Sam Houston State University
Department of Economics and International Business
Working Paper Series

Firm Decision-making Structure and Compliance with Environmental Regulations: Evidence from Environmental Auditing

Mary F. Evans
Claremont McKenna College

Lirong Liu
Sam Houston State University

Sarah Stafford
College of William and Mary

SHSU Economics & Intl. Business Working Paper No. 12-06
September 2012

Abstract:

We explore how limits to our insight on the underlying decision-making structure of regulated entities may affect the conclusions we draw about the likely impacts of environmental policies, with a focus on environmental auditing. We examine the conditions under which a multi-facility firm chooses to adopt a standardized auditing. Our firm-level empirical analysis confirms that a uniform auditing outcome is less likely among firms (1) with a more heterogeneous portfolio of facilities, and (2) that are less likely to make decision errors. We examine the implications of adding controls for firm decision-making structure in a facility-level empirical analysis of auditing's impact on compliance. Our results suggest that the estimated compliance impact of auditing varies depending on whether or not we include these controls.

SHSU ECONOMICS WORKING PAPER

Firm Decision-making Structure and Compliance with Environmental Regulations: Evidence from Environmental Auditing

July 18, 2012

Mary F. Evans[**], Lirong Liu[*], and Sarah Stafford[†]

Abstract:
We explore how limits to our insight on the underlying decision-making structure of regulated entities may affect the conclusions we draw about the likely impacts of environmental policies, with a focus on environmental auditing. We examine the conditions under which a multi-facility firm chooses to adopt a standardized auditing. Our firm-level empirical analysis confirms that a uniform auditing outcome is less likely among firms (1) with a more heterogeneous portfolio of facilities, and (2) that are less likely to make decision errors. We examine the implications of adding controls for firm decision-making structure in a facility-level empirical analysis of auditing's impact on compliance. Our results suggest that the estimated compliance impact of auditing varies depending on whether or not we include these controls.

[**] The Robert Day School of Economics and Finance, Claremont McKenna College, mevans@cmc.edu
[*] Department of Economics and International Business, Sam Houston State University, lxl004@shsu.edu
[†] Department of Economics, College of William and Mary, slstaf@wm.edu

I. Introduction

We rarely observe whether decisions relevant to compliance with environmental regulations occur in a centralized (e.g., at the firm level) or decentralized (e.g., at the facility level) manner; we merely observe an outcome, an aggregate firm-level measure or the outcome for a particular facility. Even with well-suited data, the inner workings of regulated entities remain largely a black box to the analyst. As a result, empirical efforts to examine environmental compliance require an implicit assumption about the nature of decision-making within the firm. Often, constraints imposed by the available data drive this assumption. For example, if we observe only firm-level outcomes, then it is natural to assume that the relevant decisions are made by the firm. In this paper, we explore how limits to our insight on the underlying decision-making structure of regulated entities may, in turn, limit our ability to make valid inferences about the likely impacts of environmental policies.

We focus on a particular regulatory setting, environmental auditing. In recent years, industry self-policing has received increased emphasis as a means of improving compliance with environmental regulations. This appears to have been motivated, at least in part, by significant reductions in enforcement resources at the Environmental Protection Agency (EPA), which decreased by 5 percent in real terms between 1997 and 2006 (U.S. GAO, 2007).[1] EPA's 1997 strategic plan formally cites routine environmental auditing as a means to improve environmental compliance among regulated facilities (U.S. EPA, 1997). Many

[1] Other federal agencies have also seen declines in their enforcement budgets. The enforcement budget at the Occupational Safety and Health Administration declined each year from 2001 to 2008, resulting in an overall decrease of almost 8 percent over this time period (OMB Watch, 2008). The Consumer Products Safety Commission's staff decreased from 471 full-time employees in 2005 to 401 full-time employees in 2008 (Consumer Reports, 2007). By the end of 2006, the number of enforcement cases brought by the Securities and Exchange Commission had also fallen significantly, resulting in part from a 3 percent decline in the number of employees in the enforcement unit, according to SEC Chairman Christopher Cox (Washington Post, 2006).

states also actively encourage environmental auditing through statutory privilege for environmental audit reports and immunity from penalties for violations discovered (and corrected and disclosed) during the course of an audit.[2]

In theory, environmental audits provide a mechanism to improve a regulated entity's compliance with environmental regulations as they provide a "systematic review…of facility operations and practices related to meeting environmental requirements."[3] Survey and anecdotal evidence suggests positive impacts of auditing on environmental performance (U.S. GAO, 1995; U.S. EPA, 1999). Additionally, two previous empirical studies explore the impact of environmental auditing on compliance using methods that recognize the potential endogeneity of the audit decision. Khanna and Widyawati (2011) find higher contemporaneous compliance with Clean Air Act (CAA) regulations among facilities whose S&P 500 corporate parent indicates the presence of an environmental auditing program. Evans, Liu and Stafford (2011) report no significant influence of auditing on long-term facility compliance with the Resource Conservation and Recovery Act (RCRA) among a sample of hazardous waste generators in Michigan.

The results of these two previous studies do not necessarily contradict each other as a number of important differences between the analyses could individually or collectively drive the disparate findings. The studies focus on different environmental media (i.e., air versus hazardous waste) and different time frames (i.e., contemporaneous versus long-term). The samples are distinct with Khanna and Widyawti's sample consisting of a set of facilities whose S&P 500 parent companies responded to the Investor Research Responsibility Center (IRRC) survey on environmental management practices survey and Evans et al.'s sample

[2] See Stafford (2005, 2006) and Khanna and Widyawati (2011) for further discussion of these state policies.
[3] "Interim Guidance on Environmental Auditing Policy Statement," 50 FR 46504 (November 8, 1985), Section II.A.

including large and small hazardous waste generators in Michigan. Lastly, the analyses measure environmental auditing at different levels of the firm's decision-making structure. Khanna and Widyawati use a firm-level (i.e., parent company-level) auditing measure while Evans et al. measure auditing at the facility-level. We explore the potential for this final disparity to contribute to the contrasting results. That is, we examine how the assumed decision-making structure of the regulated entity may influence insights regarding the effectiveness of auditing. Without a better understanding of the factors that drive the contrasting results, policy makers cannot effectively assess the extent to which environmental auditing, or other forms of self-policing, can substitute for more traditional compliance and enforcement mechanisms.

In the next section, we adapt Hunnicutt's (2001) model to examine the conditions under which a multi-facility firm chooses a centralized or decentralized decision making structure. The model suggests conditions under which we are likely to observe the same auditing outcome among all facilities owned by a particular firm (i.e., standardization). Section III presents our firm-level empirical analysis, which tests the hypotheses that arise from this model. Our findings highlight the potential importance of controlling for firm decision-making structure in environmental auditing analyses. In Section IV, we explore the implications of this insight for empirical analyses of environmental compliance behavior. Specifically, we use facility-level data to estimate the impact of auditing on compliance with the CAA. We find divergent results with respect to the effect of auditing depending on whether or not we include controls for decision-making structure. Section V discusses the implications of our results.

