one violation was cited are also estimated, as a check on sensitivity to outliers (inspections in which an unusually large number of violations are cited).

A number of statistical issues arise in using the inspections data to study power plant noncompliance. Of central importance is the problem of nondetection, which arises because NRC inspectors may not detect all violations at a plant. If the NRC were able to detect all violations, ensuring compliance would be much simpler. Conceptually, however, the detection problem is integral to the regulatory process: power plant management has an incentive to conceal violations that it commits intentionally, while in other cases management may not even know about violations it has been unwilling to expend the necessary resources to discover. To control for the possibility that some violations remain undetected requires specifying a detection equation, in which the probability that a violation will be detected depends on the characteristics of the particular team assigned to the inspection. Section II derives models that incorporate the detection process into the analysis and discusses their statistical properties; they are discussed in more detail in Feinstein (1987). Not only do these "Detection Controlled" models correct for biases implicit in models that erroneously assume complete detection, but they allow one to address important policy questions about the NRC's inspection program and provide an estimate of the rate of undetected violations.

A second statistical issue that arises is the extent to which violations persist over time. Persistence is important for several reasons. On the one hand, understanding how long violations persist has policy implications for the frequency and intensity of inspections. On the other hand, violations that remain undetected during one inspection may persist and be detected during a subsequent inspection; if ignored, this phenomenon can bias statistical estimates. Models that incorporate persistence flow quite naturally from the Poisson specification and are presented below.



The empirical analysis, presented in Sections IV and V, explores several important policy issues. Of primary interest are the characteristics associated with noncompliance. On the basis of the estimates obtained, financial distress, as measured by the power plant principal owner's bond rating, has little tendency to increase noncompliance. Similarly, past NRC sanctions against either a particular plant or the industry at large have little impact on noncompliance. Together, these findings call into question whether economic incentives significantly affect power plant safety. In contrast, a number of plant technological and operating characteristics, such as whether the plant is a boiling or pressurized water reactor, are found to be significantly related to noncompliance, at least in some models. Further, plant fixed effects are jointly significant in all the models, and the relative rankings of the plants by estimated propensity to noncomply accord with public perceptions. Thus overall management style and idiosyncratic technology (which cannot be readily distinguished from one another) appear to be important determinants of noncompliance, highlighting the substantial heterogeneity among U.S. power plants, particularly as compared with other countries.

The Detection Controlled models allow me to investigate the factors associated with variation in detection rates. An important finding is that NRC inspectors differ substantially in their detection of violations; in fact the variation in detection rates is comparable with the variation in noncompliance rates among plants and remains significant in models that include the plant effects (which themselves remain significant). Detection rates also increased quite sharply after the accident at Three Mile Island in 1979.

The final policy issue explored is the effectiveness of NRC regulations. If compliance is important in reducing safety risks, undetected violations should increase these risks. To test this proposition, regression models are estimated in which the dependent variable is the number of abnormal occurrences at the power plants each month, and the independent variables include an estimate of undetected violations in the recent past, as well as the lagged number of events from the previous month and plant effects. The rate of undetected violations is found to be positive, and in one of the models statistically significant, supporting the view that the regulations "matter" even after plant effects have been taken into account.

This paper's approach shares much in common with the probabilistic risk assessment (PRA) technique that nuclear engineers have developed to quantify risk (see the volume edited by Cullingford, Shah, and Gittus [1987]). Application of PRA to nuclear power entails the building of "event trees" whose branches represent particular reactor system components. The technique quantifies risk by assigning failure probabilities to each branch and tracing each possible "path" through the tree that can lead to a serious accident. Implicit in this approach is the view that serious accidents are most likely to arise not from the breakdown of a single major component but from the compound (sequential) failure of several lesser components. This emphasis on the compounding of lesser violations motivates the statistical analysis of inspections and violations data. In fact the Poisson models of this paper represent the simplest statistical framework capable of capturing the principle of compounding violations: plants with larger Poisson parameters are more prone to multiple violations and serious accidents. A more sophisticated analysis would map specific violations to specific branches on a reactor's event tree, which would allow one to calculate each violation's marginal contribution to overall risk.

The PRA methodology dates back to the *Reactor Safety Study* report (U.S. Atomic Energy Commission 1974) and was lent strong support by the accidents at Three Mile Island and Chernobyl, both of which arose from a sequence of operator errors and hardware failures similar to those modeled in event trees (see President's Commission on the Accident at Three Mile Island 1979; International Nuclear Safety Advisory Group 1986; Megaw 1987). In fact the Three Mile Island accident sequence was virtually identical to one event path mentioned in the *Reactor Safety Study* (see Konstantinov 1987). As a reflection of these experiences, the NRC has begun to incorporate PRA into its thinking. Thus the growth in what Nichols and Wildavsky (1987) have labeled "detailed prescriptive regulation" represents a shift in regulatory focus from single catastrophic breakdowns—for example, pipe ruptures—to smaller interrelated system component failures. Statistical models such as those presented below provide an additional, empirical, linkage between PRA and regulatory oversight.

The analysis also relates to a second literature, the recent theoretical work (see Baron and Besanko 1984; Laffont and Tirole 1986) on the role of asymmetric information and auditing in regulation. These papers represent a shift in focus from previous theories of regulation. Instead of evaluating preexisting regulatory practices, they study the design of optimal regulatory regimes within environments characterized by information and resource limitations. As with PRA, the statistical methods I develop represent a first step in the empirical calibration of optimally designed regulatory systems. Specifically, the Detection Controlled models capture in a natural way the inspection process through which the NRC interacts with power plant management, and the regressions relating undetected violations to future abnormal occurrences provide a means of evaluating regulatory effectiveness. Ultimately, then, these models can provide a link between economic theory and public management.



## II. Statistical Models

Since its creation in 1974, the NRC has overseen the commercial nuclear power industry. One of its primary functions is to protect the public's welfare by formulating and enforcing effective safety regulations.[1] Nuclear power plants embody a complex technology, and in response to this complexity the NRC has formulated an equally complex body of safety standards. These standards provide detailed prescriptions for most aspects of plant operations, particularly those related to the core nuclear reaction, coolant systems, accident response, emergency backup systems, and radioactive waste disposal. Examples of some of the dozens of regulations involved include those related to the calibration of equipment, maintenance and testing schedules, valve settings, control room practices, emergency preparedness, and plant logbook accuracy and detail. The NRC has also introduced multiple safety systems into reactor design. Many safety features are "built in," for example, backup emergency cooling systems, backup power generators, and containment structures. Others, mainly in the control room, monitor operations. This plethora of standards is the operational aspect of what Nichols and Wildavsky (1987) label "prescriptive regulation." Complying with all the standards is an expensive, time-consuming, and sometimes technically difficult task for plant management; noncompliance is often a tempting alternative.

Primary responsibility for ensuring compliance lies with the plant's quality control program typically run by on-site management. In turn the plant licensee monitors quality control with its own quality assurance program. Ultimate responsibility for maintaining compliance, however, rests with the NRC itself, which has developed a program of frequent inspections carried out by the Inspection and Enforcement Division, which as of 1985 consumed 31.7 percent of the commission's personnel and 21.2 percent of its budget (source: NRC annual report, 1985).

[1] Prior to 1974, the Atomic Energy Commission oversaw all nuclear activities.

*Poisson and Binary Choice Models of Noncompliance*

The inspection records comprise detailed publicly available evidence on plant violations, listing all violations cited and any remedial action the plant must take in response.[2] They provide the most natural starting point for an analysis of the factors associated with power plant noncompliance.

Given the complexity and sheer quantity of the regulations, the most useful approach to plant compliance decisions views managers as choosing an overall level of care to devote to safety, implemented via human and financial resources. By hiring more staff engineers, paying higher wages, or increasing maintenance expenditures, managers can reduce the risk of violations, and the process of isolating and correcting problems that do arise can be quickened.

Let $\mathbf{X}_1$ denote observable plant characteristics, including current financial position, plant technology, and past NRC sanctions. If we suppose the management's choice of level of care to depend on the factors $\mathbf{X}_1$, violations can be modeled as occurring (or "arriving") at the plant according to a Poisson process. Setting $\lambda = e^{\mathbf{X}_1\beta_1}$ ($\lambda$ must be positive; hence the exponential form), the probability that $n$ violations will occur, $n = 0, 1, 2, \ldots$, is $e^{-\lambda}\lambda^n/n!$. The expected number of violations is $\lambda$; the variance is also $\lambda$.

The Poisson model can be estimated conveniently in a nonlinear least squares framework. Defining $N_i$ to be the number of detected violations during the $i$th inspection, we may write

$$N_i = \lambda_i + \zeta_i, \tag{1}$$

where the variance of $\zeta_i$ is $\lambda_i$. Estimation follows a two-stage generalized least squares procedure in which the first stage produces an estimate of var($\zeta_i$), allowing a reweighting of the observations to correct for heteroscedasticity.

The statistical process that generates equation (1) is labeled "pure" Poisson, a model in which the stochastic variation in violations arises solely from the stochastic nature of the Poisson distribution; the Poisson parameter $\lambda$ is fully determined by the observables $\mathbf{X}_1$. The model can be generalized to a "random" Poisson version in which $\lambda$ itself is stochastic: $\lambda = e^{\mathbf{X}_1\beta_1+\epsilon}$, where $E(e^\epsilon) = 1$ and $\text{var}(e^\epsilon) = \eta^2$. The stochastic term $\epsilon$ can be interpreted as plant characteristics that affect the cost and implementation of compliance but are unknown to either the inspectors or the econometrician. Alternatively, $\epsilon$ can be viewed as a stochastic shock to plant safety technology that directly affects the frequency of violations. A nonlinear least squares form for estimating

[2] Plant logbooks are an important second source of information about operations.



random Poisson models is discussed in some detail in Gourieroux, Monfort, and Trognon (1984). As they show, model (1) can again be estimated using a two-stage procedure, except that the variance of $\zeta_i$ is now equal to $\lambda_i + \lambda_i^2\eta^2$, and an estimate of $\eta^2$ is computed from the first-round results.[3] When $\eta^2$ is estimated as less than or equal to zero, the random Poisson model collapses back to the pure Poisson, indicating that $\mathbf{X}_1$ explains variability in $\lambda$ sufficiently well to eliminate the need for a stochastic term.

The Poisson model captures in a natural way the possibility of multiple violations, which is important in assessing safety risks (particularly in the PRA framework). However, in some cases it may be overly sensitive to "outlier" inspections in which an unusually large number of violations are cited. One way to address this concern would be to generalize the Poisson to, for example, the compound Poisson in which violations derive from different possibly interrelated processes. Such an extension must be left to future work, however, and instead the sensitivity of the Poisson will be explored by specifying an alternative binary choice framework.

