Certainty of Punishment versus Severity of Punishment: Enforcement of Environmental Protection Laws

Dietrich Earnhart
Professor
University of Kansas - Department of Economics
earnhart@ku.edu

Lana Friesen
Professor
University of Queensland - School of Economics
l.friesen@uq.edu.au

July 20, 2022

Abstract: According to the standard enforcement model, the key deterrence components are punishment certainty and severity. While theory predicts the relative efficacy of certainty versus severity, empirical and experimental evidence are mixed. Our study represents the first to systematically compare the effects of certainty and severity in the environmental protection context. Our empirics examine wastewater discharged by chemical manufacturing facilities permitted under the Clean Water Act. We find that, when enforcement certainty and severity are high, both components effectively deter pollution, with certainty more effective. In contrast, certainty and severity increases prove counter-productive when certainty and severity are low.

JEL Codes: K32, K42, Q53

Appendix materials can be accessed online at:
https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Earnhart-appA.pdf
https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Earnhart-appB.pdf
https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Earnhart-appC.pdf
https://uwpress.wisc.edu/journals/pdfs/LE-99-2-Earnhart-appD.pdf

doi:10.3368/le.030521-0024R1

1. Introduction

According to the standard theoretical model of enforcement (Becker, 1968; Polinsky and Shavell, 2000), the two key components of deterrence are the certainty of punishment and the severity of punishment. The certainty of punishment refers to the probability of punishment being imposed if an offense is committed, while the severity of punishment reflects the magnitude of the punishment imposed.[1] Becker (1968) demonstrates that the relative efficacy of the two components in deterring crime depends on the risk preferences of the decision maker. In particular, risk averse potential offenders are more deterred by increases in severity than equivalent increases in the certainty of enforcement, with the opposite true for risk lovers, while risk neutral decision makers are equally deterred. However, evidence regarding the relative efficacy of enforcement certainty and severity is mixed. While studies using general crime data mostly find that increasing certainty is a more effective deterrent than increasing severity (e.g., Grogger, 1991; Eide, 2000; Nagin, 2013), several experimental studies find the opposite (Anderson and Stafford, 2003; Friesen, 2012). We contribute to this debate by comparing the effects of certainty and severity of enforcement in a new context: the enforcement of environmental protection laws. To our knowledge, our study is the first to compare systematically the effects of certainty and severity of enforcement on environmental compliance.[2] Our results are the necessary first step in identifying the deterrence lever which enforcement agencies should emphasize.[3]

Our second contribution is methodological. In particular, our empirical specification is novel as it allows the effects of the certainty and severity of enforcement to vary with the size of the other enforcement component. By incorporating an interaction term between the two enforcement components, we allow the deterrent effect of each enforcement component to depend on the strength of the other component. Our specification is consistent with the standard theoretical model of enforcement (Becker, 1968), where the effect of increasing enforcement severity is stronger when the certainty of enforcement is larger and vice versa; i.e., the two enforcement components are complements.[4] In addition, our specification enables a more thorough comparison of these enforcement components than a simple reliance on average effects.

In order to make these contributions, our study examines the wastewater discharged by U.S. chemical manufacturing facilities permitted within the Clean Water Act's National Pollutant Discharge Elimination System (NPDES). To control discharges from point sources, the NPDES system imposes effluent limits on wastewater pollutants. To ensure compliance with these limits, federal and state agencies monitor facilities, i.e., conduct inspections, and take enforcement actions against identified violations. Reflecting this regulatory context, we measure enforcement certainty as the number of sanctions imposed per facility-month of operation and measure enforcement severity as the conditional average sanction magnitude, i.e., monetary amount of imposed sanctions divided by the number of sanctions. We measure environmental compliance as the ratio of actual (reported) discharges to permitted discharges, i.e., discharge ratio, which captures the extent of compliance. By examining the extent of compliance, our study is able to examine both improvement toward compliance and improvement beyond compliance, which represents over-compliance (Earnhart, 2004a; Earnhart, 2009).

Our empirical results reveal that increasing either the certainty or severity of enforcement can be a significant deterrent; however, the magnitude of each effect varies with the size of the other enforcement component. In particular, when enforcement certainty and severity are high, both components are effective at deterring pollution, although certainty is more effective. This result is consistent with empirical results using general crime data (e.g., Grogger, 1991; Eide, 2000; Nagin, 2013). In contrast, when enforcement certainty and severity are weak, increasing either component is counter-productive, prompting less compliance rather than more, with the detrimental effect larger for certainty. For intermediate values, the results are more mixed.

While our empirical study cannot identify the specific mechanism driving our estimates, our results provide a cautionary tale for regulators. Maintaining a minimum level of enforcement for both deterrence levers proves important especially since the estimated effects lie in policy-relevant ranges of the enforcement parameters. In addition, our results demonstrate the importance of including interaction terms between the two components of deterrence in the empirical specification. Our estimated interactions reveal that certainty and severity of enforcement are complements, a nuance hidden by a reliance on average deterrent effects.

The rest of the study proceeds as follows. The next section describes the relevant regulatory context. Section 3 describes the data. Section 4 constructs the econometric framework and explains the econometric methods. Section 5 interprets the estimation results. Section 6 concludes.

2. Regulatory Context

The empirical component of our study examines the wastewater discharged by U.S. chemical manufacturing facilities permitted within the U.S. Clean Water Act's National Pollutant Discharge Elimination System (NPDES) between 1995 and 2001.[5] This sector generates a large amount of wastewater, with four of the 10 most polluting sub-sectors operating in the chemical manufacturing sector (EPA, 2011). As the primary form of control within the NPDES system, regulatory agencies issue facility-specific permits to facilities regulated as point sources. These permits specify the pollutant-specific discharge limits imposed on facilities. Due to considerations over local ambient water quality, discharge limits differ across facilities and time even within the same industry at the same moment in time.[6] To control for this variation, we compare actual (reported) discharges to permitted discharges and calculate the "discharge ratio", which measures the extent of compliance.

The issued NPDES permits require regulated facilities to monitor and self-report their discharges on a regular, generally monthly, basis by completing and submitting Discharge Monitoring Reports (DMRs) to regulatory agencies.[7] Thus, information on discharges is not limited by government monitoring, i.e., inspections. Since limits and discharges are pollutant-specific, our analysis must focus on certain pollutants in order to assess the extent of compliance. Our study focuses on the pollutant most common to the sampled facilities – total suspended solids (TSS).[8]

To ensure compliance with the issued permit limits, the EPA and state agencies periodically inspect facilities and take enforcement actions as needed. While the EPA retains authority to monitor and sanction facilities, state agencies are primarily responsible for monitoring and enforcement in states with primacy, which applies for all but two states in our sample. Inspections represent the backbone of environmental agencies' efforts to collect evidence for enforcement (Wasserman, 1984), maintain a regulatory presence (EPA, 1990), and offer technical assistance as opportunities arise. Federal and state agencies also take informal enforcement actions (e.g., warning letters) and formal enforcement actions (e.g., fines). For our analysis, we consider both federal and state inspections and all types of federal formal enforcement actions that impose a monetary burden (hereafter "sanctions"): fines, injunctions, and supplemental environmental projects (i.e., court-imposed orders to perform environmentally beneficial acts, e.g., wetland restoration).[9]

Our empirical analysis focuses on the 508 large or "major" chemical manufacturing facilities permitted under the U.S. Clean Water Act between 1995 and 2001. These facilities, which represented 21 % of the chemical facilities in the NPDES system in 2001, were responsible for the bulk of wastewater discharges from this sector during the sample period (Earnhart and Glicksman, 2011) and correspondingly were the focus of regulatory efforts (Earnhart, 2009; Earnhart and Segerson, 2012).[10]

3. Data

For the construction of the dependent and independent variables, we gather data from various sources: EPA Permit Compliance System database (discharge limits, actual discharges, permit features, inspections, fines); EPA Docket database (fines, injunctions, SEPs); and Compustat database (firm-level ownership structure). Appendix B provides details.

The broadest sample includes all of the monthly observations for the 508 facilities that were active at some point over the sample period: January, 1995, to June, 2001. To remain in the sample, a given facility must discharge TSS at least once during the seven-year sample period, which identifies 461 facilities. Moreover, a given facility must face an effluent limit for TSS in the particular month of discharge, which reduces the sample size from 32,378 to 29,226. Since some measures are based on a preceding 12-month period (e.g., cumulative sanction count), the regression sample period starts January, 1996, which reduces the sample size to 23,193, consisting of 406 facilities.[11]

Table 1 provides statistical summaries of the formulated dependent variable and regressors. Table 1 summarizes the environmental performance measure, the ratio of self-reported TSS discharges to the permitted discharge limit ("discharge ratio"). The mean discharge ratio of 0.315 implies that facilities on average generate TSS discharges that are 68.5

% below their monthly limits. This average indicates a need to analyze the extent of compliance rather than the status of compliance. The maximum discharge ratio of 9.080 implies that facilities allow their TSS discharges to surge as high as 808.0 % above the permitted limits. This maximum indicates a need to analyze the degree of non-compliance rather than the extent of non-compliance.