II. Conceptual model and hypotheses

We recast Hunnicutt's (2001) model of a multi-facility firm to examine decision-making with respect to environmental auditing. If the firm adopts a centralized structure, then a single location (e.g., headquarters) makes auditing decisions for all *n* facilities owned by the firm. Alternatively the firm may choose a decentralized structure in which each facility makes its own auditing decisions. Under *centralization* the firm enjoys economies of scale but faces the possibility of making decision errors. Under *decentralization* each facility makes its optimal auditing decision but the firm does not obtain economies of scale.

Let e denote the effort devoted to environmental auditing at a facility. The benefit to facility *i* of auditing effort is given by $B_i(e)$, which is increasing and strictly concave.[4] For each facility, we assume the existence of some level of effort that is optimal, denoted e_i, which depends on the characteristics of the facility's benefit function. The total benefit of auditing effort to the firm with *n* facilities is given by $\sum_{i=1}^{n} B_i(e_j)$ where e_j, the chosen level of auditing effort, may vary across facilities. Following Hunnicutt (2001), we assume that under centralization, the firm chooses a level of auditing effort for facility *i* that is optimal for some facility *j*. When $j = i$, the chosen level of auditing effort for facility *i* is optimal. A decision error occurs when $j \neq i$; we characterize this as a situation in which the optimal auditing effort for facility *j* has been *misassigned* to facility *i*. This never happens in the decentralized firm but occurs with probability ε under centralization. Thus, the probability that the centralized firm chooses the level of auditing effort that is optimal for a particular facility is $1-\varepsilon$.

[4] There are a number of reasons why a facility might benefit from auditing effort. For example, an audit may afford the facility an opportunity to correct any unknown and/or unintended violations. See Mishra et al.(1997), Pfaff and Sanchirico (2000), and Friesen (2006) for theoretical analyses of environmental auditing.

Each facility faces the same auditing cost function, $c(\cdot)$, which is increasing and concave in auditing effort. This implies that the decentralized firm, in which each facility incurs the costs associated with its auditing effort, faces higher costs than the centralized firm, in which all auditing effort occurs, and the associated costs are incurred, at a single location: $\sum_i c(e_i) > c\left(\sum_i e_i\right)$. Concave costs may, for example, derive from the elimination of duplicated auditing-related activities under centralization (e.g., centralizing auditing activities at a single environmental compliance department is lower cost than maintaining such departments at each facility).

In choosing auditing effort levels for its facilities, the centralized firm has the option of assigning the same auditing effort to all facilities. Following Hunnicutt (2001), we refer to this as *complete standardization*.[5] We use the results derived by Hunnicutt (2001) to explore which of these three possible firm structures, *centralized*, *decentralized*, or *completely standardized*, yields the highest net benefits of auditing effort under various scenarios. The model identifies two key general firm characteristics that drive the decision making structure adopted by a multi-facility firm with respect to auditing, the similarity of the facilities owned by the firm (i.e., in terms of their benefit of auditing effort functions) and the likelihood of misassigning auditing effort. Unfortunately we do not actually observe the decision making structure (decentralized versus centralized) adopted by the firm, nor do we observe auditing effort. Instead, we observe the result of the auditing effort applied at each facility, a binary auditing outcome. Data on auditing outcomes alone does not allow us to distinguish between a decentralized firm with heterogeneous facilities and a centralized firm that chooses a different level of auditing effort at each facility. In either case, we would observe

[5] Of course, identical auditing effort levels for all facilities owned by a decentralized firm can also occur. We explore this further below.

different auditing outcomes for facilities owned by the same firm. We can, however, examine the factors that increase the likelihood of observing the same auditing outcome for all facilities owned by a firm (i.e., all facilities owned by a firm audit or all facilities owned by a firm do not audit), which we call *observed standardization*, verses heterogeneous auditing outcomes. We begin by assuming that the probability of observing an audit at a particular facility is increasing in the facility's auditing effort level.

With homogeneous facilities (in terms of the benefit of auditing function), a centralized or decentralized firm structure leads to complete standardization.[6] As we move away from homogeneous facilities, the centralized firm benefits from recognizing differences among facilities provided the firm does not always misassign auditing effort (i.e. provided $\varepsilon \neq 1$). Recall that the decentralized firm correctly assigns each facility its optimal auditing effort. Since optimal auditing effort depends on the facility's benefit from auditing, a decentralized firm with heterogeneous facilities will not find it optimal to standardize. From this we have the following testable hypothesis:

H1: *Observed standardization* is less likely among firms with heterogeneous facilities.

How does the probability of misassignment affect the likelihood of *observed standardization*? To generate a testable hypothesis, we restrict attention to the firm with a portfolio of heterogeneous facilities.[7] Hunnicutt (2001) shows that when the centralized firm always misassigns auditing effort (i.e., when $\varepsilon = 1$), the optimal strategy for the firm involves complete standardization. When $\varepsilon < 1$, complete standardization is no longer optimal and the decentralized firm benefits from differentiating among facilities when assigning auditing effort levels, even though doing so may lead to decision errors. While

[6] The optimal auditing effort will, however, be different under centralization and decentralization in this case. Hunnicutt (2001) shows that with identical facilities, centralization results in higher auditing effort and higher net benefits to the firm than decentralization. See Proposition 1 of Hunnicutt (p. 540).

[7] In general the firm need own only one facility that differs from the others.

complete standardization is never optimal for the centralized firm when $\varepsilon < 1$, as ε increases in this range, the centralized firm begins to partially standardize. [8] These results suggest our second hypothesis:

H2: *Observed standardization* is less likely among firms that are unlikely to misassign auditing effort levels.

III. Firm-level analysis of standardization in auditing outcomes

Testing H1 and H2 requires facility-level data on environmental auditing. In general, data on environmental auditing are difficult to obtain due in part to the fact that EPA does not require regulated entities to indicate the presence of audit programs. Other analyses of environmental auditing have relied on self-reported firm-level measures of environmental auditing that are insufficient for our analysis. We obtain our data from the Department of Environmental Quality (DEQ) for the state of Michigan. The DEQ requires that regulated facilities provide notice of an intended environmental audit in order to be eligible for immunity from penalties for any violations discovered during the course of the audit. Specifically, the facility must file an "intent-to-audit" notice that identifies the facility at which the audit will be conducted, an indication of the time frame for the audit and a statement of the general scope of the audit.[9]

We obtained a list of the intent-to-audit notices filed with the DEQ between 1998 and 2003. The data include the company and facility name, a mailing address, and the date the notice was filed. We used this information to match each facility to EPA's Facility

[8] See Propositions 4 and 5 of Hunnicutt (p. 542).