If we continue to let $\mathbf{X}_1$ denote observable plant characteristics, noncompliance emerges from the latent variables formulation

$$\begin{gathered} Y_1^* = \mathbf{X}_1\beta_1 + \epsilon_1, \\ Y_1 = \begin{cases} 1 \text{ (violation)} & \text{if } Y_1^* > 0 \\ 0 \text{ (compliance)} & \text{if } Y_1^* \leq 0. \end{cases} \end{gathered} \tag{2}$$

The interpretation, which differs from that of the Poisson, is that the plant makes a single global decision about whether or not to comply with standards; violations are intentional. The random variable $\epsilon_1$ represents unobservable plant characteristics that affect the costs and benefits of compliance. If $\epsilon_1$ is distributed according to the distribution function $F$, the probability of a violation is $F(\mathbf{X}_1\beta_1)$, and the model can be estimated using conventional maximum likelihood techniques.

*Detection Controlled Models*

The models above focus exclusively on plant noncompliance. While this approach is appealing, especially in the case of nuclear power, where so much public attention has been focused on plant mismanagement, it ignores the inspection process generating the data.

[3] Specifically, $\eta^2$ is computed as the parameter estimate from a regression in which the dependent variable is the residual from eq. (1) squared minus the estimated $\lambda$, and the independent variable is $\lambda^2$.

Specifically, it ignores the possibility that some violations typically escape detection and therefore are never recorded.

The principal reason nondetection occurs is that the NRC has limited resources available for inspections, especially considering the sheer complexity and size of the plants. Nichols and Wildavsky (1987, p. 50) quote a senior NRC inspector as having remarked that "the head of Inspection and Enforcement said some place that we [should] inspect 1% of all construction. No way could I have looked at 1% of everything done. People can write requirements forever. But it's a case of the alligator mouth and the hummingbird stomach. Even in an operating reactor you have 250 people; you can't do a comprehensive check of everything they do."

More subtly, nondetection is endemic to the regulatory process, which imposes requirements that in the absence of enforcement power plants would not adhere to. In this sense nondetection is the empirical analogue of the asymmetric information that characterizes the firm: regulator interaction discussed, for example, in Baron and Besanko (1984) and Laffont and Tirole (1986). We thus expect the nondetection problem to be exacerbated by a number of incentive effects. Plant management will desire to conceal violations, whether these violations are intentional or arise spontaneously because of insufficient oversight and come to its attention later. In other cases, management and staff may obstruct detection by keeping poorly documented or incorrect records of operations, actions that are themselves regulatory violations but may throw inspectors offtrack for many months.

As these arguments suggest, nondetection is likely to be intrinsic to the inspection process, not just of nuclear power plants but of many similar regulatory systems. Since nondetection is endemic to inspection data, it is useful to explore how it may bias inferences drawn from models that erroneously assume complete detection. On average, models that assume complete detection underestimate the true extent of noncompliance since all plants not detected violating are assumed compliant. More generally, estimates of the factors associated with noncompliance can be systematically biased. To take a simple example, consider two plants, A and B, inspected alternately by two inspectors, named Cindy and Joe. Each inspector visits both plants, but with differing frequency; thus suppose that each performs 30 inspections, with Cindy inspecting plant A 20 times and B 10 times and Joe inspecting A 10 times and B 20 times. Each inspector is characterized by a detection rate, which represents the probability that he or she will detect a violation if one has occurred; we assume that the detection rate is the same for every inspection and for both plants. Suppose that Cindy's detection rate is 90 percent and Joe's is



50 percent. Since Cindy has a higher detection rate and performs more of the A inspections, the ratio of A's detected violation rate to B's will be biased upward relative to the ratio of their true rates of noncompliance: if each plant violates half the time, A's detected rate will be approximately 38.3 percent (23/60) and B's 31.7 percent. Such a bias can have important policy implications; for example, if the NRC assigns better detectors to less compliant plants, an "overdispersion bias" can emerge in which the less compliant plants appear to violate relatively more frequently than they really do.

This example not only points out the biases inherent in ignoring nondetection but also suggests that the problem may be resolved by modeling the detection process jointly with the violation process. Continuing to assume that the inspectors' detection abilities are the same at both plants, we can examine their relative performance at each plant. We find that Cindy detects 80 percent more violations; taking into account the inspectors' different detection rates, we can then reanalyze plant violation rates. In particular, scaling up Joe's detected violations by 80 percent leads to "Detection Controlled" estimates of noncompliance, 45 percent at each plant, which removes the earlier bias, though it does not go the whole way toward alleviating the nondetection problem.

This example can be made more rigorous and improved on, in a large data set, by specifying a detection equation as part of a formal analysis of noncompliance. As the simplest example of Detection Controlled estimation, consider the binary choice model of noncompliance given by equation (2). We can generalize this model to a two-equation system, modeling the inspection process through which violations are discovered in the simplest possible way, as a linear detection technology. Thus, conditional on a violation occurring ($Y_1 = 1$), set

$$Y_2^* = \mathbf{X}_2\beta_2 + \epsilon_2,$$
$$Y_2 = \begin{cases} 1 \text{ (detection)} & \text{if } Y_2^* > 0 \\ 0 \text{ (nondetection)} & \text{if } Y_2^* \leq 0. \end{cases} \tag{3}$$

In this specification $\mathbf{X}_2$ includes variables likely to affect detection and $\epsilon_2$ represents a stochastic shock to the detection technology. Among variables included in $\mathbf{X}_2$ will be the identity of the personnel who perform the inspection, allowing the calculation of a separate mean detection rate for each individual or team. If $\epsilon_2$ is drawn from the distribution $G$, the probability of detection is $G(\mathbf{X}_2\beta_2)$.

Since only detected violations are observed, inspection data fall into two categories: the set $A$ consisting of inspections for which a violation has been discovered and the set $A^c$ consisting of all others. The likelihood that observation $i$ will fall into set $A$ is $F(\mathbf{X}_{1i}\beta_1)G(\mathbf{X}_{2i}\beta_2)$, reflecting the fact that two events have occurred in succession: violation and detection. The set $A^c$ consists of two types of observations: compliant plants, the likelihood of which is $1 - F(\mathbf{X}_{1i}\beta_1)$, and undetected noncompliers, the likelihood of which is $F(\mathbf{X}_{1i}\beta_1)[1 - G(\mathbf{X}_{2i}\beta_2)]$. These two terms sum to $1 - F(\mathbf{X}_{1i}\beta_1)G(\mathbf{X}_{2i}\beta_2)$. It is the inability to separate the compliers from the undetected noncompliers which is at the root of the nondetection data problem. The log likelihood of the sample is

$$L = \sum_{i \in A} \log[F(\mathbf{X}_{1i}\beta_1)G(\mathbf{X}_{2i}\beta_2)] + \sum_{i \in A^c} \log[1 - F(\mathbf{X}_{1i}\beta_1)G(\mathbf{X}_{2i}\beta_2)]. \quad (4)$$

Estimation of equation (3) produces Detection Controlled Estimates (DCE), which estimate variations in noncompliance and detection simultaneously; the models are discussed more fully in Feinstein (1987). The model's symmetry captures the inherently two-sided nature of the regulatory process, which has recently been stressed in the theoretical work of Baron and Besanko (1984) and Laffont and Tirole (1986).

A Detection Controlled random Poisson model can be constructed by similar arguments, appending equation (3) to the original Poisson process generating violations. Under the assumption that each violation has a probability $G(\mathbf{X}_2\beta_2)$ of being detected and that any one violation's probability of being detected is independent of any other's chances of being detected and of the Poisson process generating violations, a nonlinear least squares model may be derived:

$$N_i = G(\mathbf{X}_{2i}\beta_2)\lambda_i + \zeta_i. \quad (5)$$

Under the independence assumptions, the process "detected violations" is itself Poisson, so that the variance of $\zeta_i$ is just $G_i\lambda_i + G_i^2\lambda_i^2\eta^2$.[4]

Detection Controlled models such as equations (4) and (5) improve on the ordinary Poisson and binary choice models of noncompliance, given by (1) and (2), in several regards. Non–Detection Controlled models replace the term $G$ by one, thereby implicitly assuming complete detection; in general this model is misspecified and produces inconsistent estimates of determinants of noncompliance, as discussed above. Since equation (3) models the detection process explicitly, it exploits available data on inspectors. This in turn allows a test for uniformity in detection across NRC personnel and helps identify particularly poor inspectors in an environment which has controlled for the different assignments inspectors receive.

[4] Alternatively, we might suppose that with probability $G$ all violations are detected, and with probability $1 - G$ none are. This leads to the same objective function but a variance of only $G\lambda$. A more flexible model would allow separate detection equations for each violation and a correlation among them.



Estimates derived from (4) and (5) allow computation of the rate of undetected violations. For the Poisson specification, this calculation is especially simple: undetected violations may be shown to follow a Poisson distribution that has expectation $(1 - G_i^*)\lambda_i^*$ (an asterisk denotes projections based on parameter estimates). Averaged over all inspections, this produces an estimate of the aggregate rate of undetected violations.

In the binary choice model, one can compute the posterior probability that a plant not detected in violation is in fact an undetected noncomplier. When Bayes's law is applied, this probability is

$$\frac{F(\mathbf{X}_{1i}\beta_1^*)[1 - G(\mathbf{X}_{2i}\beta_2^*)]}{1 - F(\mathbf{X}_{1i}\beta_1^*)G(\mathbf{X}_{2i}\beta_2^*)}. \quad (6)$$

Averaging over all plants in the set $A^c$ produces an estimate of the aggregate rate of noncompliance in the sample.

Though the Detection Controlled models are a clear improvement over models that do not include a detection equation, they do have a number of drawbacks. Most important, a statistical issue of identification arises in their estimation. The identification problem is well illustrated by the following example. Suppose that plant $i$'s probability of violating is $p_0e^{\mathbf{X}_{1i}\beta_1}$, where $p_0$ is the average level of noncompliance in the population and $e^{\mathbf{X}_{1i}\beta_1}$ fluctuates around one depending on whether $i$ is more or less likely to violate than average. Similarly define the probability of detection to be $q_0e^{\mathbf{X}_{2i}\beta_2}$. Data are available only on whether or not a violation was detected, which occurs with probability $e^{\mathbf{X}_{1i}\beta_1}p_0q_0e^{\mathbf{X}_{2i}\beta_2}$; formally this product corresponds to the $FG$ on which (4) depends. In this example, $p_0$ and $q_0$ cannot be separately identified, only their product. Conceptually the average absolute levels of violation and detection cannot be determined, a problem that also arises in the earlier discussion about inspectors Cindy and Joe: though it is clear how poor Joe is relative to Cindy, Cindy's absolute detection rate cannot be deduced. Just as in that example, the relative rates of noncompliance and detection that depend on $\mathbf{X}_{1i}$ and $\mathbf{X}_{2i}$ can be identified as long as $\mathbf{X}_1$ and $\mathbf{X}_2$ are not collinear. Variables common to $\mathbf{X}_1$ and $\mathbf{X}_2$ can be estimated only as reduced forms; however, in that case the variables will typically exert opposite effects on noncompliance and detection (e.g., the accident at Three Mile Island is likely to have increased detection but decreased noncompliance), in which case the variable's sign indicates its primary effect.