[Insert Table 1 here]

Based on our study and previous studies, regulated facilities regularly over-comply with their effluent limits (e.g., Earnhart, 2009; Earnhart and Segerson, 2012; Bandyopadhyay and Horowitz, 2006; Shimshack and Ward, 2008). The economics literature identifies various factors that may prompt over-compliance including pressure exerted by other external parties, such as local communities (Henriques and Sadorsky, 1996; Dasgupta et al., 2000), investors (Downing and Kimball, 1982), and customers (Arora and Cason, 1996). Stochastic discharge patterns and jointness in production across pollutants (Bandyopadhyay and Horowitz, 2006; Shimshack and Ward, 2008), as well as the possibility of regulatory errors (Rousseau, 2009), may also play a role. Over-compliance does not disrupt our analysis of enforcement. In the presence of stochastic discharges, enforcement remains relevant even when regulated facilities choose to over-comply with their discharge limits (Earnhart and Segerson, 2012; Shimshack and Ward, 2008).[12] Our empirical analysis incorporates regressors that attempt to control for factors that might induce over-compliance. As examples, the standard deviation of the discharge ratio proxies for discharge stochasticity and ownership structure helps to capture variation in investor pressure.[13] Table 1 also summarizes the regressors.

4. Regression Variables and Econometric Analysis

Regression Variables

To assess empirically the relative efficacy of enforcement certainty and severity, we construct an econometric relationship between our dependent variable, the facility's chosen discharge ratio (Earnhart, 2004a,b,c; Earnhart, 2009; Earnhart and Segerson, 2012), and a set of primary explanatory variables measuring the certainty and severity of enforcement, along with control factors.

Each facility must form expectations about the certainty and severity of enforcement before selecting its extent of compliance. Our empirical analysis assumes that each facility bases its expectations of future enforcement on the experiences of other similar facilities along with its own recent experiences (Shimshack and Ward, 2005; Earnhart, 2004a,b,c; Earnhart, 2009). In particular, general deterrence reflects the *ex ante* general threat of future enforcement based on the recent experiences of *other* facilities with regulatory interventions (Sah, 1991; Cohen, 2000), while specific deterrence adjusts this general threat based on the specific enforcement experiences of a particular facility in the recent past (Cohen, 2000; Earnhart and Friesen, 2013).[14] Since facilities need at least a few weeks, if not several months, to respond to others' and their own experiences with regulatory interventions, our analysis lags the measures of enforcement actions (Earnhart, 2004a,b; Earnhart, 2009; Shimshack and Ward, 2005).

For the following reasons, our study focuses exclusively on general deterrence and uses specific deterrence only as an additional control factor in an alternative model within our sensitivity analysis. First, general deterrence is more consistent with the theoretical deterrence model, given the emphasis on the *ex ante* assessment of enforcement threats. Second, specific deterrence is likely endogenously determined with respect to a facility's choice of discharge ratio. Thus, the inclusion of these likely endogenous regressors generates endogeneity bias. Since exclusion of these potentially relevant regressors generates omitted variable bias, our sensitivity analysis allows us to assess both biases.[15] Third, the evidence of enforcement-based specific deterrence is rather mixed. While one study provides evidence of specific deterrence (Earnhart, 2004b), other studies provide only mixed evidence (Foulon et al., 2002) or no evidence (Eckert, 2004; Shimshack and Ward, 2005). Most important, no study of industrial polluters finds evidence of specific deterrence when the analysis also controls for general deterrence.

For our comparison of enforcement certainty versus severity, we construct theoretically motivated measures of sanction certainty and sanction severity (Becker, 1968; Polinsky and Shavell, 2000). Since the certainty of punishment refers to the probability of punishment being imposed if an offense is committed, we use the number of sanctions imposed as our measure of sanction certainty, adjusted for the number of relevant facilities (as described below). For our measure of sanction severity, we use the average magnitude of sanctions imposed.

Given this approach, we construct the following general deterrence factors. To capture general deterrence stemming from enforcement, we compute the aggregate count and monetary value of sanctions (in dollars) imposed against all <u>other</u> similar facilities – major chemical facilities in the same state – over the preceding 12-month period. To measure enforcement certainty, we divide each aggregate count of sanctions by the number of months all other major chemical facilities were operating in each state over the given 12-month period. This division controls for differences in the extent of operation of major chemical facilities across states and over time. The resulting measure represents the likelihood of a sanction imposed on another similar facility per month of operation (i.e., number of sanctions per facility-month). To calculate the conditional average sanction as our measure of enforcement severity, we divide each aggregate monetary value of sanctions by the aggregate count of sanctions (i.e., dollars per sanction).

In our sensitivity analysis, we explore the robustness of our results to alternative specifications of these chosen key measures. First, we construct a measure of sanction certainty based on the count of violations (per violation basis) rather than the count of operating facilities (per facility basis), arguably an even more theoretically appropriate measure. Second, while our use of a preceding 12-month window is consistent with previous studies (e.g., Earnhart, 2004a; Earnhart, 2009; Earnhart and Friesen, 2017), we also explore the use of a 6-month window and a 24-month window. Third, although state boundaries are commonly used in the literature (e.g., Shimshack and Ward, 2005), we also consider an alternative geographical construction of our general deterrence measures involving EPA regional boundaries.[16]

Our theoretically grounded construction of enforcement certainty is relatively novel in the literature, as most previous studies use inspections as a proxy for enforcement certainty (e.g., Magat and Viscusi, 1990; Laplante and Rilstone, 1996; Shimshack and Ward, 2005). Sigman (1998, p.167) is the closest conceptually in constructing the "probability with which a facility expects a fine for dumping" as the ratio of the number of enforcement actions to the number of facilities in a state. (However, she exploits only cross-sectional differences in the context of illegal oil dumping.) We argue that sanction count is a more theoretically relevant measure of the likelihood of a *sanction*, as opposed to the likelihood of an inspection. Moreover, inspections serve as a weak proxy for four reasons. First, inspections very rarely are able to detect non-compliance with discharge limits. Instead, nearly all inspections are only able to gather corroborating evidence for any possible future enforcement action. Second, inspections are not necessarily prompted by non-compliance, which mutes the threat of inspections leading to sanctions. These "routine inspections" are not intended or designed to gather corroborating evidence to support a possible enforcement case in the future. Rather these routine inspections are better positioned to offer technical assistance. Third, inspections are not necessary to impose sanctions: self-reported data on non-compliance are sufficient to prompt sanctions. Thus, a focus on inspections ignores cases of flagrant non-compliance for which self-reported data may be sufficient to prompt enforcement. Fourth, inspections that identify/confirm non-compliance need not lead to sanctions due to enforcement discretion on the part of regulatory agencies. According to Earnhart and Glicksman (2011), enforcement personnel enjoy vast discretion over their decisions about who and when to sanction. While the threat of sanction may generally rise as the count of inspections grows, the opposite may hold when inspections serve as a substitute for sanctions. This negative correlation is most likely when agencies conduct inspections in order to provide technical assistance to facilities that lack the capacity rather than the desire to comply.[17] Lastly, two previous studies on regulatory enforcement support our claim that use of enforcement actions to construct a measure of enforcement likelihood dominates the use of inspections to construct this measure; Gray and Scholz (1991, 1993) find that only OSHA inspections resulting in a penalty, not all types of inspections, lower injury rates.

We do, however, control for the influence of inspections on compliance decisions. We interpret the effect of inspections as reflecting regulatory presence, technical assistance, and regulatory burden (i.e., hassle of hosting the regulatory inspector, which represent costs independent of sanctions). By controlling for the use of inspections, we effectively interpret the effect of sanction count as capturing a facility's response to an agency's willingness to translate inspections into enforcement, i.e., certainty of enforcement conditional on inspections, while the severity of punishment reflects how large that punishment is. As with enforcement, regulated facilities must gauge the threat of an inspection. To capture this threat, our analysis employs a proxy based on the annual aggregate count of inspections against <u>other</u> similar facilities – major chemical facilities in the same state – over the preceding 12-month period (i.e., inspections per facility-month). As these aggregate inspections are akin to enforcement-related general deterrence, we include this control factor in our base model.

Including measures of enforcement severity in environmental enforcement studies is a relatively recent development, with earlier studies focusing only on inspections (e.g. Magat and Viscusi, 1990; Laplante and Rilstone, 1996). Studies that do include a measure of enforcement actions typically utilize an indicator of presence (e.g., Earnhart, 2004) or the state statutory maximum (e.g., Sigman, 1998). More like our analysis, Shimshack and Ward (2005, 2008) demonstrate the general deterrent effect of state fines whether measured by an indicator or the sum of fines imposed. Earnhart and Friesen (2021a) use the same measure of sanction severity as in our study in their comparison of federal and state enforcement in four key states, along with a comparison of specific and general deterrence.

Consistent with the standard theoretical model (Becker, 1968), we allow our general deterrence measures of the certainty and severity of enforcement to influence each other by generating an interaction between the two factors. Thus, the deterrent effect of each enforcement component – certainty and severity – can depend on the size of the other enforcement component. This specification follows directly from the standard enforcement model of Becker (1968, Footnote 16). Friesen (2012) adapts the Becker model to regulatory compliance and shows the same dependence in equations (3) and (4). Intuitively, if a facility is non-compliant, its expected profits depend on the combination of the certainty and severity of enforcement (i.e., expected sanction). Therefore, the derivative of expected profit with respect to one component always depends on the other component. This focus on expected profit implicitly assumes risk neutral preferences. If instead the facility's environmental manager has risk averse preferences, the dependence remains even though the analysis involves expected utility.

Although the standard model predicts that increasing either the certainty or severity of enforcement has a deterrent effect, thus, increasing compliance, our empirical model allows these factors to prove counter-productive, which might arise when behavioral factors matter. For example, small monetary incentives might crowd out other motivations of individual environmental managers, such as an intrinsic desire to obey the law or cooperate with regulators, consistent with the substantial behavioral economics literature (Frey and Jegen, 2001; Fehr and Falk, 2002; Gneezy et al., 2011; Bowles and Polania-Reyes, 2012). While one might believe that behavioral motives should not influence corporate entities, several studies demonstrate otherwise. As examples, Nakamura et al. (2001) find that managerial attitudes towards environmental protection influence facilities' management choices, Winter and May (2001) and Earnhart and Glicksman (2015) find that coercive (rather than cooperative) enforcement styles reduce compliance, and Short and Toffel (2010) demonstrate that explicit threats by regulators undermine compliance.