[9] While it is possible that a facility might conduct an environmental audit without first notifying the DEQ, Michigan provides strong incentives for facilities to file intent-to-audit notices; a primary benefit of auditing is the potential for penalty mitigation and this benefit is available only to auditing facilities that submit the required intent-to-audit notice. Additionally, Michigan passed legislation in 1996 that grants legal privilege to all environmental auditing documents, mitigating concerns that entities would be reluctant to disclose environmental audits. See Evans et al. (2011) for further discussion of these data.

Registry System (FRS) to identify the federal facility identification number and to determine the media programs under which the facility is regulated. Our empirical analysis focuses on facilities regulated under the Clean Air Act (CAA) in Michigan and the firms that own these facilities. We obtained data on facility characteristics from EPA's Aerometric Information Retrieval System (AIRS) Facility Subsystem (or AFS) database.[10] Using DUNS numbers (when available) and owner names from the FRS, along with company name from the AFS, we matched CAA-regulated facilities in Michigan that were owned by the same firm.[11] To test H1 and H2, we of course restrict attention to multi-facility firms; we have 171 multi-facility firms represented in our data.

We identify an audit at a facility if the facility submitted at least one intent-to-audit notice between 1998 and 2003. In this case, the variable *Audit* equals one for the facility. Otherwise, we do not measure an audit at the facility and *Audit* equals zero. Our outcome variable of interest for our firm-level analysis is *Standardization*, which equals one if $Audit = 1$ **or** $Audit = 0$ for all facilities owned by a firm. Otherwise, *Standardization* equals zero. Our analysis includes 171 firms that own 730 facilities. We observe standardization for 155 of these firms (about 90%), which accounts for 590 facilities (81%). For the remaining 10 percent of firms, we observe heterogeneous auditing outcomes among their 140 facilities. Note that since our data contain facilities located in a single state, the degree of standardization we observe is likely to be higher than if our dataset included facilities located in several states. Since policies on audit privilege and audit immunity vary by state, firms

---

[10] The AFS database was downloaded from EPA's Envirofacts System in August 2007. AFS also lists 3 "portable sources" located in Michigan, but we excluded them from this analysis since they do not have a fixed facility location.

[11] As a first pass, we electronically matched facilities owned by the same firm using DUNS numbers and company names but since the owner and company name fields do not have standardized formats (e.g., no standard abbreviations or punctuation rules), we made many of the matches manually for facilities with missing DUNS numbers.

that operate in multiple states may be more likely to adopt a standardized audit policy at their facilities *within* a given state (verses adopting a standardized policy at all of their facilities).

To test H1 and H2, we estimate a simple probit model where our dependent variable is *Standardization*. As a first step in this analysis, we operationalize the notion of heterogeneous facilities (owned by the same firm) and the likelihood of misassigning auditing effort levels. We create three measures to identify firms with more heterogeneous portfolios of CAA-regulated facilities in Michigan. Each measure captures a different dimension of facility heterogeneity and takes a value between zero and one where zero denotes a homogeneous portfolio of firms with respect to that characteristic.[12] The first measure, *Product heterogeneity*, focuses on the nature of production at each facility as measured by the up to three 4-digit Standard Industrial Classification (SIC) codes associated with the facility. If the sequence of 4-digit SIC codes is identical for all facilities in the firm's portfolio, then *Product heterogeneity* equals zero. If each facility in the firm's portfolio has a unique sequence of 4-digit SIC codes, then *Product heterogeneity* is equal to one.[13] If the firm's portfolio contains a mix of unique and repeated sequences of SIC codes among its facilities, then *Product heterogeneity* takes some value between zero and one. In general, a higher value (closer to one) of *Product heterogeneity* indicates more diversity among the firm's facilities in terms of their production activities.

Our second measure, *Regulatory heterogeneity*, examines the nature of the regulatory environment at the facilities owned by a firm. While our analysis only includes facilities regulated under the CAA, our dataset contains information on other environmental

[12] See the appendix for a more detailed description of these measures.
[13] Our results are unchanged if we use an alternative product heterogeneity measure created with the sequence of 2-digit SIC codes.

programs to which the facility is subject.[14] We create five indicator variables from this information. The first two variables indicate that the facility is subject to the provisions of the Toxics Release Inventory (TRI) or the Resource Conservation and Recovery Act (RCRA), respectively. The remaining variables indicate that the facility shows up in an EPA database, the Permit Compliance System (PCS), the Integrated Compliance Information System (ICIS), or the Comprehensive Environmental Response, Compensation, and Liability Information System (CERCLIS), respectively. PCS and ICIS designate facilities that are registered with EPA's federal enforcement and compliance and that hold National Pollutant Discharge Elimination System permits. A facility in CERCLIS is on (or being considered for) Superfund's National Priorities List. *Regulatory heterogeneity* measures the diversity among facilities in terms of the sequence of these five indicator variables. If the sequence of these five indicator variables is the same for all facilities in the firm's portfolio (i.e., the facilities face the same environmental regulatory environment), then *Regulatory heterogeneity* equals zero. A value closer to one indicates more diversity among the firm's facilities in terms of their regulatory exposure.

Our final measure, *Size heterogeneity*, assesses the degree of heterogeneity in terms of the size of the facilities owned by the firm, as measured the number of employees at the facilities.[15] To construct this measure, we first create a categorical variable to characterize the number of employees at the facility as less than or equal to 10, greater than 10 but less than or equal to 100, greater than 100 but less than or equal to 500, or greater than 500. *Size*

[14] This information is available in the FRS database. For some facilities in the FRS database, there is more than one facility identified in the AFS, perhaps due to different definitions of what constitutes a facility between the two databases. Because we have no way of aggregating the AFS data, we use the AFS observations as the primary observations.

[15] The AFS provides a variable described as the number of employees at the facility. However, this variable takes a value of zero for a large fraction of facilities. If we exclude our size heterogeneity measure from the firm-level analysis below, our results with respect to the other included variables are unchanged. Alternatively, if we assume these zeros are in fact missing values, while our sample size decreases, our results are qualitatively similar.

heterogeneity equals zero if all facilities in the firm's portfolio have the same value for this categorical variable and one if each facility in the firm's portfolio has a unique value for this categorical variable. Thus, values of *Size heterogeneity* closer to one indicate more heterogeneity among the firm's facilities in terms of facility size. Based on H1, we expect negative and significant coefficients on these 3 measures of heterogeneity in our probit models.