This double exponential form is extreme; it can be shown (see Feinstein 1987) that it is the only functional form in which identification of absolute levels formally fails. Thus, for example, when $F$

and $G$ are independent normal distributions, the parameters $\beta_1$ and $\beta_2$ are fully identified, including constants. Nonetheless, it serves as a warning that relative noncompliance and detection rates will typically be better estimated than absolute levels. In turn this suggests that estimates of the undetected rate of violation (eq. [4]), which depend critically on the functional forms of $F$ and $G$, must be treated with caution.[5]

The arguments leading to the binary choice and Poisson Detection Controlled models also introduce a number of simplifying assumptions into the analysis. First, no allowance is made for false detection; that is, it is assumed that a compliant plant is never falsely accused of a violation. This assumption seems appropriate in the present context since NRC inspectors always cite a specific identified problem; but it may not be appropriate in other contexts. Second, the errors $\epsilon_{1i}$ and $\epsilon_{2i}$ have been assumed to be independent. The model can readily be extended to the case in which $\epsilon_{1i}$ and $\epsilon_{2i}$ are jointly normally distributed, with the likelihood for this model involving a one-dimensional integration over one of the $\epsilon$'s; estimates based on it are presented as an extension in Section IV. Finally, these models do not incorporate a structure of interaction between the plant and the agency inspectors. One might believe that inspectors form an expectation of the probability of noncompliance at the plant and that this affects the detection process. A linear rational expectations version of this in which the inspectors possess the same information set as the econometrician implies that the term $F(\mathbf{X}_{1i}\beta_1)$ should be appended to $\mathbf{X}_{2i}$; this extension is not pursued in the present work.

### *Persistence*

Inspections data do not refer to a cross section of power plants, each inspected only once. Instead, they refer to a panel, a collection of plants each of which is inspected regularly, on average about once every 2 weeks. This sequential feature of the inspections introduces issues of timing that the earlier models have ignored, in two regards. Consider first a random Poisson model in which detection is assumed complete. Let period $t$ refer to the time between inspections $t - 1$ and $t$. Since detection is complete, all period $t - 1$ violations are detected during inspection $t - 1$, and the plant enters period $t$ with no violations. A new Poisson process begins, characterized by parameter $\lambda_t = e^{\mathbf{X}_{1t}\beta_1 + \epsilon_t}$. This process remains active generating violations until inspection $t$. Supposing that period $t$ lasts $\Delta_t$ days, it is natural to assume that the Poisson parameter is proportional to $\Delta_t$, so that $\lambda_t = (e^{\mathbf{X}_{1t}\beta_1 + \epsilon_t})\Delta_t$. When the gap between inspections is longer, more violations can be expected to "pile up" at the plant.

[5] In fact the estimates are typically on the conservative side since violation types detected by no inspectors leave no trace in the data.



The second, more subtle issue is the persistence of undetected violations. A violation which occurs during period $t - 1$ but remains undetected during inspection $t - 1$ may persist until inspection $t$ and be detected at that time. To the extent that undetected violations persist, estimation is confounded since violations detected during inspection $t$ derive from a mixture of Poisson processes with different parameters. In addition, detection rates $G_t$ and $G_{t-1}$ differ; as an example of how this might affect the data, when inspection $t - 1$'s detection rate is unusually low, an unusually large number of violations may go undetected, persist, and show up on inspection $t$'s list of cited violations, injecting an upward bias into $\lambda_t$ and $G_t$. Understanding the extent of persistence sheds light on power plant safety technology. It is also relevant to the analysis of unsafe events and how they emerge from previously undetected violations, discussed in Section V. For example, in the Three Mile Island accident, a precipitating factor was a valve that had been left open (when it should have been closed) during maintenance several days earlier and remained undiscovered in the intervening period.

To incorporate persistence into the analysis, suppose that a period $t - 1$ violation that remains undetected during inspection $t - 1$ has probability $e^{-\rho\Delta_t}$ of persisting until inspection $t$, at which time its probability of detection is $G_t$. The parameter $\rho$, $\rho \geq 0$, measures persistence. Undetected violations are not likely to persist indefinitely since the power plant's own staff will eventually identify and correct them. To simplify, we will assume that violations persist at most until the next inspection and then, if still undetected, disappear. As a further simplification, we will derive a model of persistence assuming that violations are generated according to a pure Poisson process and that processes in adjacent time periods are independent of one another.

Suppose that $N_{t-1}$ violations were detected during inspection $t - 1$. We must then determine the number of "old" violations expected to be detected during inspection $t$, conditional on $N_{t-1}$. Fortunately, under the pure Poisson specification, the probability distribution of undetected period $t - 1$ violations is independent of $N_{t-1}$, as will now be shown; this greatly simplifies the analysis and would not be true of other processes.

Conditional on $N$ (standing for $N_{t-1}$) detected violations, the probability of $j$ undetected violations is, by Bayes's rule, equal to the probability of $N$ detected *and* $j$ undetected violations, divided by the probability of $N$ detected violations. The probability of $N$ detected and $j$ undetected violations is

$$\frac{e^{-\lambda_{t-1}}\lambda_{t-1}^{N+j}}{(N+j)!}\binom{N+j}{N}G_{t-1}^{N}(1-G_{t-1})^{j}.$$

When we sum over $j$ and simplify, the probability of $N$ detected violations is

$$\frac{e^{-\lambda_{t-1}G_{t-1}}(\lambda_{t-1}G_{t-1})^{N}}{N!}.$$

Dividing out, we get

$$\text{prob}(j \text{ undetected violations} \mid N \text{ detected}) = \frac{e^{-\lambda_{t-1}(1-G_{t-1})}[\lambda_{t-1}(1-G_{t-1})]^{j}}{j!}, \quad (7)$$

which is itself a Poisson distribution with parameter $\lambda_{t-1}(1-G_{t-1})$ and is independent of $N$.

It now follows directly that for any $N_{t-1}$ the expected number of undetected violations following inspection $t-1$ is $\lambda_{t-1}(1-G_{t-1})$, the expectation of the Poisson conditional distribution. Of these, a fraction $e^{-\rho\Delta_t}$ persist on average until inspection $t$, at which time a fraction $G_t$ are detected. Hence the expected number of old violations detected during inspection $t$ is

$$G_t e^{-\rho\Delta_t}\lambda_{t-1}(1-G_{t-1}),$$

and the appropriate nonlinear least squares model to estimate is

$$N_t = G_t\lambda_t + G_t e^{-\rho\Delta_t}\lambda_{t-1}(1-G_{t-1}) + \zeta_t. \quad (8)$$

Under the assumption that the Poisson processes generating violations in adjacent time periods are independent and that the inspection processes at $t-1$ and $t$ are also independent of one another and of the Poisson processes, the variance of $\zeta_t$ is just the sum of $G_t\lambda_t$, the variance of the new detected violations, and the variance of the old detected violations.

The variability in the number of old violations detected arises from three sources. First, the number of violations that remain undetected after inspection $t-1$ may vary around its mean. Next, the fraction of such undetected violations that persist will fluctuate around the average rate $e^{-\rho\Delta_t}$. Compounding both of these,there will be variability in the detection process at time $t$, just as for "new" violations. The exact form of the variance turns out to be

$$\lambda_{t-1}(1-G_{t-1})(G_t e^{-\rho\Delta_t}), \quad (9)$$



TABLE 1

POWER PLANTS INCLUDED IN THE STUDY

| Power Plant | State | NRC Region |
|---|---|---|
| Haddam Neck | Connecticut | East |
| Millstone #2 | Connecticut | East |
| Maine Yankee | Maine | East |
| Pilgrim #1 | Massachusetts | East |
| Salem #1 | New Jersey | East |
| Indian Point #2 | New York | East |
| Nine Mile Point #1 | New York | East |
| R. E. Ginna #1 | New York | East |
| Fitzpatrick | New York | East |
| Peach Bottom #3 | Pennsylvania | East |
| Beaver Valley #1 | Pennsylvania | East |
| Three Mile Island #1 | Pennsylvania | East |
| Susquehanna #1* | Pennsylvania | East |
| San Onofre #1 | California | West |
| Rancho Seco #1 | California | West |
| Trojan #1 | Oregon | West |
| WPPSS #2* | Washington | West |

* Began commercial operation after 1982; thus data include inspections only for 1985.

to which is added the earlier $G_t\lambda_t$ to compute var($\zeta_t$). Under the very strong independence assumptions that have been made, $\zeta_t$ and $\zeta_{t-1}$ are also independent of one another. In general, the introduction of persistence would require more involved calculations if the Poisson processes in adjacent periods were assumed correlated, if they were generalized to random Poisson models, and if detection rates across inspections, or among violations within a given inspection, were correlated.

## III. Data

As of early 1986, there were 101 commercial nuclear generating units operating in the United States, producing approximately 17 percent of the nation's commercial electric power. The generating units are clustered on site installations, each containing one, two, three, or four units. This study analyzes data on 17 generating units or, as I will continue to call them, power plants, drawn from two of the NRC's five geographic regions, East and West.[6]

The 17 power plants are listed in table 1, together with the state

[6] I focus intensively on two regions rather than randomly sampling from all five because the Detection Controlled models isolate inspector effects, which in turn require a reasonable number of cases to be estimated. While inspectors do cross regions in the data, they do not do so sufficiently to allow uniform sampling.

they reside in and their NRC region. The preponderance of plants from the East relative to the West reflects the geographic concentration of nuclear plants; in fact the plants were selected by choosing every other plant from the East region and every plant from the West, with the proviso that no more than one plant was chosen from a given site installation.

The principal data set used in the study is a compilation of all NRC inspections of operating conditions at these 17 plants over the years 1979, 1982, and 1985, except for two of the plants, Susquehanna #1 and WPPSS #2, which did not begin commercial operation until after 1982; for these two plants, data are available only for 1985. The inclusion of Three Mile Island #1 also deserves comment: this plant was shut down from April 1979 to the middle of 1985, but during that period it was manned and inspected regularly. The data were made available to me by the NRC Public Documents Room in Washington, D.C.