In addition to general deterrence, we construct measures of specific deterrence stemming from enforcement, which we include in the sensitivity analysis, in an analogous way to our general deterrence measures. Specifically, we consider sanctions imposed on an individual facility in the preceding 12-month period, consistent with previous empirical studies (Magat and Viscusi, 1990; Helland, 1998a; Helland, 1998b; Gray and Deily, 1996; Laplante and Rilstone, 1996; Earnhart, 2004a,b; Earnhart, 2009).[18] The resulting measure captures enforcement certainty and represents the likelihood of a sanction imposed on the specific facility per month of operation (i.e., sanctions per facility-month). The conditional average sanction equals the monetary value of sanctions divided by the count of sanctions (i.e., dollars per sanction), as with the general deterrence factor. We also control for inspections conducted at the individual facility in the preceding 12-month period, consistent with several previous empirical studies (Magat and Viscusi, 1990; Helland, 1998a,b; Earnhart, 2004a,b; Earnhart, 2009). Thus, the constructed regressor measures inspections per facility-month of activity. Even though the cited studies treat lagged inspections as an exogenous regressor, we exclude these regressors from our base model for the same reason that we exclude the enforcement-related specific deterrence measures.

Lastly, we control for other factors, as described in Appendix C.

Econometric Analysis and Issues

To estimate the relationship between the discharge ratio and the regressors, we employ a mixed log-log specification. We log the dependent variable because discharge ratios are distributed lognormal.[19] Logging the enforcement-related regressors simplifies the process of comparing the effects of certainty of enforcement versus severity of enforcement because the estimated coefficients reflect elasticities. We also log the inspection-related regressors because they are distributed lognormal. We do not log the other regressors.[20] As part of our sensitivity analysis, we alternatively transform the noted variables using an inverse hyperbolic sine transformation. The estimated coefficients for the enforcement-related regressors approximately reflect elasticities (Bellemare and Wichman, 2020).[21]

We address the panel structure of our data by employing a fixed effects estimator, which incorporates facility-specific indicators. Based on this depiction of dependent and independent variables, we construct the estimating (regression) equation. Let Y_{it} denote the log-transformed discharge ratio generated by facility i in time period (year-month) t. Let X_{it} capture log-transformed general deterrence-related enforcement certainty and Z_{it} capture log-transformed general deterrence-related enforcement severity. Let K_{it} denote the facility-specific, time-variant control factors and G_t denote the year and seasonal control factors. And let F_i capture the facility fixed effects. The regression equation then follows:

$$Y_{it} = \beta X_{it} + \theta Z_{it} + \sigma(X_{it} \times Z_{it}) + \rho K_{it} + \psi G_t + F_i + \varepsilon_{it},$$

[1]

where ε_{it} denotes the error term.

As our base approach, we cluster the standard errors on the state in which a facility operates. Within our sensitivity analysis, we alternatively cluster the standard errors on the regulated facility, which helps to address serial correlation.[22]

We consider four regressor sets or model sets that differ based on the non-deterrence factors, i.e., control factors. Our base model uses Model Set A, which includes facility-specific indicators, year indicators, seasonal indicators, non-deterrence regulatory factors (permit conditions and budgetary resources). Within our sensitivity analysis, we use alternative model sets. Model Set B includes only the facility-specific indicators. Model Set C includes facility-specific indicators and year and seasonal indicators. Relative to Model Set A, Model Set D further adds facility and firm characteristics. Fortunately, our primary regressor conclusions prove fully robust to the choice of model set.

As the identifying variation for capturing the effects of general deterrence, we exploit the substantial discretion enjoyed by EPA regional offices and state regulatory agencies over their CWA enforcement and monitoring decisions (Earnhart and Glicksman, 2011). This discretion allows agencies to shape their sanction decisions to fit an enforcement strategy based on the policy concerns of individual regions and states, which vary across space and over time. For example, we exploit the switch from a more conservation-oriented EPA regional administrator to a more business-oriented EPA regional administrator. The broad nature of enforcement strategies assures the exogeneity of the general deterrence measures. Given this exogeneity, we comfortably assume that, absent any variation in the general deterrence measures, facilities chose the same extent of compliance, as reflected in the discharge ratio, conditional on control factors, which represents our key identifying assumption.

Our measures of general deterrence allow the analysis to exploit variation in the sanctioning of other similar facilities across states and variation over time for each facility operating in a given state. This general variation is driven by regulatory priorities (e.g., greater emphasis on clean water) and the extent of compliance delivered by other facilities in the recent past. We assume that no one facility is sufficiently to drive regulatory priorities especially in the case of benefits from cleaner water. Similarly, we assume that the extent of compliance delivered by one facility in the recent past, which may be correlated with current compliance, does not influence meaningfully the extent of compliance delivered by other facilities in the recent past, i.e., no one facility's compliance is sufficiently prominent to shape the compliance decisions of other facilities. These conditions serve as additional identifying assumptions.

Conditional on the three noted identifying assumptions, our study identifies the causal link from enforcement certainty and severity to a facility's discharge ratio choice.

Lastly, sub-section 6.4 offers sensitivity analysis that assesses the robustness of the conclusions drawn from estimation of our base model.

5. Estimation Results

Table 2 reports our estimates of the effects of enforcement certainty and severity, including both the main and interactive coefficients for our base model, which uses Model Set A. Table 3 reports the fixed effects estimates for the control factors drawn from Model Set A.

Deterrence Factors: Individual Effects

We begin by interpreting the enforcement-related coefficients, as shown in Table 2. The general deterrence results reveal the following. First, the interactive term reveals that synergies exist between the certainty and severity of enforcement. Thus, inclusion of this interactive term is warranted given the strong statistical significance. The negative interactive term reveals that the two enforcement components are complements: as one enforcement component increases, the marginal effect of the other component becomes more negative (i.e., more effective at lowering the discharge ratio), which is fully consistent with the standard theoretical model of enforcement.

[Insert Tables 2 and 3 here]

Given the statistical significance of the interaction term, we must calculate each enforcement component's marginal effect based on a particular value for the other enforcement component. The main coefficients are only interpretable when the paired enforcement component equals zero, which is severely limited. In order to explore the marginal effect of each enforcement component conditioned on other values of the paired enforcement component, we construct Figure 1. For each enforcement component, Figure 1 displays the estimated marginal effects, along with the associated 90 % confidence intervals, across the full range of actual observed values for the other enforcement component. The x-axis in each panel uses a log scale.

[Insert Figure 1 here]

The marginal effect of general deterrence-related enforcement certainty is significantly positive, statistically zero, and significantly negative across roughly equal portions of the relevant range, as shown in Figure 1. Thus, increases in the certainty of enforcement based on other similar facilities' experiences are counter-productive, ineffective, or effective at inducing better compliance across the relevant range. In strong contrast, the marginal effect of general deterrence-related enforcement severity is significantly negative for most of the relevant range and only significantly positive at the very bottom of the range. Therefore, Figure 1 reveals that increases in the severity of enforcement based on other similar facilities' experiences are mostly effective at inducing better compliance across the relevant range.

For illustrative purposes only, Table 4 reports the statistics associated with the endpoints of the curves shown in Figure 1. Consistent with the complementary nature of the two enforcement components, the marginal effect of each enforcement component is positive when the other component is set at its minimum value, yet the marginal effect is negative when the other component is set at its maximum. A positive marginal effect demonstrates that enforcement is counter-productive. Both counter-productive marginal effects prove statistically significant. Therefore, regardless of the component of enforcement – certainty or severity, greater enforcement leads to higher discharge ratios, i.e., worse compliance, when the other factor is set at its minimum value. A negative marginal effect demonstrates that deterrence is effective at improving compliance. Both of the two effective marginal effects prove statistically significant. Thus, regardless of the component of enforcement, greater enforcement leads to lower discharge ratios, i.e., better compliance, when the other factor is set at its maximum value. Since we employ a (mixed) log-log specification, the estimated coefficients represent elasticities, which offer a natural interpretation. For example, a 1 % increase in the certainty of enforcement raises the discharge ratio by 0.19 % when enforcement severity takes its minimum value.

[Insert Table 4 here]

While Figure 1 and Table 4 display analysis that spans the entire parameter space, this analysis does not indicate how policy relevant these findings are. That is, this analysis does not demonstrate where the preponderance of the observations lies. A counter-productive effect, for example, is less important if it only occurs very infrequently in the data. To address this point, we separately calculate the certainty marginal effect and severity marginal effect for each of the 23,193 observations in the sample and then assess whether the calculated effect significantly differs from zero. Table 5 displays the results. As shown, the certainty effect is significantly counter-productive for 71 % of the observations, yet significantly productive for 12 % of the observations, and statistically zero for 17 % of the observations. In contrast, the severity effect is significantly productive for 31.1 % of the observations, significantly counter-productive for 68.8 % of the observations, and statistically zero for 0.1 % of the observations.

[Insert Table 5 here]

Overall, these results demonstrate that counter-productive enforcement is not just a theoretical possibility but also a reality in the examined enforcement regime. While one might hope that greater enforcement would never prove counter-productive, if not always effective, our empirical results demonstrate that, under particular conditions, greater enforcement appears to generate counter-productive effects. In particular, we conclude that, based on other facilities' experiences with enforcement, an increase in the certainty of enforcement reduces the chosen extent of compliance over a significant portion of the actual policy space.