Unfortunately, our data do not provide an ideal measure for a firm's propensity to misassign auditing effort levels. To proxy for this measure, our model includes a dummy variable, *Public*, that equals one if the firm is publicly traded and zero otherwise. Relative to private firms, publicly traded firms may have more to lose by misassigning auditing effort (e.g., if misassignment results in the firm failing to identify and correct violations that are later discovered by the regulator). If so, then publicly traded firms would be less likely to assign auditing efforts levels and H2 would predict a negative and significant coefficient on *Public*. We also include a variable that measures the number of CAA-regulated facilities in Michigan owned by the firm, *# facilities*. Firms that own more facilities may have more resources available or may have greater access to specialized auditing resources, which may decrease the likelihood of misassignment. On the other hand, all else equal, a higher number of facilities complicates the assignment of auditing effort and thus could increase the likelihood of misassignment. Therefore, the coefficient on *# facilities* could be positive or negative.

Table 1 presents descriptive statistics as well as the results of our probit analysis. The first column reports means and standard deviations. While the sample average for *Product heterogeneity* is 0.67, 35 firms (about 20%) are homogenous along this dimension and 98 firms (about 57%) have a value of *Product heterogeneity* equal to one. 52 firms are homogeneous with respect to regulatory heterogeneity and 66 firms have the maximum value of regulatory heterogeneity. With respect to size, 71 firms are homogeneous and 50 firms are maximally heterogeneous. Thirty percent of the firms in our sample are publicly traded. The average firm owns just over four facilities (although the median number of facilities is two).

The third column reports estimated coefficients and robust standard errors from our probit model. The last column reports conditional marginal effects and standard errors obtained using the Delta method. The results suggest that product and regulatory heterogeneity significantly reduce the likelihood of observing standardization at a multi-facility firm. The coefficient on *Size heterogeneity* is also negative, as expected, but not significantly different from zero. The conditional marginal effect for *Product heterogeneity* suggests that the most heterogeneous firm (i.e., *Product heterogeneity=1*) is 2.6 percentage points less likely to standardize than the homogeneous firm (p-value=0.058). In terms of regulatory heterogeneity, the most heterogeneous firm is 3.6 percentage points less likely to standardize than the firm with *Regulatory heterogeneity* equal to zero (p-value=0.014). While these estimated effects are not huge, they confirm the predictions of our conceptual model. Standardization is also less likely among publicly traded firms; publicly traded firms are 5 percentage points less likely to standardize than privately held firms. If publicly held firms are less likely to misassign auditing effort levels, then this result is consistent with H2. Lastly, the probability of standardization decreases with the number of facilities owned by the firm although the estimated conditional marginal effect is small. The results of our firm-level empirical analysis are broadly consistent with the predictions of our conceptual model.

What are the implications of our firm-level analysis for empirical models of environmental auditing? First, our results suggest that the measurement error introduced by assigning the same auditing outcome to all facilities owned by a firm, as is necessary with only a firm-level auditing outcome, is likely to be greater for (1) firms with heterogeneous portfolios of facilities and (2) firms that are unlikely to misassign auditing effort levels. For how many facilities would we incorrectly assign an auditing outcome were we to substitute our facility-level audit outcome with a firm-level outcome? To explore this question, we create two different firm-level auditing outcomes with our data. *Firm audit1* equals one if the firm owns at least one facility that submitted an intent-to-audit notice and zero otherwise. *Firm audit2* equals one if all of the facilities owned by the firm submitted intent-to-audit notices and zero otherwise.[16] *Firm audit1* differs from *Audit*, the facility-level audit measure, for 71 of the 730 facilities (about 10%) in our data. *Firm audit2* differs from *Audit* for 69 facilities. The distributions of both *Firm audit1* and *Firm audit2* are significantly different from that of *Audit* (both with p-value=0.00). Ultimately we do not know whether either of these measures would align with a firm-level auditing outcome obtained from a survey of firms, as used in Khanna and Widyawati (2011). Regardless, this exercise suggests caution in assuming complete standardization in auditing outcomes among facilities owned by the same firm.

Second, our findings hint at the potential importance of controlling for firm decision-making structure in environmental auditing analyses, whether at the firm- or facility-level. In the next section, we explore the implications of this second insight in a facility-level analysis of the impact of environmental auditing on compliance with the Clean Air Act (CAA).

[16] It remains possible that (1) a firm with *Firm audit1* equal to zero audits at one of its facilities located outside the state of Michigan or (2) a firm with *Firm audit2* equal to one chooses not to audit at a facility located in another state. Our two firm-level audit measures are defined based on the sample of facilities we observe, which are all located in Michigan.

IV. Facility-level analysis of environmental auditing and compliance with the Clean Air Act

Our facility-level analysis examines the factors that encourage environmental auditing at a facility as well as the impact of auditing on the facility's long term compliance with the CAA. To do so, we expand our dataset beyond the multi-facility firms included in our analysis above to include stand-alone facilities. We extract additional facility characteristics, enforcement history, and current compliance status from the AFS. We also linked the AFS database to EPA's Enforcement and Compliance History Online (ECHO) and Integrated Data for Enforcement Analysis (IDEA) databases to obtain additional enforcement and compliance measures.

The CAA requires facilities to self-report their compliance status on an on-going basis (i.e., each quarter). The self-reported nature of the compliance data introduces the potential for bias. While existing evidence on the accuracy of self-reported environmental compliance data is mixed, analyses of self-reported compliance measures that are used for enforcement purposes have been unable to reject the accuracy of these data (Laplante and Rilstone, 1996; Shimshack and Ward, 2005).[17] The use of self-reported data for enforcement purposes provides strong incentives, such as criminal fines, to encourage truthful reporting. The CAA self-reported data are used for enforcement. We use these data to create a binary compliance variable, *Comply*, which equals one if the facility is in compliance for each of the 12 quarters between the second quarter of 2004 and the first quarter of 2007.[18] *Comply* equals zero if the facility is out of compliance for at least one of these 12 quarters. In

[17] deMarchi and Hamilton (2006) find reporting irregularities in self-reported data from the Toxics Release Inventory but these data are not used by regulators for enforcement purposes.

[18] Facilities are included in our analysis provided we observe compliance or non-compliance status for all 12 quarters. Facilities with missing compliance data during this time period are excluded from our analysis. We re-estimated our models including all facilities with at least 2 quarters of data in this time period and we obtain qualitatively similar results.

focusing on the longer run impacts of auditing on compliance with the CAA, our analysis differs from Khanna and Widyawati (2011), who examine contemporaneous compliance effects.

A key concern in conducting an empirical analysis of the effect of audits on compliance behavior is the potential for endogeneity. Specifically, because violations discovered during the course of an audit are eligible for penalty mitigation (with the possibility of a complete waiver of penalties), facilities that are concerned that they are non-compliant may be more likely to implement an environmental auditing program. If so, then the audit decision is correlated with factors that affect the compliance status of the facility. While this may be less of a concern given our emphasis on the longer run effects of auditing, we allow for this possibility by estimating a bivariate probit model.