The total number of inspections is 1,178, of which 1,013 have proven appropriate for a study of safety regulation; the criteria that eliminated the other 165 are discussed below. The 1,013 inspections comprise a panel, consisting of the complete sequence of inspections at each plant for each year; typically the inspections are spaced quite uniformly apart over the year, averaging about two inspections per month.[7] Figure 1 illustrates the distribution of numbers of inspections at the different plants; on the basis of these data, the frequency of inspections varies from a high of just below three per month at Rancho Seco to a low of slightly more than one per month at Haddam Neck. Undoubtedly Rancho Seco (and some of the other plants) is inspected frequently by design since it has acquired a reputation as possessing an especially error-prone safety technology; it would be interesting, but beyond this paper's scope, to investigate the process through which the NRC allocates inspection resources across plants. For each inspection the data include the dates the inspection began and ended, a brief description of the nature of the inspection, and a listing and description of any violations cited.

Important for the Detection Controlled methods, the data also identify the names of all inspectors who participated in the inspection. A total of 321 inspectors are identified in the data, with the typical inspection including two to four individuals. Of these 321, I have selected 40 on whom to focus attention, specifying a fixed effect for each of the 40. By and large these 40 are the inspectors who performed the most inspections; figure 2 depicts the distribution of num-

[7] However, some of the inspections are "resident" inspector monthly reports; see below.

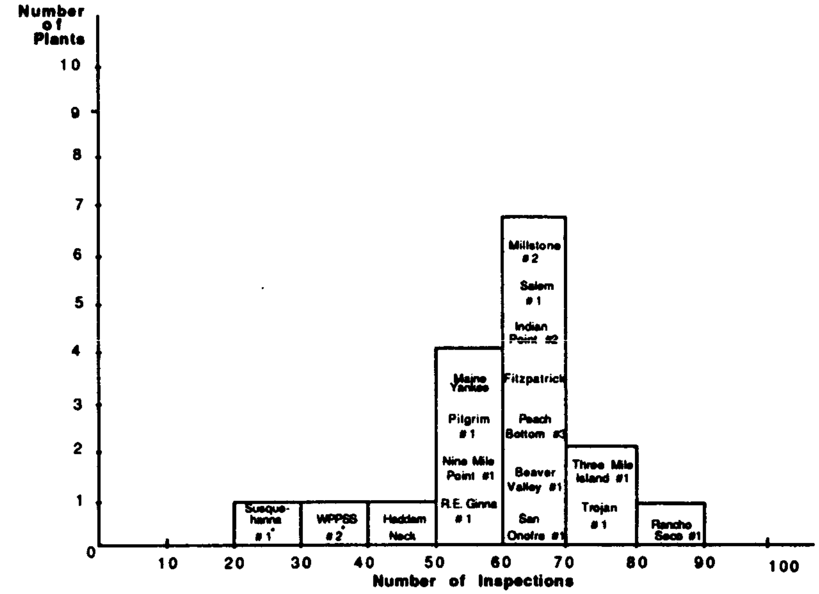


FIG. 1.—Histogram depicting distribution of inspections among plants. Susquehanna #1 and WPPSS #2 began commercial operation after 1982; thus data include inspections only for 1985.

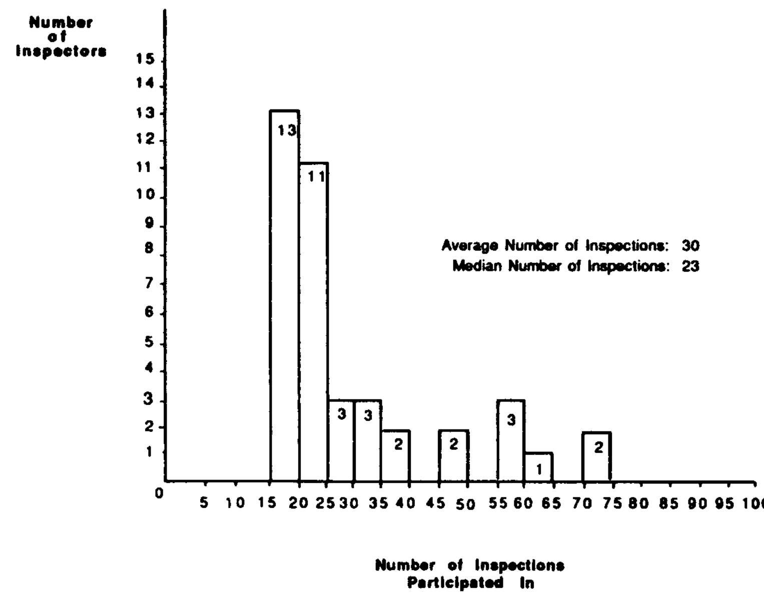


FIG. 2.—Histogram depicting distribution of inspections among 40 chosen inspectors

TABLE 2

TYPES OF INSPECTIONS

| Inspection Type | Number of Inspections |
|---|---|
| 1. Resident inspection | 343 |
| 2. Specific hardware issues | 264 |
| 3. Management practices | 341 |
| 4. Personnel training/emergency preparedness | 64 |
| Total | 1,013 |
| 5. Plant security and restricted access checks | 61 |
| 6. Exposure to radiation/radioactive mail | 105 |
| Grand total | 1,179 |

bers of inspections performed by each of the 40, with the average number of inspections being 30 and the median number 23. This group of 40 is responsible for 45 percent of the inspection "slots" recorded in the data.

In selection of the 40 inspectors, two technical issues arose. Since inspectors work in teams, it was important to guarantee that no two of the chosen 40 inspected together too much of the time. In most cases the teams are quite fluid, with membership rotating through a large pool of inspectors. Nonetheless this criterion did eliminate a few inspectors, and in the final group of 40, six remain who performed just over 50 percent of their inspections with one other partner. In all six cases, however, that partner performed at least 49 inspections himself and no more than 25 percent of these with this particular inspector. The assignment of inspectors to plants motivated a second selection criterion, that no inspector perform too many inspections at any one plant; in the final group, eight inspectors performed just over 50 percent of their inspections at one plant.

The data describe six basic types of inspections, listed in table 2 together with the frequency of each type. Category 1, resident inspection, deserves special comment. Each power plant is assigned a resident inspector who has primary responsibility for plant safety during his tenure. On a monthly or bimonthly basis, the resident inspector files an inspection report. Conceptually these resident inspections are different from the other inspections, all of which refer to brief visits (1–3 days is common) by NRC personnel; hence we would expect detection rates and possibly types of violations to differ as well. Categories 5, plant security and restricted access (to unauthorized personnel) checks, and 6, radiation exposure and radioactive mail procedures, do not pertain directly to accident safety; inspections in these categories are removed, leaving a data base of 1,013.



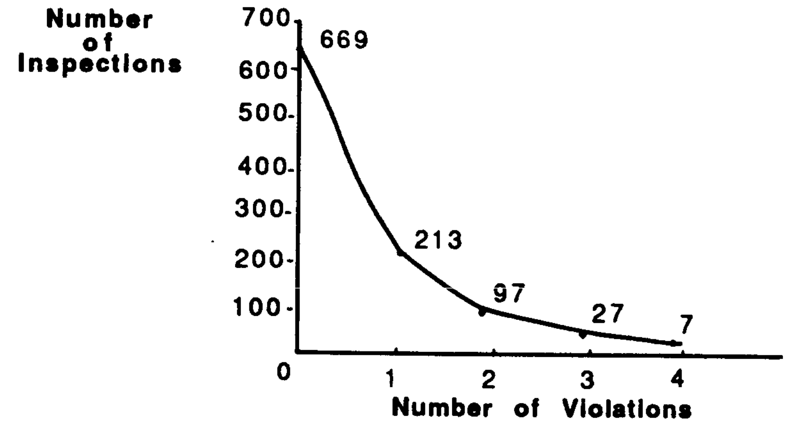


FIG. 3.—Distribution of numbers of violations cited per inspection. Only violations cited in categories 1–4 of table 2 are included.

Figure 3 graphs the distribution of number of violations cited among these 1,013 inspections; although these inspections focus primarily on safety issues, occasionally violations related to categories 5 and 6 of table 2 arise; such violations are not included in figure 3 or any of the remaining analysis. Most inspections (roughly two-thirds) do not lead to any citations, and the distribution resembles the tail of an exponential or Poisson distribution. While a large number of specific types of violations are cited, they fall broadly into the four safety-related categories of table 2. Examples of prevalent violations are "surveillance equipment miscalibrated," "valves left open when they should be closed," and "maintenance not performed as scheduled," all of which refer to hardware or equipment violations; and "failure to properly review proposed safety design modifications," "no fire watch posted," and "inadequacies in emergency preparedness training," which refer to management practices. No attempt has been made to rank these violations by degree of severity, in large part because the experiences at Three Mile Island and Chernobyl and the PRA approach all emphasize that interaction among apparently minor violations can produce a serious safety hazard. Nonetheless, all results obtained in this paper must be interpreted with the knowledge that they derive from large-scale statistical analysis of all violations rather than from case studies of specific plant malfunctions.

What do the raw data reveal about noncompliance and detection? As a simple means of answering this question, the inspections have been grouped into the two categories "no violations cited" and "one or more violations cited"; this distinction parallels the simple binary choice models discussed earlier and estimated in the next section. Figure 4 depicts two histograms based on this characterization. The

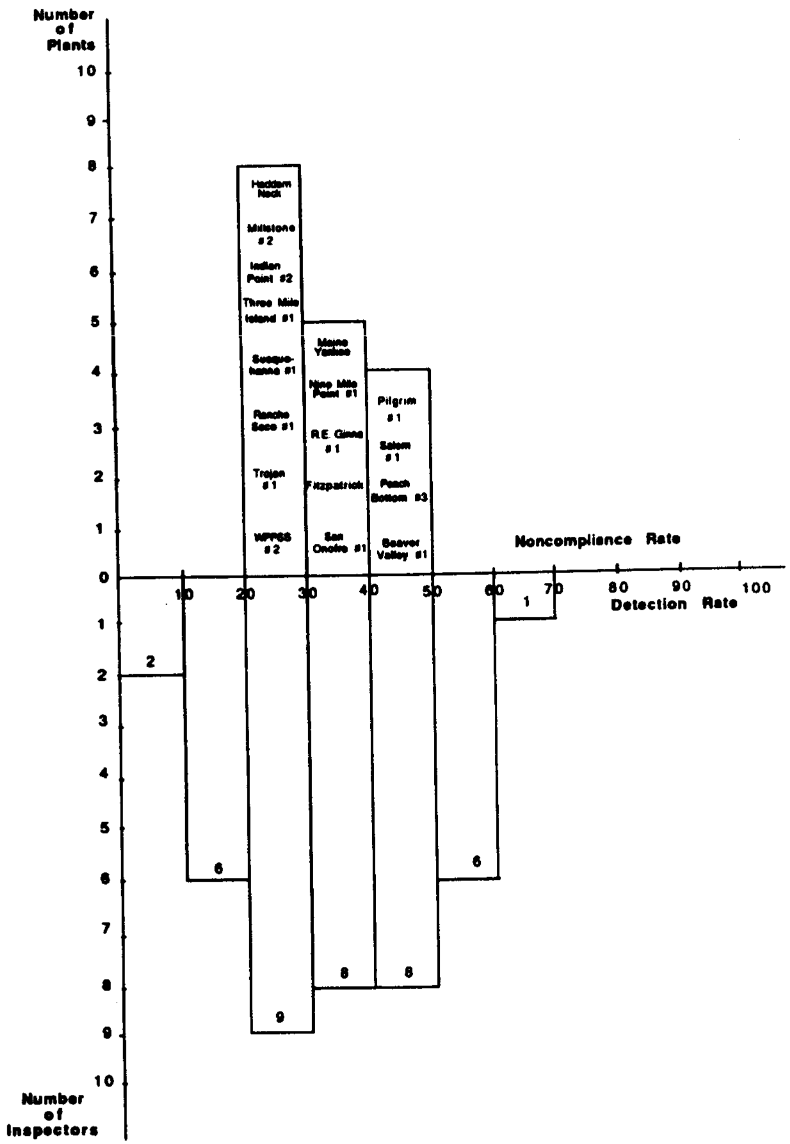


FIG. 4.—Dual histograms depicting the variation in raw compliance rates among plants and raw detection rates among inspectors.