Deterrence Factors: Comparison of Marginal Effects

We next compare the marginal effects of certainty and severity, which represents our study's primary objective. Throughout our comparison, we conclude that one marginal effect "dominates" the other marginal effect whenever the former proves the more effective deterrent, i.e., the marginal effect is more negative or at least less positive. Table 4 provides comparisons based on sample extremes by comparing the two marginal effects when each enforcement component is set at either its minimum or maximum. Again, we use these comparisons *only* to illustrate the full span of our sample. More importantly, Table 6 provides comparisons based on observation-specific marginal effects.

[Insert Table 6 here]

We first compare the general deterrence-based marginal effects at the sample extrema as shown in Table 4. As the final column of Table 4 shows, the severity effect significantly dominates the certainty effect when severity is set at a low level. When both certainty and severity are set at their minima, increasing severity is significantly *less* counter-productive than an equivalent increase in certainty. In contrast, if enforcement severity lies at an extremely high level, an increase in certainty proves significantly more effective than an increase in severity, regardless of the certainty value.

To assess the policy relevance of these comparisons, we evaluate whether the certainty and severity marginal effects calculated for each observation in the sample are significantly different, as shown in Table 6. Both marginal effects are counter-productive for nearly 70 % of the observations (Region 1). In all of these cases, the severity effect significantly dominates since its magnitude is less counter-productive. In contrast, both marginal effects are productive in 26 % of the observations (Regions 4 and 5). However, in only 30 % of these cases is the difference significant, in which case the certainty effect dominates the severity effect (Region 4). For 5 % of observations, the severity effect dominates the certainty effect since the former is productive, while the latter is counter-productive (Region 6), which proves statistically significant in 95 % of the cases. In sum, this assessment of general deterrence reveals that the severity effect significantly dominates the certainty effect in many more cases than the reverse. Yet when both effects are productive, the certainty effect significantly dominates the severity effect in more cases than the reverse.

As a final means of comparing the marginal effects of certainty and severity, Figure 2 graphically displays the comparison of the two marginal effects across the range of values for the case of general deterrence. In Figure 2, the horizontal axis captures the level of the certainty component, while the vertical axis captures the level of the severity component. Three lines divide the graph into distinctive regions. The (vertical) line of zero elasticity with respect to severity divides the graph into two regions: to the left of this line, the severity effect is counter-productive; to the right of this line, the severity effect is productive. Similarly, the (horizontal) line of zero elasticity with respect to certainty divides the graph into two regions: below this line, the certainty effect is counter-productive; above this line, the certainty effect is productive. The (positively sloped) line of equal elasticities divides the graph into two regions: left/above this line, the certainty effect dominates the severity effect; right/below this line, the severity effect dominates the certainty effect. Drawing upon all three lines, the graph divides into six distinctive regions based on the productive/counter-productive aspect of each marginal effect and the comparison of the two effects, as shown in Table 6. None of the described lines consider statistical significance; they rely exclusively upon the magnitudes of the marginal effects; statistical significance is assessed in Table 6. Finally, Figure 2 overlays data on the sample distribution of each pairing of certainty level and severity level, with each pairing shown as a diamond. By overlaying these data, we are able to assess whether any given region proves relevant in the sample. Nevertheless, the graph merely displays the distribution of the sample data; they do <u>not</u> show the frequency of observations within each region.[23]

[Insert Figure 2 here]

As shown in Figure 2, the marginal effect of severity is only counter-productive when the certainty value equals zero. In contrast, the marginal effect of certainty is counter-productive for several cases within the sample. In such cases, the severity effect dominates the certainty effect.

In all other cases, both marginal effects prove productive. In several cases, the severity effect dominates, while in many other cases, the certainty effect dominates. This graphical analysis helps to display the conclusions supported by Table 6 and to generalize the conclusions supported by Table 5.[24]

Sensitivity Analysis

Lastly, we offer sensitivity analysis to assess the robustness of our base model results. In this sub-section, we assess robustness by employing various alternative estimation strategies. First, we add EPA region-by-year fixed effects to the regressor set (Alternative Model 1). Second, we divide the count of sanctions by the count of violations over the preceding 12 months rather than the count of active months (Alternative Model 2).[25] We report these alternative estimates in Table 2 in order to facilitate comparison with our base model estimates. Third, we construct general deterrence measures based on two alternative time periods: (1) preceding 6 months, and (2) preceding 24 months. We report these alternative estimates in Table 7. Fourth, we use alternative model sets of control factors: Model Sets B, C, and D. Appendix Table D1 displays these alternative estimates. Fifth, we employ a variety of robustness checks: cluster the standard errors on the facility (Alternative Model 3), remove from the sample all facilities that exit the NPDES system during the sample period (Alternative Model 4),[26] exclude the interaction between sanction count and sanction conditional average value (Alternative Model 5), include the enforcement-related specific deterrence measures as control factors (Alternative Model 6), use general deterrence measures based on EPA regional boundaries rather than state boundaries (Alternative Model 7), exclude facilities affected the EPA Priority Sector program (Alternative Model 8), implement an inverse hyperbolic sine transformation of the dependent variable and the primary regressors (Alternative Model 9), remove the inspection-related regressor (Alternative Model 10), and modify the replacement values of zero in the dependent variable and general deterrence measures (Alternative Model 11). For the last model, we reduce the replacement values by 50 % (Alternative Model 11a) and increase the replacement values by 50 % (Alternative Model 11b). Appendix Table D2 displays the results of these nine alternative strategies. Use of the inverse hyperbolic sine transformation generates main and interactive coefficients for the general deterrence measures that are not comparable to the base model results. Thus, we re-evaluate the marginal effects of the individual general deterrence enforcement components and re-assess the comparison of these marginal effects; Appendix Table D3 displays this evaluation and comparison based on sample minima and maxima. Use of observation-specific values of the enforcement components supports similar conclusions.

We assess the alternative estimates in turn. Except in the cases of excluding the interaction between enforcement certainty and severity and inverse hyperbolic sine transformation, assessment of the coefficient magnitudes and p-values is sufficient. If these numbers are sufficiently similar, we would draw conclusions identical those reported above if we were to re-assess marginal effects. We assess the exceptional cases separately.

First, the addition of EPA region-by-year fixed effects (Alternative Model 1) generates highly similar coefficient magnitudes and p-values. Second, division by the count of sanctions (Alternative Model 2) generates highly similar p-values; the coefficient magnitudes differ because the denominator of the sanction certainty measure differs from the base model. Third, use of preceding 24-month general deterrence measures generates strongly significant and bigger coefficients, yet use of preceding 6-month general deterrence measures does not generate statistically significant coefficients. This pair of results seems to reveal that facilities need more than six months to learn of sanctions against other facilities and react to this information and, given more time to respond, facilities react more strongly. Fourth, Model Sets B, C, and D generate highly similar results. Fifth, clustering the standard errors on the facility reveals robustness to the choice of clustering approach (Alternative Model 3), removal of facilities that exit the NPDES system generates estimates nearly identical to the base model results (Alternative Model 4), inclusion of the enforcement-related specific deterrence measures as control factors leads to nearly identical results (Alternative Model 6), use of EPA region-based general deterrence measures generates results that support the same conclusions even though the level of statistical significance is slightly lower (Alternative Model 7), excluding facilities affected the EPA Priority Sector program generates results nearly identical to our base model estimates (Alternative Model 8), removal of the inspection-related regressor generates nearly identical results (Alternative Model 9), and the use of alternative replacement values for zeros in the log transformation generates highly similar results supporting identical conclusions (Alternative Model 11).[27]

Lastly, we assess the exceptional cases. Alternative Model 5 excludes the interaction term between sanction count (enforcement certainty) and sanction conditional average value (enforcement severity). As shown in Appendix Table D2, neither enforcement certainty nor severity proves statistically significant. Thus, the average marginal effect does not appear to statistically differ from zero. However, this alternative model is mis-specified because it excludes the highly significant interaction term. The base model offers a fully nuanced assessment of the enforcement marginal effects across the full range of values and those present in the sample.

Alternative Model 9 implements an inverse hyperbolic sine transformation of the dependent variable and the primary regressors. As shown in Appendix Table D2, two of three primary regressor coefficients prove statistically significant. The two main coefficients reflect the marginal effect when the other enforcement component equals zero. Based on the main enforcement certainty coefficient, the marginal effect of greater certainty is significantly counter-productive when severity equals zero. Based on the main enforcement severity coefficient, the marginal effect of greater severity is insignificantly positive at a very small level (β=0.0001) when certainty equals zero. More important, Appendix Table D3 assesses and compares the marginal effects based on sample minima and maxima. In general, the marginal effects support conclusions identical to those supported by our base model results. As the single exception, the marginal effect of severity never proves significantly counter-productive.

6. Conclusions

Our study explores the relative efficacy of increasing the certainty of enforcement versus the severity of enforcement in prompting better compliance with wastewater discharge limits. The findings of our exploration contribute to an ongoing debate over this comparison since theory and previous empirical findings from criminology and experiments are mixed. We offer a novel contribution by examining the context of regulatory compliance and allowing the two enforcement effects to differ with the size of the other enforcement component. Our empirical results reveal that greater enforcement can prove a significant deterrent for U.S. chemical manufacturing facilities permitted within the Clean Water Act's National Pollutant Discharge Elimination System (NPDES). However, the magnitudes of the certainty and severity effects vary according to the perceived enforcement threat. When enforcement certainty and severity are high, both components are effective at deterring pollution, although certainty is more effective. This set of results is consistent with previous results from criminology (e.g., Grogger, 1991; Eide, 2000; Nagin, 2013). In contrast, for low values of enforcement certainty and severity, both components are counter-productive, i.e., lead to greater pollution, with the counter-productive effect of certainty larger in magnitude. For intermediate values, the results are more mixed.