Let q_i represent the binary compliance outcome for facility *i*, which depends on the latent variable q_i^* representing facility *i*'s net benefit from compliance where

$$q_i^* = x'_{qi}\beta_q + a_i\delta_q + \varepsilon_{qi}$$

with

$$q_i = \begin{cases} 1\, if\;\; q_i^* \geq 0 \\ 0\; otherwise \end{cases}.$$

x_{qi} is a vector of explanatory variables that proxy for the costs and benefits of compliance and a_i is a binary variable representing facility *i*'s observed auditing status (i.e. *Audit* as defined above). a_i depends on an underlying latent variable a_i^*, which represents facility *i*'s net benefit from conducting an audit during the period of the analysis and is given by:

$$a_i{*} = x'_{ai}\beta_a + \varepsilon_{ai}$$

with

$$a_i = \begin{cases} 1 \; if \;\; a_i^* \geq 0 \\ 0 \; otherwise \end{cases}.$$

x_{ai} denotes a vector of explanatory variables that proxy for the costs and benefits of environmental auditing. We assume the error terms, ε_{qi} and ε_{ai}, follow a bivariate normal distribution with zero means and covariance matrix $\begin{bmatrix} 1 & \rho \\ \rho & \sigma_q \end{bmatrix}$. Maddala (1983) derives the full information maximum likelihood estimator for this model, which we estimate in Stata using the biprobit command. According to Wilde (2000), the model is identified with sufficient variation in the independent variables.[19]

We can broadly classify the primary independent variables included in our analysis as those related to the facility's characteristics, its inspection and compliance history, the stringency of the CAA regulations to which the facility is subject, its environmental exposure, and county-level characteristics for the county in which the facility is located. To explore the potential importance of controlling for the firm's decision-making structure, some of our specifications also include firm-level variables related to firm structure. Table 2 provides variable descriptions and summary statistics for the variables included in our facility-level analysis.

Of the 2811 facilities included in our analysis, 111 (about 4%) submitted intent-to-audit notices during the time frame we consider.[20] We control for facility size using *Employees*, which measures the number of employees at each facility. Broad industry differences are captured by the variable *Manufacturing*, which equals one for facilities

[19] Khanna and Widyawati (2011) adopt the same identification strategy.
[20] The 4 percent audit rate is roughly consistent with Potoski and Prakash's (2005) estimate that approximately 4 percent of "major" CAA-regulated facilities participate in the ISO14001 certification program, a program which requires (among other things) adoption of an environmental auditing protocol.

classified as manufacturing (i.e., facilities that have 2-digit SIC codes between 20 and 39) and zero otherwise. The majority of facilities included in our analysis are manufacturing.

We include four variables to control for the facility's inspection and compliance history. *Past inspection* equals one if the facility was inspected at least once between 1994 and 1998. We measure an inspection as any federal or state inspection, compliance evaluation (onsite or offsite), or any source test to check for compliance. *Count past inspection* measures the number of times the facility was inspected between 1994 and 1998. *Past violation* denotes a violation between 1994 and 1998. In particular, *Past violation* equals one if the facility received a federal or state notice of violation, a notice of non-compliance, or if a federal or state administrative order related to non-compliance was filed against the facility. *Past penalty* measures the sum of any penalties under CAA the facility paid between 1994 to 1998 (in $1,000s). Forty three percent of facilities in our sample were inspected at least once during this time period with about five percent of facilities in violation during this timeframe. The average penalty paid is $31,000 but the median penalty is zero.[21]

The variable *Major* denotes facilities that are classified as a "major" stationary source of air pollution, a source that emits at least 10 tons per year of any of the listed toxic air pollutants or 25 tons of a mixture of air toxics. Because regulations for major air sources are more complex and stringent than those for other sources, this designation may affect both the audit and compliance decisions for a facility. Fifteen percent of facilities in our sample have this designation. The next set of variables in Table 2, *MACT* through *CFC*, indicates the particular regulatory air program(s) to which the facility is subject. These variables were derived from the air program code information available in the AFS database. Some of these programs are quite large, while others are more specialized. For example, 94 percent

[21] The mean is misleading here given that the vast majority of facilities did not pay a penalty. In fact, *Past penalty* is nonzero for only 53 facilities. Among these facilities, the mean penalty is about $1700.

of facilities are classified as SIP sources while less than one percent is subject to new source review. As different regulatory programs within CAA entail different standards, these variables are potentially important controls in our analysis.

The next five variables, *CERCLIS* through *TRI*, which were defined above, indicate other EPA programs to which the facility is subject. About one third of the facilities in our analysis are also subject to the reporting requirements of the Toxics Release Inventory and over half are also regulated under the Resource Conservation and Recovery Act.

Our analysis includes four county-level variables. The variable *Nonattainment* denotes facilities located in one of 25 Michigan counties that were designated as non-attainment for the 8-hour ozone standard in 2004. The county non-attainment designation may affect facilities' expectations about the likelihood of inspection and/or sanction and therefore the compliance decision. The next three variables are proxies for the general environmental, political, and economic climate of the county in which the facility is located. To capture the strength of the environmental constituency in each county we include *County conservancy*, which measures the number of Nature Conservancy members per 1000 residents of the county.[22] As noted by Innes and Sam (2008), a larger environmental constituency may suggest a higher degree of public awareness of a facility's environmental performance and more successful lobbying of local government by environmental interest groups. *County Republicans* indicates the percentage of voters in the 2000 Presidential election that voted for the Republican candidate, George W. Bush and *County education* indicates the percent of the

[22] County-level data on membership in environmental organizations is not readily available. We thank Mary Thomas and Donald Zeilstra from the Nature Conservancy, Michigan Field Office, for providing these data.

county's population aged 25 and older whose highest level of education achieved was high school.[23]

Some of our specifications include two additional variables that measure firm characteristics. In order to explore some of the insight gleaned from our firm-level analysis, we would ideally examine the implications of including controls for the firm's chosen decision-making structure. Unfortunately, our data do not contain such measures. Instead, we include two variables that we argue are likely related to this characteristic. The first variable, *Public*, equals one if the firm that owns the facility (i.e., the facility's parent company) is publicly-traded and zero if it is privately held. The increased visibility of publicly-traded companies as well as their responsibility to shareholders may affect their adoption of a particular structure. We also know from the firm-level analysis that the likelihood of adopting a standardized auditing policy is different for publicly-traded and privately-held firms. The second variable, *Multifacility*, takes the value of one if the firm owns more than one CAA-regulated facility in Michigan. The distinction between centralized and decentralized structure is irrelevant for stand-alone facilities but potentially important for multi-facility firms as we saw in our firm-level analysis. Ten percent of facilities in our sample belong to publicly-traded firms while about 20 percent belong to a multi-facility firm.[24]

We follow Evans et al. (2011) and Khanna and Widyawati (2011), the two existing analyses of environmental auditing of which we are aware, in selecting facility- and county-level variables for inclusion in the audit and compliance equations. Table 3 provides the

[23] The data used to create *County Republican* were extracted from the Michigan Department of State's website (http://miboecfr.nicusa.com/election/results/00gen/01000000.html). *County education* is from the 2000 Census.