top (upward-pointing) histogram illustrates the variation in detected noncompliance across plants, listing, for example, the plants for which 20–30 percent of all inspections led to at least one citation. The bottom (downward-pointing) histogram summarizes the variation in detection rates among the 40 inspectors. It is interesting to notice that the variation in detection rates across NRC personnel is comparable to and possibly exceeds the variation in noncompliance rates across plants, which suggests that the Detection Controlled methods may substantially improve on models that focus exclusively on noncompliance.

The substantial variability in both noncompliance and detection motivates the estimation of more complete statistical models that include explanatory variables and control simultaneously for variations in these two dimensions. Appropriate models have been discussed at length earlier and are presented in the next section. The remainder of this section describes the construction of explanatory variables; since the data are organized by inspection, all variables take a specified value for each inspection, usually depending on the inspection's date and the power plant being inspected.

According to the deterrence hypothesis, higher expected sanctions, as measured by increases in the level of sanctions imposed in the recent past, should reduce noncompliance. To investigate this hypothesis the NRC monthly series *Enforcement Actions: Significant Actions Resolved,* which lists all sanctions imposed by the commission, has been used to create two variables: SANCNATIONAL, a 3-month moving average of the aggregate fines (normalized by 100,000) levied by the NRC against commercial power plants found violating safety standards; and SANCPLANT, a plant-specific dummy variable set to one if the plant being inspected has been sanctioned in the preceding 3 months.

A second hypothesis to be considered is whether financially troubled plant owners increase noncompliance, either as a cost-saving device or because their financial liability in case of accident is reduced.[8] The variable BONDRATING has been constructed on the basis of Moody's bond ratings;[9] a plant the majority of whose bonds rated Aaa in the year of the inspections has been assigned a BONDRATING value of 16, bonds rated Aa1 have been assigned 15, and so on down to the lowest rating in the sample, B2, assigned to the Gen-

[8] As Dubin and Rothwell (1987) point out, even though power plants are well covered by insurance, they still face liability from an accident in the form of equity losses and a number of lesser losses.

[9] The source for this is Moody's *Public Utilities Manual* and Moody's *Municipal and Government Manual.* The plant WPPSS #2 was not rated in 1985.

eral Public Utilities Corporation (which operates Three Mile Island) in 1982, which translates into a BONDRATING value of 2.

Plant technology is also likely to affect noncompliance. The variable AGE measures the age of the power plant in years as of 1985 (see the 1985 NRC annual report). Age can affect compliance in two ways. Older power plants may be expected to comply more readily to the extent that a "learning curve" (discussed in Joskow and Rozanski [1979]) effect increases managerial efficiency over time. On the other hand, new plants possess better-engineered safety systems, and old plants have often been retroactively fit with recent systems, which can make compliance more costly. The dummy variable BWR is set to one if the power plant being inspected is a boiling water reactor (BWR). There are two main types of reactors in commercial use in the United States, BWRs and pressurized water reactors (PWR). They differ in their coolant system technology: the BWR coolant system's water converts directly into steam to drive its turbine generators (hence the term boiling), while the PWR coolant system is kept under sufficiently high pressure to prevent it from boiling and instead exchanges its heat with a secondary coolant system, which in turn drives the turbines.[10] Hence BWRs possess a somewhat less stable cooling system than PWRs, which creates the need for more complex monitoring technology and may make compliance more costly. The NRC monthly publication *Licensed Operating Reactors Status Summary Report* has been used to construct the variable DOWN, a dummy set to one if the plant either is shut down or is running at too low a level to generate salable electric power during the month of the inspection. In the sample, plants are down approximately 30 percent of the time. One might expect down plants to comply more readily with regulations, for example because they have more time available for maintenance activities. Collectively these three variables represent only a rudimentary characterization of plant technology and operations; other data sources such as plant logs (which record specific hardware maintenance activities and problem areas) would provide valuable additional information.

As mentioned above, power plants are assigned resident inspectors. The dummy variable CHANGERESINS is set to one if there has been a change in the resident inspector during the month of the inspection. Such a change may increase noncompliance if the new inspector is less experienced with the plant's safety practices; alternatively, noncompliance may decrease if the old inspector had been "captured" by the power plant staff.

A major factor expected to influence detection rates is the identities

[10] For a useful and simple discussion, see Megaw (1987).



of the inspectors; fixed effects have been constructed for each of the 40 chosen inspectors. These are the inspectors for whom the quantity of inspections was judged sufficient to estimate such effects with reasonable precision.[11] As a second detection variable, the dummy RESINSPECTION is set to one if the inspection is a resident inspector's report; the detection rates of these reports may differ systematically from other inspections.

The last two variables explore whether the NRC is a reactive agency that responds to adverse incidents, or adverse publicity, by tightening detection. The accident at Three Mile Island received widespread press coverage and elicited criticism of the commission from several quarters, notably the President's Commission on the Accident at Three Mile Island (1979). Hence we might expect the NRC to have increased citations in the months following the accident; to investigate this possibility, the dummy variable POST-TMI is set to one for all inspections performed during April–December 1979. The variable LAG.EVENTS is constructed as the number of safety events reported by the power plant in the 30 days preceding the inspection's initiation date.[12] If the NRC is a reactive agency, an increase in LAG.EVENTS should increase detection rates.

It can be argued that the last four variables discussed, CHANGERESINS, RESINSPECTION, POST-TMI, and LAG.EVENTS, might each affect both noncompliance and detection rates. They have been included in one or the other equation on the basis of a priori reasoning, but in general the expected signs with which they affect the propensity to violate and detection will be opposite, which allows identification of their primary effect.

## IV. Empirical Results

Empirical results are presented in the following order. First the binary choice analysis of noncompliance is discussed; table 3 provides the actual estimates, including both ordinary and Detection Controlled models; table 4 presents rankings of power plants by propensity to violate; and figure 5 depicts the histogram of the distribution of implied inspector detection rates. Next estimates of Poisson models are discussed, with parameter estimates provided in table 5 and inspector detection rates in figure 6; power plant noncompliance rankings for the Poisson models are included in table 4. Finally, extensions

[11] Since the models are nonlinear in detection, consistency requires an asymptotic argument, requiring many inspections per inspector.

[12] The source for this is the NRC monthly publication *Licensee Event Reports* discussed in Sec. V.

that incorporate persistence and an additional binary choice model that allows nonzero correlation between the violation and detection processes are presented in table 6.

Table 3 presents four binary choice models of power plant noncompliance. Model 1 is a basic probit model, while model 2 is a probit model that includes plant effects. Models 3 and 4 are Detection Controlled analyses in which probability distributions that govern the rates of noncompliance and detection (*F* and *G*) are taken to be normal and independent of one another; model 3 is a base case (no plant or inspector effects), while model 4 is the most general of the four models and includes both plant effects and inspector effects for the 40 chosen inspectors.

Results are broadly consistent across these four models and are also broadly consistent with the Poisson models presented shortly. Collectively, the results suggest a number of conclusions about noncompliance and detection. First, the explanatory variables included in the violation equation are by and large only weakly associated with noncompliance. This is especially true of the deterrence variables SANCNATIONAL and SANCPLANT, which do not attain significance in any of the four models, and the financial variable BONDRATING, which is of the wrong sign and is marginally significant when plant effects are omitted, but otherwise is insignificant. There is little evidence to indicate that economic incentives influence plant behavior.

The variables related to power plant technology perform slightly better. The dummy variable BWR is positive and significant when plant effects are omitted but insignificant when they are included. The dummy variable DOWN is negative in all four models, significant in models 1 and 3, and marginally significant in model 2; apparently power plants not generating salable electricity do have an easier time complying with regulations. The variable CHANGERESINS is marginally significant only when plant effects are omitted. The variable AGE is insignificant in models 1 and 3; it is omitted from the models that include plant effects because of multicollinearity problems that arise between it, BWR, and the plant effects.

A second result that emerges from table 3 is that idiosyncratic plant effects are important determinants of noncompliance. In both models 2 and 4 the plant effects are jointly significant at the 95 percent level.[13] Table 4 presents the rankings of plants according to estimated propensity to noncomply. The first column reproduces in a more detailed form the information contained in the histogram of figure 4, presenting the rankings of the plants in the raw data. Columns 2 and

[13] Only 15 plant effects are specified since an overall constant and the dummy variable BWR are also included.



3 present the rankings estimated from the binary choice models 2 and 4; these rankings are computed directly from the estimates in table 3, with the proviso that for BWR plants the BWR coefficient was added to the plant effect, and for Trojan the net effect (abstracting from the overall constant) is zero, while for WPPSS #2 the net effect is the BWR coefficient. The rankings are quite similar among these three columns and reflect public perceptions about the relative propensity to noncomply among the different plants. Thus, for example, the Pilgrim and Peach Bottom plants, both of which have a long history of mismanagement, rate at the top of the list, while the Indian Point plant, which is widely regarded as a model for good management, ranks near the bottom. One surprise in the rankings is the Rancho Seco plant, which ranks near the bottom despite its reputation for safety problems. One interpretation of this finding is that while Rancho Seco's technology is erratic, its management is not, a conclusion that does not contradict the earlier finding (see fig. 1) that it is the most frequently inspected plant in the sample; it may simply require an unusually large amount of monitoring.