Our empirical results possess meaningful policy implications especially since enforcement agencies should place greater emphasis on the more effective lever – certainty of enforcement or severity of enforcement. First, enforcement agencies should be cautious when increasing the frequency of enforcement actions with low severity as the general deterrent effects of such actions are likely to be counter-productive. On the other hand, increasing severity can be an effective general deterrent but might prove counter-productive if such large punishments are infrequently imposed. In contrast, when enforcement certainty and severity are high, both instruments are effective in deterring pollution, although certainty is more effective. More broadly, environmental protection agencies should not rely upon a single component of enforcement – certainty or severity. Instead, environmental agencies should employ both levers meaningfully so that both tools prove effective. Nevertheless, the agencies should emphasize use of the certainty of enforcement over the severity of enforcement when both levers are meaningfully employed.

Our results also demonstrate the importance of including an interaction term between the two components of deterrence: certainty of enforcement and severity of enforcement. This interaction term reveals that certainty and severity are complements for general deterrence, as well as specific deterrence. Future research should explore whether or not these interactive effects generalize to other environmental protection and regulatory settings.

Our research focuses exclusively on the benefits of greater enforcement, while ignoring the costs. We encourage future research to exploit our estimates on the relative effectiveness of enforcement components – certainty and severity – to conduct a benefit-cost analysis.

While our empirical analysis uncovers policy-relevant ranges of counter-productive enforcement, our analysis is not designed to uncover the mechanism. One possible mechanism is that the small monetary incentives associated with low severity might crowd out other motivations of individual environmental managers, such as an intrinsic desire to obey the law or cooperate with regulators. Such an interpretation is consistent with the substantial behavioral economics literature (Frey and Jegen, 2001; Fehr and Falk, 2002; Gneezy et al., 2011; Bowles and Polania-Reyes, 2012). While one might believe that behavioral motives should not influence corporate entities, several studies demonstrate otherwise. For example, Nakamura et al. (2001) find that managerial attitudes towards environmental protection influence facilities' management choices, both Winter and May (2001) and Earnhart and Glicksman (2015) find that coercive (rather than cooperative) enforcement styles reduce compliance, and Short and Toffel (2010) demonstrate that explicit threats by regulators undermine compliance. Future research should also examine further these motives in the context of regulatory compliance.

Lastly, we acknowledge that we examine only one sample that comprises major facilities operating in a single sector (chemical manufacturing), single medium (surface water), and a single pollutant (total suspended solids) during a specific period. Thus, our results need not generalize to other sectors, other media, and/or other pollutants, as well as more recent periods and smaller facilities. Future research should explore additional sectors (e.g., oil and gas extraction), additional media (e.g., air pollutants), and additional pollutants (e.g., particulate matter). One might expect that more recent data might reveal even stronger general deterrence impacts given increased information access and dissemination via online media.

Acknowledgments: We thank Dylan Rassier and Trisha Shrum for their excellent research assistance. We also thank the University of Queensland for their financial support of our collaboration.

Disclaimer: This manuscript was partially developed under a STAR Research Assistance Agreement No. R-82882801-0 awarded by the U.S. Environmental Protection Agency. It has not been formally reviewed by the EPA. The views expressed in this document are solely those of the authors. The EPA does not endorse any products or commercial services mentioned in this manuscript.

References

Anderson, Lisa, and Sarah Stafford. 2003. "Punishment in a Regulatory Setting: Experimental Evidence from the VCM." *Journal of Regulatory Economics* 24 (1): 91-110.

Angrist, Joshua, and Jörn-Steffen Pischke. 2009. *Mostly Harmless Econometrics*. Princeton University Press: Princeton, NJ.

Arora, Seema, and Timothy Cason. 1996. “Why Do Firms Volunteer to Exceed Environmental Regulations? Understanding Participation in EPA’s 33/50 Program.” *Land Economics* 72 (4): 413-432.

Bandyopadhyay, Sushenjit, and John Horowitz. 2006. “Do Plants Overcomply with Water Pollution Regulations? The Role of Discharge Variability." *B.E. Journal of Economic Analysis and Policy: Topics in Economic Analysis and Policy* 6(1): 1-32.

Baum, Christopher, Mark Schaffer, and S. Stillman. 2003. “Instrumental Variables and GMM: Estimation and Testing.” *Stata Journal* 3: 1-31.

Becker, Gary. 1968. "Crime and Punishment: An Economic Approach." *Journal of Political Economy* 76: 169-217.

Bellemare, Marc, and Casey Wichman. 2020. "Elasticities and the Inverse Hyperbolic Sine Transformation." *Oxford Bulletin of Economics and Statistics* 82 (1): 50-61.

Bowles, Samuel, and Sandra Polania-Reyes. 2012. "Economic Incentives and Social Preferences: Substitutes or Complements?." *Journal of Economic Literature* 50: 368-425.

Cohen, Mark. 2000. "Empirical Research on the Deterrent Effect of Environmental Monitoring and Enforcement." *Environmental Law Reporter* 30: 10245-10252.

Dasgupta, Susmita, Hemamala Hettige, and David Wheeler. 2000. "What Improves Environmental Compliance? Evidence from Mexican Industry." *Journal of Environmental Economics and Management* 39: 39-66.

Downing, P., and J. Kimball. 1982. "Enforcing Pollution Control Laws in the United States." *Policy Studies Journal* 11 (1): 15-22.

Earnhart, Dietrich. 2004a. "Panel Data Analysis of Regulatory Factors Shaping Environmental Performance." *Review of Economics and Statistics* 86 (1): 391-401.

Earnhart, Dietrich. 2004b. "Regulatory Factors Shaping Environmental Performance at Publicly-Owned Treatment Plants." *Journal of Environmental Economics and Management* 48 (1): 655-681.

Earnhart, Dietrich. 2004c. "The Effects of Community Characteristics on Polluter Compliance Levels." *Land Economics* 80 (3): 408-432.

Earnhart, Dietrich. 2009. "The Influence of Facility Characteristics and Permit Conditions on the Effects of Environmental Regulatory Deterrence." *Journal of Regulatory Economics* 36 (3): 247-273.

Earnhart, Dietrich, and Lana Friesen. 2013. “Can Punishment Generate Specific Deterrence without Updating? Analysis of a Stated Choice Scenario.” *Environmental and Resource Economics* 56 (3): 379-397.

Earnhart, Dietrich, and Lana Friesen. 2017. “The Role of Regulated Facilities’ Perceptions of the Effectiveness of Regulatory Monitoring and Enforcement: Compliance with the Clean Water Act.” *Ecological Economics* 142: 282-294.

Earnhart, Dietrich, and Lana Friesen. 2021a. “Enforcement Federalism: Comparing the Effectiveness of Federal Punishment versus State Punishment.” *Environmental and Resource Economics* 78: 227-255.

Earnhart, Dietrich, and Lana Friesen. 2021b. “Use of Competitive Endogenous Audit Mechanisms by Federal and State Inspectors within Environmental Protection Agencies.” *Journal of Environmental Economics and Management*, 109: 102476.

Earnhart, Dietrich, and Robert Glicksman. 2011. *Pollution Limits and Polluters’ Efforts to Comply: The Role of Government Monitoring and Enforcement*. Palo Alto, CA: Stanford University Press.

Earnhart, Dietrich, and Robert Glicksman. 2015. “Extent of Cooperative Enforcement: Effect of the Regulator-Regulated Facility Relationship on Audit Frequency.” *Strategic Behavior and the Environment* 5 (2): 111-156.

Earnhart, Dietrich, and Kathleen Segerson. 2012. “The Influence of Financial Status on the Effectiveness of Environmental Enforcement.” *Journal of Public Economics* 96: 670-684.

Eckert, Heather. 2004. “Inspections, Warnings, and Compliance: The Case of Petroleum Storage Regulation.” *Journal of Environmental Economics and Management* 47 (2): 232-259.

Eide, Erling. 2000. "Economics of Criminal Behavior." In *Encyclopedia of Law and Economics*. Boudewijn Bouckaert and Gerrit De Geest (eds), Cheltenham: Edward Elgar.

Eide, Erling, P.H. Rubin, and J.M. Shepherd. 2006. "Economics of Crime." *Foundations and Trends in Microeconomics* 2: 205-279.

EPA. 1990. "A Primer on the Office of Water Enforcement and Permits and Its Programs." Office of Water, Environmental Protection Agency, Washington, DC, March 1990.

EPA. 1997. *Chemical Industry National Environmental Baseline Report 1990-1994*. Office of Enforcement and Compliance Assurance, Environmental Protection Agency, EPA 305-R-96-002, October 1997.

EPA. 1999a. *EPA/CMA Root Cause Analysis Pilot Project: An Industry Survey*. Environmental Protection Agency, EPA-305-R-99-001, May 1999.

EPA. 1999b. "Major Findings from the CEIS Review of EPA's Permit Compliance System Database." EPA Center for Environmental Information and Statistics, Washington, DC.

EPA. 2011. *Technical Support Document for the 2010 Effluent Guidelines Program Plan*. Environmental Protection Agency, EPA 820-R-10-021.

Fehr, E., and A. Falk. 2002. "Psychological Foundations of Incentives." *European Economic Review* 46: 687-724.

Foulon, Jérôme, Paul Lanoie, and Benôit Laplante. 2002. "Incentives for Pollution Control: Regulation or Information." *Journal of Environmental Economics and Management* 44 (1): 169-187.

Frey, B.S., and R. Jegen. 2001. "Motivation Crowding Theory." *Journal of Economic Surveys* 15: 589-611.

Friesen, Lana. 2012. "Certainty of Punishment versus Severity of Punishment: An Experimental

Investigation." *Southern Economic Journal* 79 (2): 399-421.