[24] Note that due to missing data for some of the variables include in our facility-level analysis, not all of the 730 facilities whose parent companies are included in our firm-level analysis are present in our facility-level analysis. 542 facilities whose firms were included in our firm-level analysis are present in our facility-level sample.

results of our facility-level bivariate probit analysis. The results of the audit and compliance equations are given in the left-hand and right-hand panels, respectively. We report results for two specifications that differ in terms of whether or not we include the two firm-level measures intended to act as controls for firm decision-making structure. Because Model I excludes these firm-level controls, it is similar to those considered in previous analyses of environmental auditing.[25] Model II is identical to Model I other than the inclusion of *Public* and *Multifacility*, our controls for the decision-making structure adopted by the facility's parent company. Model I is reported in the second and fourth columns while Model II is reported in the third and fifth columns.

The results of Models I and II are broadly consistent with respect to the variables that measure facility characteristics, inspection and compliance history, stringency of the CAA regulations the facility faces, environmental exposure, and county-level characteristics. Our results suggest that larger facilities (as measured by the number of employees) are more likely to audit but are less likely to comply. The coefficients on *Past inspection*, *Past violation*, the interaction of these variables, and *Past penalties* have the same signs as comparable measures in Khanna and Widyawati (2011) but the estimated effects are insignificant in our models. The variables that measure the stringency of the CAA regulations and the other environmental regulations to which the facility is subject are important controls in our models. In particular, facilities classified as major stationary sources under CAA are more likely to audit but less likely to comply. The estimated coefficients on *NSPS*, *SIP*, *CFC* in the compliance equation also indicate significant effects. Facilities listed in CERCLIS are

[25] This analysis and that of Khanna and Widywati (2011) examine the effects of auditing on compliance with the CAA. However, the significant differences between the dataset we use here and the dataset used by Khanna and Widyawati make replicating their specifications unfeasible. While our data are closer to the sample used by Evans et al. (2011) as both samples focus on facilities in Michigan, Evans et al.'s sample includes manufacturing facilities regulated under RCRA while our sample restricts attention to CAA-regulated facilities.

less likely to audit while those subject to TRI and RCRA are more likely to audit. The positive coefficients on *County education* suggest that facilities located in counties with more educated residents are more likely to audit and comply although only the former effect is significant (in Model II). The significant coefficients on *County Republicans* suggest that facilities located in counties with a larger share of Republican voters are less likely to audit and more likely to comply. While somewhat unexpected, these results are consistent with Evans et al. (2011).

The results from Model I, which excludes controls for firm structure, and Model II, which includes these controls, differ along two important dimensions. First, the primary variable of interest in the compliance equation, *Audit*, is positive and significant in Model I but insignificant (and negative) in Model II. The former suggests that, controlling for the endogeneity of *Audit*, auditing facilities are more likely to be in compliance. The latter result suggests no significant impact of auditing on long-term compliance. Second, the estimate for ρ in Model I is negative and significant, which suggests a correlation between the unobserved factors that affect *Audit* and *Comply*. The estimate for ρ is insignificant in Model II. Along both of these dimensions, Model I is consistent with the findings of Khanna and Widyawati (2011) while Model II aligns with Evans et al.'s (2011) results.

A Wald test confirms that we are able to reject the null hypothesis that *Public* and *Multifacility* are jointly equal to zero ($\chi^2(4) = 40.92$, p-value = 0.00). Thus, including these variables significantly improves the fit of the model. If, as we argue, these variables are related to the decision-making structure chosen by the firm, then our results suggest that such controls are important in facility-level analyses of environmental auditing.

V. Discussion

As external observers, we rely on data provided by regulated entities, through surveys or through self-reports, or by regulators to measure environmental compliance and related outcomes. Our empirical analyses of these data require an assumption, which often goes unstated, regarding the nature of decision-making within the regulated entity. Our results speak to the importance of explicitly recognizing this feature in empirical analyses.

Our firm-level analysis suggests that assigning the same auditing outcome for all facilities owned by the same firm (i.e., assuming standardization) is more problematic for firms with heterogeneous portfolios of facilities and for firms that are unlikely to make decision errors. This systematic measurement error has the potential to lead to erroneous conclusions regarding the effectiveness of programs designed to encourage regulated entities to undertake environmental audits.

Our facility-level analysis finds that the estimated impact of auditing on compliance varies depending on whether or not we include controls for firm decision-making structure. When such controls are excluded, our empirical results suggest a positive impact of auditing on long-run compliance with the CAA. This effect disappears once these controls are added. Of course, our controls are not perfect measures of firm decision-making structure; the firm is likely to remain a black box in this regard. However, as in our dataset, many datasets are likely to contain a handful of observable firm characteristics that may be related to firm structure and can therefore be used as proxies.

Our data permit an investigation of these issues within the context of environmental auditing. However, note that the decision to conduct an environmental audit is closely related to the decision to participate in other voluntary environmental initiatives (VEIs).[26] If

[26] VEIs include voluntary programs sponsored by regulatory agencies (e.g., EPA's 33/50 program) as well as industry associations (e.g., Responsible Care, a program sponsored by the U.S. Chemical Manufacturers

firms and facilities face similar incentives in these two settings, then the lessons from our facility-level empirical analysis extend to empirical analyses of VEI participation. Our results are also relevant for empirical analyses of environmental compliance more generally.

Association), third-party programs (e.g., ISO14001), and firm-specific initiatives (e.g., adoption of environmental management systems).

REFERENCES

Consumer Reports. 2007. "CPSC Budget: Reading Between the Lines." February 27.

de Marchi, Scott and James T. Hamilton. 2006. "Assessing the Accuracy of Self-Reported Data: An Evaluation of the Toxics Release Inventory." *Journal of Risk and Uncertainty* (32): 57-76.

Evans, Mary F., Lirong Liu, and Sarah Stafford. 2011. "Do Environmental Audits Improve Long-term Compliance? Evidence from Manufacturing Facilities in Michigan." *Journal of Regulatory Economics* (40): 279-302.

Friesen, Lana. 2006. "The Social Welfare Implications of Industry Self-Auditing." *Journal of Environmental Economics and Management* (51): 280-294.

Hunnicutt, Lynn. 2001. "Mixups in the Warehouse: Centralized and Decentralized Multi-plant Firms." *Economic Inquiry* 39(4): 537-548.