A third conclusion, based on models 3 and 4, is that detection matters. The hypothesis of complete detection is strongly rejected by a likelihood ratio test that compares the fits of models 3 and 1, and 4 and 2. Furthermore, there is strong evidence of heterogeneity among the 40 chosen inspectors. Figure 5 presents a histogram illustrating the distribution of estimated detection rates among the 40 based on model 4; since this figure refers to absolute levels instead of relative rankings, it must be interpreted with caution, but it does indicate the presence of a tail of underperformers. The figure also presents the raw data histogram of detection rates as a point of comparison; the general rightward shift in the estimated rate histogram relative to the raw data histogram is to be expected since the estimated rates derive from a model that allows for the possibility of compliance. A likelihood ratio test comparing the fit of models 4 and 3 strongly rejects the hypothesis that detection rates are similar across the 40 and between them and the other inspectors. This result is strengthened by the fact that model 4 also includes plant effects; the variability among inspectors apparently does not derive primarily from the different plant assignments they receive. While this heterogeneity result requires confirmation, especially to investigate whether it derives from omitted noncompliance variables or variations in the types of inspections performed by different inspectors, it does raise important policy questions about the training of inspectors.

Finally, models 3 and 4 provide evidence that after the Three Mile Island accident, detection rates, or at least citation rates, increased markedly: the variable POST-TMI is positive and significant. The

TABLE 3

BINARY CHOICE MODELS OF NONCOMPLIANCE AND DETECTION

| | Model 1: Basic Probit | Model 2: Probit Plant Effects | Model 3: Basic DCE | Model 4: DCE Plant Inspection Effects |
|---|---|---|---|---|
| Violation equation: | | | | |
| CONSTANT | −.656 (.168) | −.738 (.398) | −.577 (.212) | −.683 (.481) |
| SANCNATIONAL | −.0393 (.168) | −.128 (.172) | .206 (.269) | −.0409 (.226) |
| SANCPLANT | .0646 (.168) | .0808 (.180) | .0820 (.225) | .156 (.242) |
| BONDRATING | .0247 (.0137)* | −.00167 (.0383) | .0294 (.0176)* | .0109 (.0503) |
| BWR | .231 (.0915)** | −.0291 (.433) | .345 (.122)** | −.0190 (.516) |
| CHANGERESINS | −.548 (.318)* | −.512 (.325) | −.660 (.358)* | −.422 (.447) |
| DOWN | −.242 (.115)** | −.224 (.123)* | −.295 (.145)** | −.119 (.164) |
| AGE | −.0113 (.0109) | ... | .00268 (.0149) | ... |
| Plant effects: | | | | |
| Haddam Neck | | .258 (.256)[a] | | .282 (.609)[a] |
| Millstone #2 | | .130 (.232) | | .144 (.332) |
| Maine Yankee | | .361 (.239) | | .474 (.369) |
| Pilgrim #1 | | .851 (.445) | | 1.26 (.564) |
| Salem #1 | | .723 (.276) | | .883 (.439) |
| Indian Point #2 | | .107 (.265) | | .0563 (.404) |
| Nine Mile Point #1 | | .478 (.474) | | .558 (.594) |
| R. E. Ginna #1 | | .310 (.251) | | .329 (.350) |
| Fitzpatrick | | .406 (.623) | | .631 (.798) |
| Peach Bottom #3 | | .684 (.391) | | 1.18 (.513) |
| Beaver Valley #1 | | .560 (.235) | | 1.16 (.425) |
| Three Mile Island #1 | | .0625 (.267) | | .186 (.350) |
| Susquehanna #1 | | .249 (.538) | | .501 (.731) |
| San Onofre #1 | | .549 (.292) | | .473 (.473) |
| Rancho Seco #1 | | .0546 (.294) | | .0271 (.446) |
| Detection equation: | | | | |
| CONSTANT | | | .0647 (.205) | .412 (.396) |
| RESINSPECTION | | | .830 (.359)** | 1.36 (.596)** |
| POST-TMI | | | 3.25**[b] | 1.34 (.560)** |
| LAG.EVENTS | | | .0218 (.0393) | −.121 (.0700)* |
| Log likelihood | −621.7 | −606.7 | −608.7 | −538.0 |

* Statistically significant at the 90 percent level.
** Statistically significant at the 95 percent level.
[a] Plant effects jointly statistically significant at the 95 percent level.
[b] Likelihood flat. Statistically significant at the 95 percent level.

## TABLE 4

RANKINGS OF POWER PLANTS IN STUDY ACCORDING TO ESTIMATED PROPENSITY TO NONCOMPLY

| Rank | Raw Data (1) | Binary 2 (2) | Binary 4 (3) | Poisson 2 (4) | Poisson 4 (5) |
|---|---|---|---|---|---|
| 1 (highest) | Pilgrim #1* | Pilgrim #1* | Pilgrim #1* | Pilgrim #1* | Pilgrim #1* |
| 2 | Salem #1 | Salem #1* | Peach Bottom #3* | Salem #1 | Beaver Valley #1 |
| 3 | Peach Bottom #3* | Peach Bottom #3* | Beaver Valley #1 | Peach Bottom #3* | Salem #1 |
| 4 | Beaver Valley #1 | Beaver Valley #1 | Salem #1 | Beaver Valley #1 | R. E. Ginna #1 |
| 5 | San Onofre #1 | San Onofre #1 | Fitzpatrick* | San Onofre #1 | Fitzpatrick* |
| 6 | Nine Mile Point #1 | Nine Mile Point #1* | Nine Mile Point #1* | Fitzpatrick* | Peach Bottom #3* |
| 7 | Maine Yankee | Fitzpatrick* | Susquehanna #1* | Nine Mile Point #1* | Millstone #2 |
| 8 | Fitzpatrick* | Maine Yankee | Maine Yankee | Maine Yankee | San Onofre #1 |
| 9 | R. E. Ginna #1 | R. E. Ginna #1 | San Onofre #1 | WPPSS #2* | WPPSS #2* |
| 10 | Haddam Neck | Haddam Neck | R. E. Ginna #1 | R. E. Ginna #1 | Maine Yankee |
| 11 | Susquehanna #1* | Susquehanna #1* | Haddam Neck | Susquehanna #1* | Susquehanna #1* |
| 12 | Millstone #2 | Millstone #2 | Three Mile Island #1 | Haddam Neck | Nine Mile Point #1* |
| 13 | Three Mile Island #1 | Indian Point #2 | Millstone #2 | Three Mile Island #1 | Three Mile Island #1 |
| 14 | Indian Point #2 | Three Mile Island #1 | Indian Point #2 | Millstone #2 | Haddam Neck |
| 15 | WPPSS #2* | Rancho Seco #1 | Rancho Seco #1 | Rancho Seco #1 | Trojan #1 |
| 16 | Rancho Seco #1 | Trojan #1 | Trojan #1 | Indian Point #2 | Rancho Seco #1 |
| 17 (lowest) | Trojan #1 | WPPSS #2* | WPPSS #2* | Trojan #1 | Indian Point #2 |

* Boiling water reactor.

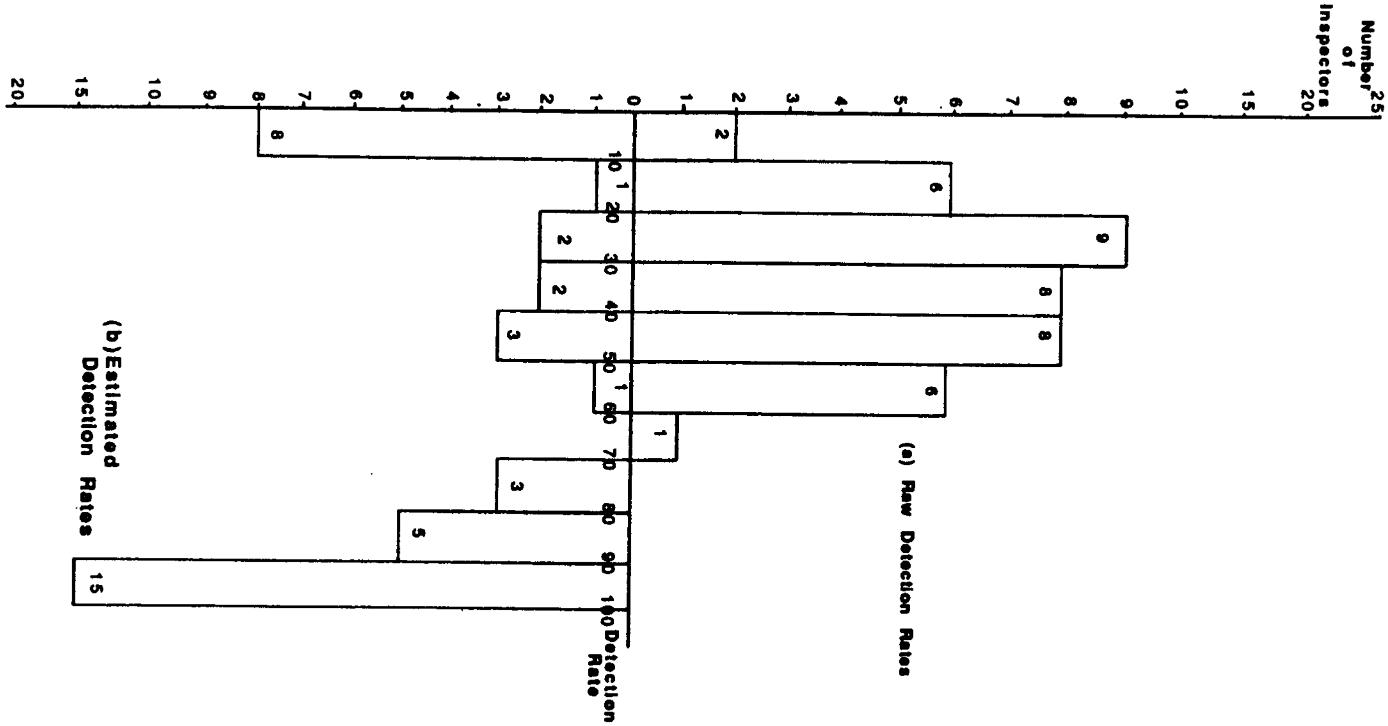


FIG. 5.—Inspector detection rates (probit model 4)

detection variable RESINSPECTION is also positive and significant, while LAG.EVENTS achieves only marginal significance, in model 4.

Table 5 presents estimates of four random Poisson models that parallel the four binary choice models of table 3. These models are estimated using a two-stage nonlinear least squares approach; the first-round estimates provide a means of estimating $\eta^2$, the variance of the random effect. In models 1, 2, and 3, $\eta^2$ is positive, while in model 4, it is estimated as zero, indicating that sufficient explanatory variables have been included in the model to reduce the random Poisson specification to the simpler pure Poisson.

The results in table 5 are qualitatively similar to those in table 3, though typically coefficients are of larger magnitude (reflecting the change in dependent variable), and will not be discussed in detail. One difference is that the deterrence variables attain significance of the wrong sign in model 4.