GAO. 1996. *Water Pollution: Many Violations Have Not Received Appropriate Enforcement Attention*. General Accounting Office, Report # GAO/RCED-96-23, Washington, DC.

Gneezy, U., S. Meier, and P. Rey-Biel. 2011. "When and Why Incentive (Don't) Work to Modify Behavior." *Journal of Economic Perspectives* 25 (4): 191-210.

Gray, Wayne, and Mary Deily. 1996. "Compliance and Enforcement: Air Pollution Regulation in the U.S. Steel Industry." *Journal of Environmental Economics and Management* 31: 96-111.

Gray, Wayne, and John Scholz. 1991. "Analysing the Equity and Efficiency of OSHA Enforcement." *Law & Policy* 13: 185-214.

Gray, Wayne, and John Scholz. 1993. "Does Regulatory Enforcement Work? A Panel Analysis of OSHA Enforcement." *Law & Society Review* 27: 177-213.

Gray, Wayne, and Ronald Shadbegian. 2004. "'Optimal' Pollution Abatement - Whose Benefits Matter, and How Much?." *Journal of Environmental Economics and Management* 47: 510-534.

Gray, Wayne, and Ronald Shadbegian. 2005. "When and Why do Plants Comply? Paper Mills in the 1980s." *Law and Policy* 27 (2): 238-261.

Gray, Wayne, and Ronald Shadbegian. 2007. "The Environmental Performance of Polluting Plants: A Spatial Analysis." *Journal of Regional Science* 47 (1): 63-84.

Gray, Wayne, and Jay Shimshack. 2011. "The Effectiveness of Environmental Monitoring and Enforcement: A Review of the Empirical Evidence." *Review of Environmental Economics and Policy* 5: 3-24.

Grogger, J. 1991. "Certainty vs. Severity of Punishment." *Economic Inquiry* 29 (2): 297-309.

Hamilton, James. 1993. “Politics and Social Costs: Estimating the Impact of Collective Action on Hazardous Waste Facilities.” *RAND Journal of Economics* 24 (1): 101-125.

Helland, Eric. 1998a. “The Enforcement of Pollution Control Laws: Inspections, Violations and Self-Reporting.” *Review of Economics and Statistics* 80 (1): 141-153.

Helland, Eric. 1998b. “The Revealed Preferences of State EPAs: Stringency, Enforcement, and Substitution.” *Journal of Environmental Economics and Management* 35 (3): 242-261.

Helland, Eric, and Andrew Whitford. 2003. “Pollution Incidence and Political Jurisdiction: Evidence from the TRI." *Journal of Environmental Economics and Management* 46: 403-424.

Henriques, Irene, and Perry Sadorsky. 1996. “The Determinants of an Environmentally Responsive Firm: An Empirical Approach.” *Journal of Environmental Economics and Management* 30: 381-395.

Jacobson, Sarah. 2015. “Deterrence in the Public and Private Sectors.” Paper presented at the University of Colorado Environmental and Resource Economics workshop, Vail, Colorado, September 12, 2015.

Laplante, Benoît, and Paul Rilstone. 1996. “Environmental Inspections and Emissions of the Pulp and Paper Industry in Quebec.” *Journal of Environmental Economics and Management* 31 (1): 19-36.

Magat, Wesley A., and W. Kip Viscusi. 1990. “Effectiveness of the EPA’s Regulatory Enforcement: The Case of Industrial Effluent Standards.” *Journal of Law and Economics* 33 (2): 331-360.

Nadeau, Louis. 1997. “EPA Effectiveness at Reducing the Duration of Plant-Level Noncompliance.” *Journal of Environmental Economics and Management* 34 (1): 54-78.

Nagin, Daniel. 2013. “Deterrence: A Review of the Evidence by a Criminologist for Economists.” *Annual Review of Economics* 5(1): 83-105.

Nakamura, Masao, Takuya Takahashi, and Ilan Vertinsky. 2001. “Why Japanese Firms Choose to Certify: A Study of Managerial Responses to Environmental Issues." *Journal of Environmental Economics and Management* 42 (1): 23-52.

Polinsky, A. Mitchell, and Steven Shavell. 2000. “The Economic Theory of Public Enforcement of Law." *Journal of Economic Literature* 38: 45-76.

Rousseau, Sandra. 2009. “The Use of Warnings in the Presence of Errors.” *International Review of Law and Economics* 29: 191-201.

Sah, Raj. 1991. "Social Osmosis and Patterns of Crime." *Journal of Political Economy* 99: 1272-1295.

Scholz, John, and Wayne Gray. 1990. "OSHA Enforcement and Workplace Injuries: A Behavioral Approach to Risk Assessment." *Journal of Risk and Uncertainty* 3: 283-305.

Shimshack, Jay, and Michael Ward. 2005. "Regulator Reputation, Enforcement, and Environmental Compliance." *Journal of Environmental Economics and Management* 50: 519-540.

Shimshack, Jay, and Michael Ward. 2008. "Enforcement and Over-Compliance." *Journal of Environmental Economics and Management* 55: 90-105.

Short, Jodi, and Michael Toffel. 2010. "Making Self-Regulation More Than Merely Symbolic: The Critical Role of the Legal Environment." *Administrative Science Quarterly* 55: 361-396.

Sigman, Hilary. 1998. “Midnight Dumping: Public Policies and Illegal Disposal of Used Oil.” *RAND Journal of Economics* 29: 157–178.

Stock, J., and M. Yogo. 2005. “Testing for Weak Instruments in Linear IV Regression.” In *Identification and Inference for Econometric Models: Essays in Honor of Thomas Rothenberg*. eds. D.W.K. Andrews and J.H. Stock, 80-108, New York: Cambridge University Press.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment under Uncertainty: Heuristics and Biases." *Science* 185: 1124-1131.

Wasserman, Cheryl. 1984. “Improving the Efficiency and Effectiveness of Compliance Monitoring and Enforcement of Environmental Policies, United States: A National Review,” OECD.

Winter, S., and Peter May. 2001. "Motivation for Compliance with Environmental Regulations." *Journal of Policy Analysis and Management* 20: 675-698.

Wooldridge, Jeffrey. 2002. *Econometric Analysis of Cross Section and Panel Data*. MIT Press: Cambridge, MA.

Table 1
Summary Statistics (N=23,193)

Variable	Mean	Std Dev	Min	Max
Environmental Performance Measures				
TSS Discharge Ratio	0.315	0.356	0.001	9.080
TSS Discharge Ratio (logs)	-1.691	1.551	-11.364	2.206
General Deterrence Factors				
General Deterrence - Enforcement Certainty: Sanction Count per active month over 12-month preceding period against *other* facilities (#/facility-month)	0.0010	0.0030	0	0.0313
General Deterrence - Enforcement Severity: Conditional Average Sanction Magnitude over 12-month preceding period against *other* facilities ($/sanction)	228,750	837,659	0	8,225,931
Control Factors				
Specific Deterrence - Enforcement Certainty: Sanction Count per active month over 12-month preceding period against facility (#/facility-month)	0.0020	0.0148	0	0.3333
Specific Deterrence - Enforcement Severity: Conditional Average Sanction Magnitude over 12-month preceding period against facility ($/sanction)	8,133.64	226,821	0	8,225,931
Inspections per active month over 12-month preceding period at *other* facilities (#/facility-month)	0.1242	0.1052	0	0.8438
Inspections per active month over 12-month preceding period at own facility (#/facility-month)	0.1292	0.1614	0	3
Year 1997 (1,0) [1]	0.1845	0.3879	0	1
Year 1998 (1,0) [1]	0.1856	0.3888	0	1
Year 1999 (1,0) [1]	0.1796	0.3838	0	1
Year 2000 (1,0) [1]	0.1782	0.3827	0	1
Year 2001 (1,0) [1]	0.0870	0.2824	0	1
Winter Season (1,0) [2]	0.2576	0.4373	0	1
Spring Season (1,0) [2]	0.2726	0.4453	0	1
Summer Season (1,0) [2]	0.2425	0.4286	0	1
Monthly Effluent Limit (000s lbs/day)	1.2742	4.0362	0	50
Initial or Interim Limit Type (1,0) [3]	0.0160	0.1264	0	1
Modification to Permit (1,0) [4]	0.0870	0.2815	0	1
State and Local Budget / # of Manufacturers ($ per) [5]	0.0500	0.0320	0.0090	0.1868
EPA Regional Budget / # of Manufacturers ($ per) [6]	0.6727	0.1483	0.4738	1.2293
Flow Capacity (million gallons / day) [7]	2.1237	3.2437	0	26.29
Flow to Flow Capacity (ratio) [7]	1.0399	1.7972	0	20.35
Standard Deviation of Discharge Ratio	0.1738	0.3745	0	17.05

Publicly Held Ownership [8]	0.6875	0.4635	0	1

[1] The omitted category is year 1996.
[2] The omitted category is autumn.
[3] The omitted category is final limit type.
[4] The omitted category is no modification to the permit.
[5] Data on state and local natural resource-related budgets are available only for the years 1995 to 1999; the study extrapolates these data to cover the years 2000 and 2001.
[6] EPA regional data exist only for the years 1998 to 2002; the study backward extrapolates these data to cover the years 1995 to 1997.
[7] When no monthly measurement of wastewater flow is available, the study imputes a replacement value based on the following hierarchy depending on data availability: (1) facility-specific annual average, (2) facility-specific sample average, (3) sample-wide average. This imputation affects less than 3 % of the sample. Sensitivity analysis reveals that exclusion of the observations with imputed wastewater flow values does not meaningfully alter the estimation results.
[8] The omitted category is non-publicly held ownership.