Khanna, Madhu and Diah Widyawati. 2011. "Fostering Regulatory Compliance: The Role of Environmental Self-Auditing and Audit Policies." *Review of Law and Economics* (7): 129–163.

Laplante, Benoit and Paul Rilstone. 1996. "Environmental Inspections and Emissions of the Pulp and Paper Industry in Quebec." *Journal of Environmental Economics and Management* (31): 16-36.

Maddala, G.S. 1983. *Limited-Dependent and Qualitative Variables in Econometrics*. Cambridge, England: Cambridge University Press.

Mishra, Birendra K., D. Paul Newman and Christopher H. Stinson. 1997. "Environmental Regulations and Incentives for Compliance Audits." *Journal of Accounting and Public Policy* (16): 187–214

OMB Watch. 2008. "Workers Threatened by Decline in OSHA Budget, Enforcement Activity." January 23.

Pfaff, Alexander S. P. and Chris William Sanchirico. 2000. "Environmental Self-Auditing: Setting the Proper Incentives for Discovery and Correction of Environmental Harm." *Journal of Law, Economics, and Organization* (16): 189-208.

Potoski, Matthew and Aseem Prakash. 2005. "Green Clubs and Voluntary Governance: ISO 14001 and Firms' Regulatory Compliance." *American Journal of Political Science* (49): 235-248.

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

Stafford, Sarah L. 2005. "Does Self-Policing Help the Environment? EPA's Audit Policy and Hazardous Waste Compliance." *Vermont Journal of Environmental Law* 6.

Stafford, Sarah L. 2006. "State Adoption of Environmental Audit Initiatives." *Contemporary Economic Policy* 24(1): 172-187.

U.S. General Accounting Office (GAO). 1995. "Environmental Auditing: A Useful Tool that Can Improve Environmental Performance and Reduce Cost Savings." GAO/RCED 95-37.

U.S. Government Accountability Office (GAO). 2007. "EPA-State Enforcement Partnership Has Improved, but EPA's Oversight Needs Further Enhancement." GAO-07-883.

U.S. Environmental Protection Agency (EPA). 1997. "EPA Strategic Plan." EPA/190-R-97 002.

U.S. Environmental Protection Agency (EPA). 1999. "EPA/CMA Root Cause Analysis Pilot Project: An Industry Survey,"

Washington Post. 2006. "SEC Enforcement Cases Decline 9% Staff Reduced Because of Budget Crunch." November 3.

Wilde, Joachim. 2000. "Identification of Multiple Equation Probit Models with Endogenous Dummy Regressors." *Economics Letters* 69(3): 309-312.

Table 1. Probit results for firm-level analysis of standardization in auditing outcomes

	Sample mean (standard deviation)	Estimated coefficient (robust standard error)	Conditional marginal effect (Delta-method standard error)
Product heterogeneity	0.67 (0.42)	-0.89* (0.46)	-0.026 (0.014)
Regulatory heterogeneity	0.51 (0.43)	-0.88* (0.41)	-0.036* (0.015)
Size heterogeneity	0.39 (0.42)	-0.091 (0.80)	-0.0080 (0.031)
Public	0.30 (0.46)	-0.80** (0.31)	-0.050** (0.019)
# facilities	4.27 (6.47)	-0.060** (0.018)	-0.0057** (0.0022)
Constant	--	3.25** (0.53)	--

N=171, pseudo $R2=0.22$

For all variables except *# facilities*, conditional marginal effects are calculated as the response for the change of going from a value of zero to a value of one (at the means of the other variables). For *# facilities*, the conditional marginal effect is calculated at the means of all the variables.

Table 2. Variable description and summary statistics for facility-level analysis of environmental auditing

Variable name	Variable description	Mean	Standard deviation
Facility-level variables			
Comply	= 1 if facility reported complete compliance from 2nd quarter 2004 through 1st quarter 2007, = 0 if facility was out of compliance for one or more quarters	0.80	0.40
Audit	= 1 if facility submitted at least one intent to audit notice between 1998 and 2003	0.040	0.19
Employees	Number of employees at facility	169.24	1013.35
Manufacturing	= 1 if facility is classified as manufacturing (SIC codes 20 – 39)	0.55	0.50
Past inspection	= 1 if facility was inspected at least once between 1994 and 1997	0.42	0.49
Count past inspection	Number of times facility was inspected between 1994 and 1998	0.97	1.62
Past violation	= 1 if the facility was cited for a violation at least once between 1994 and 1998	0.049	0.22
Past penalties	Total CAA penalties paid from 1994 to 1998, in $1,000s	31.91	571.45
Major	= 1 if facility is classified as a major air source	0.15	0.36
MACT	= 1 if facility is subject to MACT (maximum achievable control technology).	0.071	0.26
PSD	= 1 if facility is subject to PSD (prevention of significant deterioration)	0.049	0.22
NSR	= 1 if facility is subject to NSR (new source review)	0.0068	0.082
NSPS	= 1 if facility is subject to NSPS (new source performance standards)	0.17	0.38
SIP	= 1 if facility is classified as a SIP (state implementation plan) source	0.94	0.24
CFC	= 1 if facility is subject to CFC tracking	0.052	0.22
CERCLIS	=1 if the facility is tracked in CERCLIS	0.015	0.12
ICIS	= 1 if facility is tracked ICIS	0.13	0.34
PCS	=1 if the facility is tracked in PCS	0.11	0.32
RCRA	= 1 if facility is regulated under RCRA	0.57	0.49
TRI	=1 if the facility is subject to TRI reporting	0.31	0.46
County-level variables			
Nonattainment	= 1 if facility is located in county that was classified as non-attainment for ozone in 2004	0.65	0.48
County conservancy	Number of individuals in the county that belonged to the Nature Conservancy in 2005 per 1000 residents	2.49	1.05
County Republicans	Percentage of the voters in the county in the 2000 Presidential election that voted Republican	0.48	0.11
County	Percent of the county's population aged 25 and older	83.32	4.22

education	whose highest level of education achieved was high school in 1990		
Firm-level variables			
Public	= 1 if firm (parent company) is publicly traded	0.10	0.31
Multi-facility	=1 if the firm (parent company) owns more than one CAA-regulated facility in Michigan	0.19	0.39