Plant noncompliance rankings based on models 2 and 4 are presented in table 4. Notice that the rankings of several plants have changed; most notably, WPPSS #2 is ranked sharply higher in both sets of new rankings, reflecting the fact that it was charged with multiple violations unusually frequently.

Figure 6 presents the distribution of estimated detection rates among the 40 chosen inspectors; the distribution is quite similar to that of figure 5 but reflects a slightly higher estimated average detection rate.

Models 4 of tables 3 and 5 allow computation of the rate of undetected violations, which was discussed in Section II and is computed as a projection based on the models' estimates. In the binary choice model 4, 318 inspections (31 percent) led to at least one detected citation, and it is estimated that in an additional 137 inspections (13.5 percent), at least one infraction occurred but none was detected; the standard error of this estimate is 0.2 percent. Estimates based on the Poisson model 4 must be interpreted as expected rates of violation; thus the number of detected violations in the data is 443 (0.437 per inspection), and the expected number of undetected violations is computed to be 154 or 0.15 per inspection, with a standard error of 0.0047. These estimates must be treated with caution, as we discussed earlier, but suggest that nondetection may be a problem, particularly in combination with the earlier finding of heterogeneity among inspectors.

Table 6 presents the final set of estimates on noncompliance and detection, a series of model extensions. Of special interest is model 3, a Detection Controlled model that incorporates persistence.[14] According to this model, the half-life of violations that escape detection during an initial inspection is approximately 3–4 days; the probability is roughly one-half that such an undetected violation will persist that long. Interestingly, this number accords well with experience garnered from the Three Mile Island accident, at which a contributing factor was a misset value that had apparently not been noticed during an inspection 3 days earlier (President's Commission 1979). Estimates of this model differ somewhat from those of the earlier models, especially in that the overall levels of noncompliance are considerably lower; this is not surprising when it is recognized that the persistence model exposes a single violation to the possibility of detection during more than one inspection; hence each violation "goes further."

[14] In computation of these estimates, the gap in days between inspections (see Sec. II) was divided by 7.0 to improve numerical procedures.



Model 1 of table 6 is a binary choice model that allows nonzero correlation between the errors of the violation and detection equations; the correlation parameter was found through grid search.[15] The correlation is positive but not significantly different from zero on the basis of a likelihood ratio test comparing this model with model 4 of table 3. Finally, model 2 is a persistence model without detection, the only difference between this model and Poisson model 2 being that the gap between inspections factors into the Poisson parameter as discussed in Section II.

## V. The Relationship between Undetected Violations and Events

The previous analysis has focused on power plant violations of safety regulations and the ability of the NRC to detect these violations. While analyzing and controlling noncompliance is an important aspect of regulatory effectiveness, it is not the only aspect. A second issue is the relevance of the safety standards themselves. From the viewpoint of statistical analysis, this second topic may be posed as the question, Does noncompliance significantly increase safety risks?

To address this question, a number of causality tests of the type proposed by Granger (1969) and Sims (1972) are presented in table 7. In these regressions the dependent variable is EVENTS, a monthly series whose value is the number of abnormal occurrences reported by the power plant that month. Abnormal occurrences as summarized by EVENTS are the best measure of safety risks available; nonetheless EVENTS is not equivalent to those risks, and so the results must be interpreted only as a rough guide to policy-making. The models reported in table 7 are random Poisson models similar to

[15] This model requires a one-dimensional integration since it includes a bivariate normal; integration was performed numerically using Bode's eight-point rule, as discussed in Abramowitz and Stegun (1964).

TABLE 5

RANDOM POISSON MODELS OF NONCOMPLIANCE AND DETECTION

| | Model 1: Basic Poisson | Model 2: Poisson Plant Effects | Model 3: Basic DCE | Model 4: DCE Plant Inspection Effects |
|---|---|---|---|---|
| Violation equation: | | | | |
| CONSTANT | −1.00 (.202) | −1.46 (.498) | −.862 (.197) | −1.28 (.500) |
| SANCNATIONAL | −.00620 (.212) | −.165 (.213) | .119 (.212) | .455 (.165)** |
| SANCPLANT | −.0725 (.208) | −.0953 (.230) | −.0644 (.196) | .308 (.145)** |
| BONDRATING | .0327 (.016)** | .00945 (.0463) | .0314 (.0157)** | −.00530 (.0469) |
| BWR | .327 (.110)** | .532 (.547) | .340 (.107)** | .754 (.470) |
| CHANGERESINS | −.858 (.499) | −.781 (.485) | −.886 (.510) | −.781 (.639) |
| DOWN | −.369 (.150)** | −.350 (.159)** | −.334 (.147)** | −.538 (.154)** |
| AGE | −.0230 (.0138) | −.0163 (.0137) | ... | ... |
| Plant effects: | | | | |
| Haddam Neck | | .469 (.364)[a] | | .197 (.352)[a] |
| Millstone #2 | | .281 (.345) | | .860 (.312) |
| Maine Yankee | | .569 (.343) | | .553 (.363) |
| Pilgrim #1 | | .610 (.535) | | .406 (.435) |
| Salem #1 | | 1.06 (.367) | | 1.02 (.399) |
| Indian Point #2 | | .122 (.396) | | −.253 (.472) |
| Nine Mile Point #1 | | .0794 (.579) | | −.493 (.486) |
| R. E. Ginna #1 | | .521 (.354) | | .950 (.321) |
| Fitzpatrick | | .225 (.759) | | .150 (.719) |
| Peach Bottom #3 | | .506 (.469) | | .145 (.389) |
| Beaver Valley #1 | | .982 (.316) | | 1.13 (.356) |
| Three Mile Island #1 | | .349 (.374) | | .209 (.410) |
| Susquehanna #1 | | −.0242 (.674) | | −.412 (.613) |
| San Onofre #1 | | .874 (.391) | | .804 (.444) |
| Rancho Seco #1 | | .230 (.412) | | −.191 (.490) |
| Detection equation: | | | | |
| CONSTANT | | | .355 (.208) | 1.31 (.418) |
| RESINSPECTION | | | 3.57**[b] | 2.23 (.766)** |
| POST-TMI | | | 3.97**[b] | 1.84 (.675)** |
| LAG.EVENTS | | | −.0306 (.0397) | −.239 (.066)** |
| $\eta^2$ (first round) | .470 | .299 | .321 | .0 |
| SSR | 541 | 524 | 527 | 493 |
| $R^2$ | .276 | .298 | .294 | .341 |

** Statistically significant at the 95 percent level.
[a] Plant effects jointly statistically significant at the 95 percent level.
[b] Likelihood flat. Statistically significant at the 95 percent level.

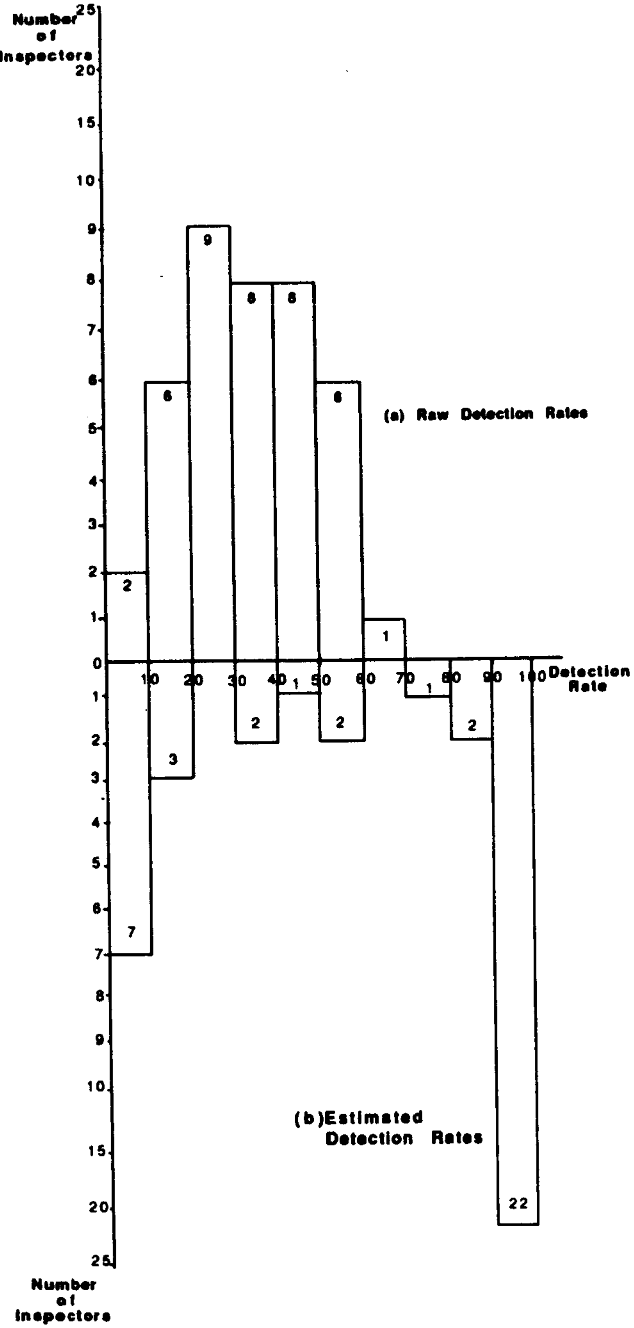


FIG. 6—Inspector detection rates (Poisson model 4)



those discussed earlier, in which EVENTS is the dependent variable, and the Poisson parameter is specified as $(\mathbf{X}\beta)e^{\eta}$, where $\eta$ is a random effect with variance $\zeta^2$. The vector $\mathbf{X}$ includes a constant, the lagged value of EVENTS from the previous month (LAG.EVENTS), plant effects, and a measure of undetected violations (UNDETECTED-VIOLS). The variable UNDETECTEDVIOLS is computed on the basis of the estimates of Poisson model 4, according to two different specifications: in the first a violation that remains undetected is assumed to persist 7 days; in the second, 30 days. To calculate UNDETECTEDVIOLS, each day of the month was taken in turn, a window of either 7 or 30 days was established backward in time from that date, and the cumulative number of undetected violations estimated (as discussed in Sec. II) from all inspections occurring in that window was computed; UNDETECTEDVIOLS is the sum of these "exposure" rates divided by 30. Estimation is complicated by the possibility of serial correlation in the random effects $\eta$ from one month to the next (especially since the models include a lagged dependent variable) and is further complicated by the fact that UNDETECTEDVIOLS is not directly observed, but only its estimate. Under the assumption that the covariance in the random effects is $\rho$, the models of table 7 are estimated using a two-stage procedure in which the first stage optimizes jointly over $\beta$ and $\rho$ and is used to compute an estimate of $\zeta^2$; the second stage reweights the observations, partially differences the observations, and reestimates the model.[16]

The variable UNDETECTEDVIOLS is positive and statistically significant when the window is 7 days and positive but insignificant when the window is 30 days. Overall, there is support for the view that violations do increase safety risks. In addition, LAG.EVENTS is positive and statistically significant in both regressions, providing evidence that the system processes generating failures are highly serially correlated, a finding that itself has policy implications. Both of these findings are made more persuasive by the fact that they emerge within models that include plant effects. The plant effects themselves are jointly statistically significant in both models.