Table 2
Fixed Effects Estimation of TSS Discharge Ratio: Baseline and Alternative Specifications
Enforcement in Preceding 12-month Period – General Deterrence-related Coefficients
(N = 23,193)

ß = coefficient magnitude
p = coefficient p-value

Regressor	Base Model		Alt Model 1		Alt Model 2	
	ß	p	ß	p	ß	p
Main Effects						
Sanction Count	0.1888	0.002	0.1745	0.021	0.0483	0.048
Sanction Conditional Avg $	-0.1299	0.000	-0.1152	0.001	-0.0359	0.003
Interactive Effect						
Sanction Count ⊗ Sanction Conditional Avg $	-0.0199	0.000	-0.0179	0.001	-0.0100	0.001
Control Factor Inclusion						
Baseline Controls	X		X		X	
Region-by-Year Fixed Effects			X			
Violation Adjusted Sanction Count					X	

Notes:
Regressions include facility-specific indicators.
Standard errors are clustered on the state on which a facility operates.
Clustering the standard errors on the facility generates the following p-values in the baseline model for sanction count, conditional average sanction magnitude, and their interaction, respectively: 0.004, 000, and 0.000.

Table 3
Fixed Effects Estimation of TSS Discharge Ratio – Baseline Specification
Control Factor Coefficients: Magnitudes and p-values (N = 23,193)

Regressor	Coefficient	
	Magnitude	p-value
Inspections conducted at other facilities	-0.0127	0.741
Year 1997	-0.0730	0.121
Year 1998	-0.0845	0.192
Year 1999	-0.2054	0.038
Year 2000	-0.1305	0.136
Year 2001	-0.2367	0.081
Winter Season [1]	0.1391	0.000
Spring Season [1]	0.1208	0.000
Summer Season [1]	0.0481	0.028
Initial/Interim Limit Type [2]	-0.0105	0.947
Modification to Permit [3]	-0.0660	0.645
Monthly Effluent Limit	0.0173	0.762
State and Local Budget / # of Manufacturers	-1.8623	0.576
EPA Regional Budget / # of Manufacturers	0.4652	0.538
Intercept	-0.6385	0.224
Regression Elements		
F-Test: Zero Slopes	21.95	0.000
F-test: Fixed Effects	76.89	0.000

[1] The omitted category is autumn season.
[2] The omitted category is final limit type.
[3] The omitted category is the lack of modification to an NPDES wastewater permit.

Table 4

Marginal Effects of Individual General Deterrence Enforcement Components and Comparisons: Evaluated at Sample Minima and Maxima – Estimates based on Baseline Specification

(significantly negative effects shown in **bold**, significantly positive effects shown in *italics*)

ß = marginal effect magnitude
p = marginal effect p-value or test statistic p-value

Conditional Value		Individual Marginal Effects				Comparison Test		Conclusion
Certainty	Severity	Certainty		Severity		Statistic	p	
		ß	p	ß	p			
Min	Min	*0.1887*	*0.001*	*0.0075*	*0.030*	10.55	0.001	Certainty $\prec$ Severity
Min	Max	**-0.1281**	**0.001**	*0.0075*	*0.030*	11.10	0.001	Certainty $\succ$ Severity
Max	Min	*0.1887*	*0.001*	**-0.0616**	**0.000**	12.90	0.000	Certainty $\prec$ Severity
Max	Max	**-0.1281**	**0.001**	**-0.0616**	**0.000**	4.18	0.041	Certainty $\succ$ Severity

Table 5

Observation-specific Marginal Effects of General Deterrence Enforcement Certainty and Severity: Individual Marginal Effects

(percent of observations shown in each cell)

Enforcement Component	Sign	Significance	
		Significant [p≤0.10]	Insignificant [p>0.10]
Certainty	> 0	70.9	3.0
	< 0	12.0	14.0
Severity	> 0	68.8	0
	< 0	31.1	0.1

Table 6

Observation-specific Marginal Effects of General Deterrence Enforcement Certainty and Severity: Comparison of Marginal Effects
(count of observations with percent shown in parentheses)

Region	Certainty Effect Category	Severity Effect Category	Dominant Effect	Number of Overall Cases	Number of Cases by Significance Status (row percent)	
					Significant [p≤0.10]	Insignificant [p>0.10]
1	counter-productive	counter-productive	severity	15,965 (68.8)	15,965 (100.0)	0
2	counter-productive	counter-productive	certainty	0	0	0
3	productive	counter-productive	certainty	3 (0.001)	3 (100.0)	0
4	productive	productive	certainty	3,858 (16.6)	1,825 (47.3)	2,033 (52.7)
5	productive	productive	severity	2,181 (9.4)	0	2,181 (100.0)
6	counter-productive	productive	severity	1,193 (5.1)	1,128 (94.6)	65 (5.4)

Table 7

Fixed Effects Estimation of TSS Discharge Ratio:
Enforcement in Preceding 6-month Period (N=23,193) **or 24-month Period** (N=18,916)
General Deterrence-related Coefficients

Regressor	6-month period		24-month period	
	Coefficient	p-value	Coefficient	p-value
Main Effects				
Sanction Count	0.0761	0.109	0.5281	0.000
Sanction Conditional Avg $	-0.0288	0.365	-0.2738	0.000
Interactive Effect				
Sanction Count ⊗ Sanction Conditional Avg $	-0.0026	0.620	-0.0402	0.000
Control Factor Inclusion				
Baseline Controls	X		X	

Notes:
Regressions include facility-specific indicators.
Standard errors are clustered on the state on which a facility operates.

Figure Captions

Figure 1
Estimated Marginal Effects with respect to Certainty and Severity including 90% Confidence Intervals

Figure 2
General Deterrence: Elasticities with Respect to Certainty and Severity

Endnotes

[1] While less common in the environmental enforcement literature, the terminology of certainty versus severity is standard in the criminology literature (e.g., Grogger, 1991; Nagin, 2013).

[2] While several empirical studies of regulatory compliance separately estimate the effects of certainty and severity (Scholz and Gray, 1990; Sigman, 1998; Shimshack and Ward, 2005), none formally compares the two effects.

[3] In the standard deterrence model of Becker (1968), increases in the severity of monetary-based punishment is costless. Consequently, Becker (1968) demonstrates that the optimal level of severity is the legal maximum. Of course, once increasing enforcement severity is costly, a comparison of the two levers' effectiveness is relevant for any benefit-cost analysis.

[4] This dependence is shown in footnote 16 of Becker (1968), as well as in equations (3) and (4) of Friesen (2012), who adapts the Becker model to regulatory compliance.

[5] Given the fundamental nature of our research endeavor, the choice of sample period is not critical. We merely wish to avoid a time period in which special circumstances apply broadly, such as a sector-wide policy initiative from the EPA. Our chosen sample period is sufficient since only two small sub-sectors, industrial organics and chemical preparations, faced an EPA policy initiative for only a portion of the sample period.

[6] When establishing discharge limits, agencies must also consider sector-specific Effluent Limitation Guidelines, which apply uniformly to all facilities. Agencies do <u>not</u> base limits on facility characteristics, past discharge patterns, or expected future discharge patterns (Earnhart and Glicksman, 2011). Consequently, limits are not endogenously determined with respect to the extent of compliance, i.e., discharge ratio.

[7] Appendix A explores the potential problems of non-reporting and strategic misreporting.

[8] This pollutant represents one of the five EPA conventional pollutants, which are the focus of EPA efforts. Several previous studies of wastewater discharges examine TSS (e.g., Earnhart, 2009; Laplante and Rilstone, 1996; Earnhart and Segerson, 2012). We examine only TSS in order to explore the broadest sample. Limiting the sample to those facilities that discharge TSS and other pollutants, such as biological oxygen demand (BOD), reduces the sample size. This restriction reduces the generalizability of our results since those facilities that discharge both TSS and some other pollutant may differ from the broadest sample of facilities that discharge at least TSS.

[9] We focus on federal enforcement actions for two reasons. First, the EPA did not and still does not offer a central database with systematically recorded data on state-issued enforcement actions. Acquisition of complete data on state enforcement requires exploitation of state-specific databases that vary substantially in their structure and recording procedures. See Jacobson (2015) for a thorough assessment of several states' enforcement data. Second, federal sanctions clearly impose financial penalties on violating facilities. In contrast, most state enforcement actions do not impose financial penalties and, when imposed, the state penalties are much smaller than federal penalties. Earnhart and Segerson (2012) make these same two points. Earnhart and Friesen (2021a) compare the effectiveness of federal and statement enforcement in four key states.

[10] The EPA offers no comprehensive data on minor facilities.

[11] The sample includes facilities operating in all 10 EPA regions except EPA Region 8.

[12] In this context, over-compliance represents a regulated facility's *ex ante* decision to lower the likelihood of being non-compliant *ex post*.

[13] Additional regression analysis includes local community characteristics (e.g., per capita income) as regressors to control for local community pressure. Regression results demonstrate that these characteristics do not prove statistically significant (F-test statistic=1.14, p=0.338) and their inclusion does not disrupt the general pattern of results reported in Table 2. These results are available upon request from the authors.

[14] Consistent with many studies examining general deterrence, facilities learn of others' experiences through various channels. As the most prominent channel, the EPA and state agencies regularly and intentionally announce large or important sanctions.

[15] We highlight that the general deterrence regressors are not potentially endogenous as they measure interventions imposed against <u>other</u> facilities, which should depend on <u>other</u> facilities' compliance. Thus, general deterrence measures are not vulnerable to any concern over the targeting of individual facilities. Nevertheless, we acknowledge that an individual facility's discharge decision may depend on sector-level unobservable factors that may also influence an agency's enforcement decisions.