Table 3. Results of the facility-level analysis of environmental auditing

	Audit equation		Compliance equation	
Variable name	**Model I**	**Model II**	**Model I**	**Model II**
Audit	--	--	1.41** (0.25)	-0.51 (0.87)
Employees	0.00011* (0.000058)	0.000094* (0.000053)	-0.00021** (0.000078)	-0.00015 (0.000090)
Manufacturing	-0.015 (0.12)	0.0061 (0.13)	-0.015 (0.077)	-0.026 (0.082)
Past inspection	--	--	-0.088 (0.074)	-0.10 (0.082)
Past violation	--	--	-0.023 (0.41)	-0.020 (0.44)
Past inspection*Past violation	--	--	0.13 (0.44)	0.12 (0.48)
Count past inspections	-0.013 (0.027)	-0.026 (0.027)	--	--
Past penalties	-0.00062 (0.00049)	-0.00057 (0.00044)	-0.00014 (0.00017)	-0.00017 (0.00018)
Major	0.37** (0.14)	0.32** (0.15)	-1.95** (0.11)	-2.07** (0.11)
MACT	0.13 (0.16)	0.25 (0.16)	-0.13 (0.13)	-0.10 (0.14)
PSD	0.32* (0.18)	0.37* (0.19)	-0.32* (0.18)	-0.28 (0.20)
NSR	-0.93* (0.53)	-1.019 (0.56)	-0.23 (0.50)	-0.37 (0.52)
NSPS	0.078 (0.14)	0.034 (0.15)	-0.24** (0.092)	-0.21** (0.097)
SIP	0.56 (0.35)	0.52 (0.37)	0.71** (0.18)	0.83** (0.18)
CFC	0.47* (0.24)	0.38 (0.28)	-1.04** (0.20)	-1.04** (0.21)
CERCLIS	-5.08** (0.35)	-6.38** (0.93)	0.12 (0.34)	-0.13 (0.33)
ICIS	0.15 (0.11)	0.14 (0.13)	0.054 (0.11)	0.085 (0.12)
PCS	-0.041 (0.13)	-0.086 (0.14)	-0.086 (0.11)	-0.14 (0.12)
RCRA	0.44** (0.15)	0.45** (0.16)	0.077 (0.081)	0.15* (0.08)
TRI	0.65** (0.13)	0.58** (0.14)	-0.19** (0.095)	-0.038 (0.10)
Nonattainment	--	--	-0.057 (0.073)	-0.080 (0.080)

County conservancy	-0.077 (0.072)	-0.11 (0.071)	0.0092 (0.041)	0.0040 (0.041)
County Republicans	-1.99** (0.55)	-2.046** (0.56)	1.64** (0.39)	1.36** (0.44)
County education	0.029 (0.020)	0.035* (0.021)	0.014 (0.011)	0.018 (0.012)
Public	--	0.21 (0.15)	--	0.27* (0.15)
Multi-facility	--	0.53** (0.12)	--	-0.24** (0.11)
Constant	-4.43** (1.47)	-4.99** (1.53)	-1.05 (0.84)	-1.34 (0.87)
Rho	-0.82** (0.096)	0.14 (0.41)	--	--
Log L	-1168.60	-1150.76	--	--

Table reports estimated coefficients and robust standard errors, below coefficients in parentheses. Sample size is 2811. * indicates significance at 10% and ** indicates significance at 5%.

APPENDIX. Description of heterogeneity measures

Product heterogeneity

The dataset contains up to three two-digit SIC codes for each facility. Let SIC1, SIC2, SIC3 represent these variables. Three steps are required to create *Product heterogeneity*.

Step 1:

Sort facilities by parent company identifier, SIC1, SIC2, SIC3. Create a variable, called *Product duplicates*, that equals zero if a particular sequence of SIC1, SIC2, SIC3 is unique among facilities within a parent company. Otherwise, *Product duplicates* assigns a count to identify the number of times a particular sequence of SIC1, SIC2, SIC3 shows up within facilities owned by the parent company (i.e., the first facility with the sequence will have *Product duplicates* equal to one, the second facility with that same sequence will have *Product duplicates* equal to two, and so on). Consider a hypothetical example firm that owns five facilities with values for SIC1, SIC2, SIC3 given in the following table (-- indicates a missing value):

Facility ID	SIC1	SIC2	SIC3	Duplicates
1	30	37	--	0
2	20	--	--	1
3	20	24	26	0
4	20	--	--	2
5	20	--	--	3

Two sequences of SIC1, SIC2, SIC3 are unique within the facilities owned by this firm (30, 37, --; 20, 24, 26) so the facilities with these sequences (1 and 3) are assigned values of *Product duplicates* equal to zero. One sequence, 20, --, --, occurs three times, first for facility 2 (so *Product duplicates*=1), second for facility 4 (so *Product duplicates*=2), and third for facility 5 (so *Duplicates*=3).

Step 2:

Identify the maximum value of *Product duplicates* among facilities owned by each parent company. Call this *Max product duplicates*. Note that if a firm has *Max product duplicates* equal to *Num facilities* (the # of facility owned by the parent company), then all facilities owned by that parent company have the same values for the sequence SIC1, SIC2, SIC3. If *Max product duplicates* is equal to zero, then each facility owned by the firm has a unique sequence of SIC1, SIC2, SIC3.

Step 3:

Form *Product heterogeneity* for each firm using the following:

$$\text{Product heterogeneity} = \frac{Num\ \ facilities - Max\ \ product\ \ duplicates}{Num\ \ facilities}.$$

Since the range of *Max product duplicates* is zero to *Num facilities* for each firm, product heterogeneity is bounded between zero and one where zero indicates homogeneity (i.e., $Max\ \ product\ \ duplicates = Num\ \ facilities$) and one indicates the largest degree of heterogeneity (i.e., $Max\ \ product\ \ duplicates = 0$).

Regulatory heterogeneity

This measure is formed using an analogous three step method. However, we are interested in the values of the sequence of five variables: *CERCLIS, ICIS, PCS, RCRA, TRI*. If *Max* *regulatory duplicates* represent the maximum value of *Regulatory duplicates* (where this measure is formed as described above), then

$$\textit{Regulatory heterogeneity} = \frac{\textit{Num facilities} - \textit{Max regulatory duplicates}}{\textit{Num facilities}}.$$

Size heterogeneity

This measure uses only one variable, rather than a sequence of variables as the other two measures so forming it is more straightforward. We use *Employees*, the number of employees at the facility, to create a categorical variable, *Cat employees* where

$$\textit{Cat employees} = \begin{cases} 1 \textit{ if } \textit{Employees} \le 10 \\ 2 \textit{ if } 10 < \textit{Employees} \le 100 \\ 3 \textit{ if } 100 < \textit{Employees} \le 500 \\ 4 \textit{ if } 500 < \textit{Employees} \end{cases}$$

We then create a variable for each facility, *Employee duplicates*, formed using the same method as described above, based on the values of *Cat employees* at the facilities owned by a given parent company. Letting *Max employee duplicates* represent the maximum value of *Employee duplicates* among facilities owned by a parent company, we form *Size heterogeneity* using the same technique as with the other two heterogeneity measures:

$$\textit{Size heterogeneity} = \frac{\textit{Num facilities} - \textit{Max employee duplicates}}{\textit{Num facilities}}$$