While these findings are interesting, a number of objections to the analysis can be raised. First, a full-equilibrium model that includes the allocation of inspectors to plants has not been set up: undetected

[16] Specifically, the variance of an observation is $(\mathbf{X}\beta)^2\zeta^2 + \mathbf{X}\beta + \gamma\Sigma\, w_j^2\lambda_j$, where $\gamma$ is the coefficient on UNDETECTEDVIOLS, $w_j$ is the number of days the $j$th inspection is in the window of, and $\lambda_j$ is a variance term. Similarly the covariance is

$$(\mathbf{X}_t\beta)(\mathbf{X}_{t-1}\beta)\rho + \gamma^2(\Sigma\, w_{kt}w_{kt-1}\lambda_k)(1 + \rho),$$

where the sum runs over all inspections shared in common between the current month and the last.

## TABLE 6

**Extensions**

| | Model 1: Binary Correlation | | Model 2: Non-DCE Persistence | | Model 3: DCE Persistence | |
|---|---|---|---|---|---|---|
| Violation equation: | | | | | | |
| CONSTANT | −.0169 | (.716) | −1.96 | (.368) | −1.30 | (.713) |
| SANCNATIONAL | −.369 | (.421) | −.0981 | (.170) | −1.19 | (.512)** |
| SANCPLANT | .410 | (.502) | .102 | (.127) | 1.30 | (.169)** |
| BONDRATING | −.00497 | (.113) | .00643 | (.0395) | .179 | (.0503)** |
| BWR | −.272 | (.721) | .641 | (.395) | −.689 | (.700) |
| CHANGERESINS | −.646 | (.627) | −.643 | (.160)** | .557 | (.368) |
| DOWN | −.363 | (.251) | −.0948 | (.103) | −1.19 | (.681) |
| Plant effects: | | | | | | |
| Haddam Neck | .431 | (2.57)[a] | −.0890 | (.290)[a] | −2.03[a,b] | |
| Millstone #2 | .277 | (.829) | .0397 | (.211) | −2.05 | (.417) |
| Maine Yankee | .824 | (1.07) | .353 | (.254) | −1.60 | (.283) |
| Pilgrim #1 | 3.25[b] | | .160 | (.402) | −.809 | (.828) |
| Salem #1 | 3.20[b] | | .701 | (.291) | −2.02 | (.221) |
| Indian Point #2 | .0773 | (1.14) | −.112 | (.270) | −2.83 | (.348) |
| Nine Mile Point #1 | 1.08 | (1.23) | −.541 | (.481) | −.659 | (.869) |
| R. E. Ginna #1 | .479 | (.994) | −.0535 | (.270) | −1.07 | (.333) |
| Fitzpatrick | 1.20 | (1.68) | −.0278 | (.620) | −1.76 | (.679) |
| Peach Bottom #3 | 4.29[b] | | −.0212 | (.371) | −.797 | (1.01) |
| Beaver Valley #1 | 2.18 | (2.34) | .455 | (.244) | −1.391 | (.345) |
| Three Mile Island #1 | .166 | (.595) | −.103 | (.336) | −.673 | (.406) |
| Susquehanna #1 | 1.21 | (.749) | −.200 | (.470) | 1.27[b] | |
| San Onofre #1 | .883 | (1.41) | .791 | (.290) | −1.85 | (.266) |
| Rancho Seco #1 | .127 | (1.21) | .0230 | (.307) | −1.81 | (.217) |
| Detection equation: | | | | | | |
| CONSTANT | .480 | (1.07) | | | −2.27 | (.430) |
| RESINSPECTION | 1.80 | (.788)** | | | −2.20 | (.402)** |
| POST-TMI | 1.51 | (.702)** | | | .665 | (.419) |
| LAG.EVENTS | −.125 | (.0918) | | | .687 | (.0931)** |
| Log likelihood | −537.9 | | ... | | ... | |
| SSR | ... | | 557 | | 2,958 | |
| ρ | .50 | | ... | | 1.6 | |

** Statistically significant at the 95 percent level.
[a] Plant effects jointly statistically significant at the 95 percent level.
[b] Likelihood flat. Statistically significant at the 95 percent level.

TABLE 7

Causality Tests for the Variable Events

| | 7-Day Window | | 30-Day Window | |
|---|---|---|---|---|
| CONSTANT | .799 | (.235) | .877 | (.246) |
| LAG.EVENTS | .430 | (.0603)** | .415 | (.0629)** |
| UNDETECTEDVIOLS | .0449 | (.0199)** | .00627 | (.00587) |
| Plant effects: | | | | |
| Haddam Neck | .134 | (.299)[a] | .104 | (.308)[a] |
| Millstone #2 | .556 | (.424) | .697 | (.439) |
| Maine Yankee | .663 | (.365) | .764 | (.384) |
| Pilgrim #1 | 1.26 | (.518) | 1.45 | (.545) |
| Salem #1 | 1.03 | (.547) | 1.18 | (.573) |
| Indian Point #2 | .462 | (.366) | .411 | (.375) |
| Nine Mile Point #1 | .824 | (.393) | .836 | (.405) |
| R. E. Ginna #1 | −.491 | (.281) | −.492 | (.289) |
| Fitzpatrick | 2.26 | (.612) | 2.50 | (.654) |
| Peach Bottom #3 | .392 | (.384) | .417 | (.409) |
| Beaver Valley #1 | .376 | (.539) | .848 | (.569) |
| Three Mile Island #1 | −.499 | (.269) | −.468 | (.279) |
| Susquehanna #1 | .832 | (.600) | .806 | (.619) |
| San Onofre #1 | .0919 | (.333) | .0849 | (.345) |
| Rancho Seco #1 | .720 | (.390) | .701 | (.399) |
| WPPSS #2 | .993 | (.261) | 1.05 | (.277) |
| $\zeta^2$ | .234 | | .238 | |
| $\rho$ | −.00343 | | .00468 | |
| SSR | 534 | | 521 | |

** Statistically significant at the 95 percent level.
[a] Plant effects jointly significant at the 95 percent level.

violations at some plants may be more or less likely to lead to serious hazards, and the NRC may be aware of this. Second, not all regulations are equivalent: some regulations may never be violated, and those that are violated may represent the "tail," that is, those that are least important for safety. Both of these objections, however, would seem to imply that the impact of UNDETECTEDVIOLS might be understated; since it remains statistically significant in one of the models, the hypothesis that the standards do affect safety remains viable. A final objection, which tends to imply that the impact of UNDETECTEDVIOLS might be overstated, is that the NRC responds to events in one month by citing many violations the next month; since events are so highly serially correlated, UNDETECTEDVIOLS may appear to be significant only because of an errors-in-variables problem. In response to this objection, it should be recalled that the earlier analysis in fact found a negative relationship between LAG.EVENTS and detection rates; nonetheless a more fully simultaneous model has not been estimated.



## VI. Conclusions

This paper has explored a number of issues central to NRC safety regulation of U.S. commercial nuclear power plants. Among the findings, those especially of interest include the rankings of power plants by propensity to noncomply, which by and large accord well with public perceptions, considerable heterogeneity among NRC inspectors in detection and citation practices, a sharp increase in detection following the Three Mile Island accident, and a positive relationship between undetected violations and future abnormal occurrences.

The statistical methods of this paper bear a strong resemblance to the probabilistic risk assessment method of quantifying risk at nuclear power plants. Extensions of the models to more fully reflect specific reactor event trees would help align the regulatory agenda more fully with this engineering approach to safety.

## References

Abramowitz, Milton, and Stegun, Irene A., eds. *Handbook of Mathematical Functions, with Formulas, Graphs, and Mathematical Tables*. Washington: Nat. Bur. Standards, 1964.

Baron, David P., and Besanko, David. "Regulation, Asymmetric Information, and Auditing." *Rand J. Econ.* 15 (Winter 1984): 447–70.

Cullingford, M. C.; Shah, S. M.; and Gittus, J. H., eds. *Implications of Probabilistic Risk Assessment*. Essex: Elsevier, 1987.

Dubin, Jeffrey A., and Rothwell, Geoffrey S. "Risk and Reactor Safety Systems Adoption." Manuscript. Stanford, Calif.: Stanford Univ., 1987.

Feinstein, Jonathan S. "Detection Controlled Estimation." Manuscript. Stanford, Calif.: Stanford Univ., 1987.

Gourieroux, Christian; Monfort, Alain; and Trognon, Alain. "Pseudo Maximum Likelihood Methods: Applications to Poisson Models." *Econometrica* 52 (May 1984): 701–20.

Granger, Clive W. J. "Estimating Causal Relations by Econometric Models and Cross-spectral Methods." *Econometrica* 37 (July 1969): 424–38.

International Nuclear Safety Advisory Group. *Summary Report on the Post-Accident Review Meeting on the Chernobyl Accident*. Safety Series, no. 75-INSAG-1. Vienna: Internat. Nuclear Safety Advisory Group, 1986.

Joskow, Paul L., and Rozanski, George A. "The Effects of Learning by Doing on Nuclear Plant Operating Reliability." *Rev. Econ. and Statis.* 61 (May 1979): 161–68.

Konstantinov, L. V. "Probabilistic Safety Assessment in Nuclear Safety: International Developments." In *Implications of Probabilistic Risk Assessment*, edited by M. C. Cullingford, S. M. Shah, and J. H. Gittus. Essex: Elsevier, 1987.

Laffont, Jean-Jacques, and Tirole, Jean. "Using Cost Observation to Regulate Firms." *J.P.E.* 94, no. 3, pt. 1 (June 1986): 614–41.

Megaw, James. *How Safe?* Toronto: Stoddart, 1987.

Nichols, Elizabeth, and Wildavsky, Aaron. "Nuclear Power Regulation: Seeking Safety, Doing Harm?" *Regulation*, no. 1 (1987), pp. 45–53.
President's Commission on the Accident at Three Mile Island. *The Need for Change, the Legacy TMI: Report*. New York: Pergamon, 1979.
Sims, Christopher A. "Money, Income, and Causality." *A.E.R.* 61 (September 1972): 540–52.
U.S. Atomic Energy Commission. *Reactor Safety Study: An Assessment of Accident Risks in U.S. Commercial Nuclear Power Plants*. Washington: U.S. Atomic Energy Comm., 1974.