[16] We are only aware of one study that examines general deterrence using measures derived from a more local set of facilities. Gray and Shadbegian (2007) consider inspections conducted at other manufacturing plants within 10 miles of a plant, finding general deterrence effects only from inspections conducted within the same state and not without. Earnhart and Friesen (2021b) demonstrate that EPA regional offices apparently focus on individual states when designing CWA monitoring/enforcement strategies.

[17] Earnhart and Glicksman (2015) discuss the difference between a coercive enforcement approach, which seeks to deter non-compliance, and a cooperative enforcement approach, which seeks to facilitate compliance.

[18] These same studies treat lagged sanctions as an exogenous regressor.

[19] To address the presence of zero values, we replace zeroes with the midpoint between 0 and the smallest positive value (0.000012). As part of our sensitivity analysis, we explore alternative replacement values.

[20] To address the presence of zero values in our enforcement measures, we use a very small value, relative to the sample distribution, to serve as a reasonable approximation of some minimal threat of enforcement or inspection. (The specific values for enforcement certainty, enforcement severity, and inspections are, respectively, 0.01, 1, and 0.01.) In reality, a threat is always present even when no government interventions were recently conducted. Our adjustment avoids the error of ignoring this reality. As part of our sensitivity analysis, we explore alternative values for this adjustment.

[21] Bellemare and Wichman (2020) recommend a more exact formulation of elasticities when the transformed variables include zero-value observations; see page 54. Implementation of this recommended formulation proves complicated in our interactive specification. Thus, we focus on the approximate formulation.

[22] This alternative clustering approach helps to strengthen our conclusions since the sample includes 40 states. Thus, the number of clusters in the primary clustering approach is 40, which lies slightly below the rule of thumb of 42 when implementing standard clustering (Angrist and Pischke, 2009).

[23] To display the lines and data pairings effectively, Figure 2 curtails the vertical and horizontal ranges of the graph; in the process, a very few data pairings are not displayed.

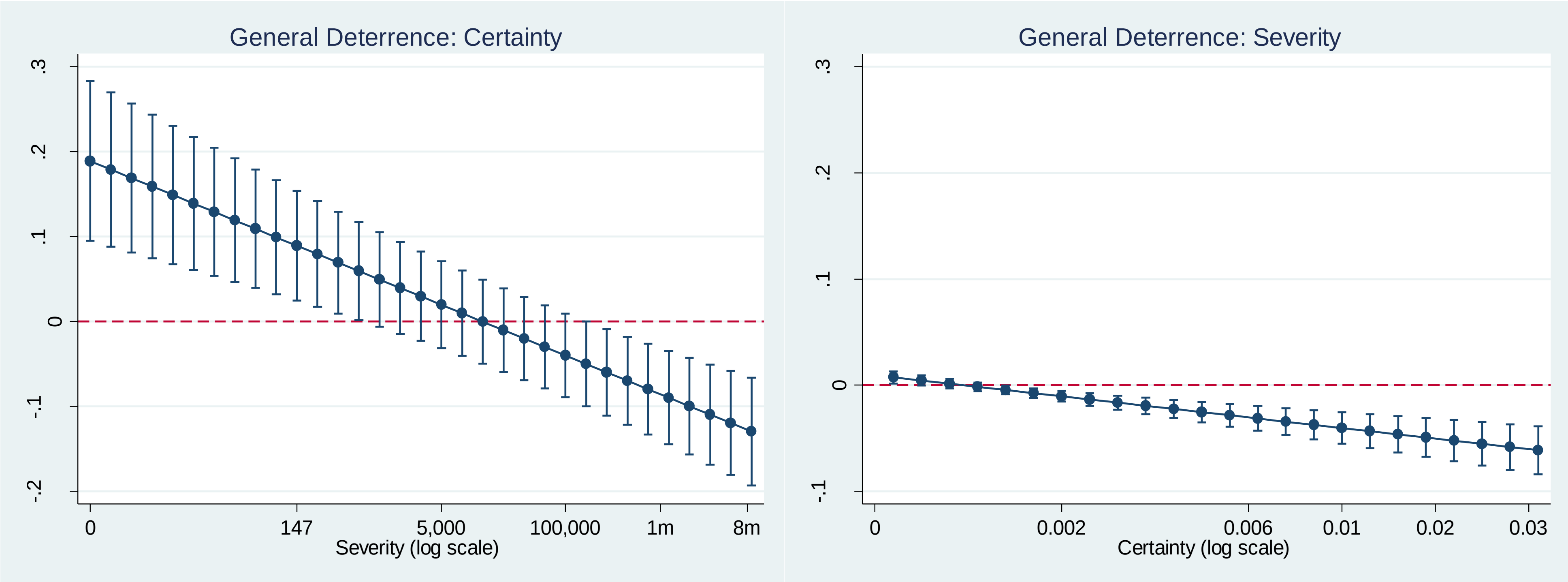
[24] We briefly interpret the coefficients relating to the control factors shown in Table 3: (1) discharge ratios vary over the sample period from year to year, (2) discharge ratios vary over the calendar year from season to season.

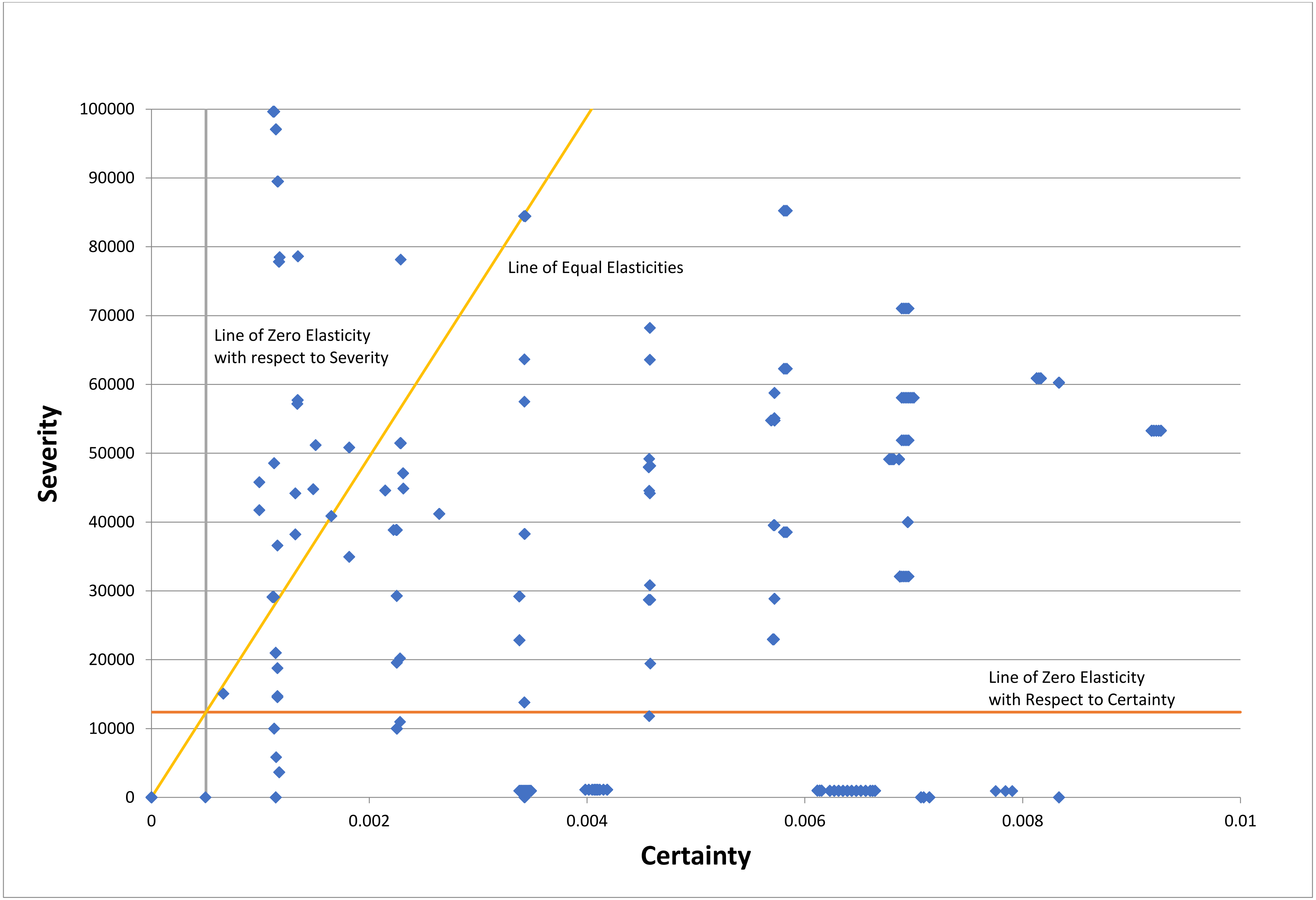
[25] This alternative approach avoids division by zero by adding one to each violation count. The base model does not face this issue since each state has active facilities throughout the sample period.

[26] Of the 406 facilities included in our sample, nine exit the system, representing 2 % of overall facilities.

[27] In the case of Alternative Model 11, we re-assess the marginal effects shown in Table 4. Use of the

Alternative Model 11 coefficients generate highly similar marginal effects that support identical conclusions.


General Deterrence: Certainty
.3
.2
.1
0
-.1
-.2
0
147
5,000
100,000
1m
8m
Severity (log scale)
General Deterrence: Severity
.3
.2
.1
0
-.1
0
0.002
0.006
0.01
0.02
0.03
Certainty (log scale)


Severity
Certainty
Line of Equal Elasticities
Line of Zero Elasticity
with respect to Severity
Line of Zero Elasticity
with Respect to Certainty
100000
90000
80000
70000
60000
50000
40000
30000
20000
10000
0
0
0.002
0.004
0.006
0.008
0.01