07-023

Resolving Information Asymmetries in Markets: The Role of Certified Management Programs

Michael W. Toffel
Harvard Business School
Soldiers Field
Boston, MA 02163
(617) 384-8043
mtoffel@hbs.edu

February 1, 2006

Resolving Information Asymmetries In Markets: The Role Of Certified Management Programs

Michael W. Toffel[*]
Harvard Business School
Soldiers Field
Boston, MA 02163
(617) 384-8043
mtoffel@hbs.edu

February 1, 2006

Abstract

Firms and regulators are increasingly relying on voluntary mechanisms to signal and infer quality of difficult-to-observe management practices. Prior evaluations of voluntary management programs have focused on those that lack verification mechanisms and have found little evidence that they legitimately distinguish adopters as having superior management practices or performance. In this paper, I conduct one of the first evaluations to determine whether a voluntary management program that features an independent verification mechanism is achieving its ultimate objectives. Using a sample of thousands of manufacturing facilities across the United States, I find evidence that the ISO 14001 Environmental Management System Standard has attracted companies with superior environmental performance. After developing quasi-control groups using propensity score matching, I also find that adopters subsequently improve their environmental performance. These results suggest that robust verification mechanisms such as independent certification may be necessary for voluntary management programs to mitigate information asymmetries surrounding management practices. Implications are discussed for the industry-associations, government agencies, and the non-governmental organizations that design these programs, the companies that are investing resources to adopt them, and those that are relying on them to infer the quality of management practices.

Keywords: auditing, monitoring, supply chain, evaluation, voluntary programs, environmental management, propensity score matching, ISO 14001

JEL Classification Codes: D820, L150, M110, M140, Q530

INTRODUCTION

Company management practices that govern quality, financial, environmental, and labor issues are of increasing importance to many customers. In part, this is due to potential negative spillovers. For example, a supplier's inconsistent production process can reduce the quality of the buyer's final goods, which can lead to higher costs and damaged reputations (Krueger & Mas, 2004). In addition, suppliers who poorly manage regulatory compliance may be more likely to be shut down by regulatory inspectors, which can impose costly business interruption on their customers (Medina-Ross, 2002). Furthermore, the media and end-consumers are increasingly holding firms accountable for their suppliers' labor and environmental management practices (e.g., CAFOD, 2004; McNeil, 2004; Strom, 1996). Beyond seeking to mitigate risks to their end-product quality, brand image, and corporate reputation, some companies seek suppliers with superior environmental practices to promote their environmental image or objectives (NEETF, 2001; Walton, Handfield, & Melnyk, 1998).

Despite their importance, suppliers' management practices remain quite difficult for buyers to observe, which presents an information asymmetry problem. Conventional solutions have included branding strategies and vertical integration (Alchian & Demsetz, 1972; Williamson, 1985). In addition, mandatory disclosure regulations have emerged in a few domains where difficult-to-observe management practices can directly affect public health, such as in hospital care and restaurant hygiene (Jin & Leslie, 2003; Mukamel & Mushlin, 1998). More recently, hundreds of thousands of organizations have turned to a burgeoning number of voluntary management programs in an attempt to reduce information asymmetries in business-to-business transactions. While these programs call for participants to adopt particular management practices, procedures, and frameworks, they do not impose any performance requirements. Furthermore, these programs vary dramatically in their verification requirements: most allow any organization to simply self-declare its participation, while a few require periodic audits by an independent monitor to verify that participants have fully implemented the program's procedural requirements.

The absence of performance standards and, in most cases, verification requirements has led critics to dismiss voluntary management programs as marketing gimmicks or "greenwash". Indeed, prior evaluations of voluntary management programs have found little evidence that better-than-average performers are more likely to adopt them, or that participating in these programs is associated with performance improvement. However, because the programs that have been evaluated lack robust verification mechanisms, it remains unclear whether participants failed to implement their procedural requirements, or whether the programs' focus on processes to the exclusion of outcomes is simply inadequate to attract superior performers or to elicit performance improvement.

I address this issue by evaluating the ISO 14001 Environmental Management System Standard, a voluntary management program that requires periodic independent monitoring to ensure that adopters fully conform to its requirements. I find that ISO 14001 attracted manufacturers with greater facility-wide toxic air emissions, and that these emissions were imposing greater health hazard on their communities than non-adopters. At the same time, however, these adopting facilities were less pollution-intensive. In other words, they had lower emissions (both in terms of pounds and health hazard) than non-adopters once differences in facility size were accounted for. To investigate whether adoption is associated with subsequent performance improvement, I create two matched samples and employ a difference-in-differences analysis. I find that adopters subsequently reduced both their facility-wide emissions and their pollution intensity more than their matched control groups.

Overall, these findings suggest that a voluntary management program with robust verification can indeed be a useful mechanism to differentiate organizations as having superior management practices, both through a selection effect and a treatment effect. This finding is in sharp contrast with prior evaluations of voluntary management programs that had weak or no verification mechanisms. This disparity suggests that verification requirements may be a vital feature to ensure that voluntary management programs legitimately distinguish participants.

VOLUNTARY MANAGEMENT PROGRAMS

Voluntary management programs share a common focus on production processes rather than end-results. These programs have been initiated by a wide variety of organizations. Industry-association programs include Responsible Care (chemicals), Sustainable Forestry Initiative (forest products), and Sustainable Slopes (skiing). Non-governmental organization (NGOs) programs include Social Accountability International's SA 8000 and the Forest Stewardship Council's Forest Management Certification. In addition, international and national standards bodies have issued process management standards such as the International Organization for Standardization's ISO 9000 Quality Management System Standard and ISO 14001 Environmental Management System Standard, and the British Standards Institution's OHSAS 18001 Occupational Health and Safety Management Systems Specification. National and supranational governmental bodies have also sponsored voluntary management programs, such as the US Occupational Safety and Health Administration's (OSHA) Voluntary Protection Program, the European Union's Eco-Management and Audit Scheme (EMAS), and the United Nations Global Compact. Each of these programs seeks to differentiate participants as possessing superior management practices related to quality, the environment, workers, or human rights.

Prior Evaluations of Voluntary Management Programs

Despite their intentions, there remains great uncertainty about whether voluntary management programs actually distinguish adopters as having superior management practices (Melnyk, Sroufe, Calantone, & Montabon, 2002; O'Rourke, 2003). For these programs to legitimately distinguish participants, they must either disproportionately attract participants with superior *ex ante* performance (a positive selection effect), or participants must subsequently improve their performance faster than non-participants (a treatment effect), or both. The few robust evaluations of voluntary management programs have found little evidence of a positive selection effect. Instead, there is more evidence that companies with inferior performance are more likely to participate (King & Lenox, 2000; Lenox & Nash, 2003; Naimon, Shastri, & Sten, 1997; Rivera & de Leon, 2004). The few studies that evaluated the effects of voluntary management programs found no evidence that participants subsequently performed better than non-participants (King & Lenox, 2000; Naimon et al., 1997; Rivera, de Leon, & Koerber, 2005). All of these studies, however, focused on voluntary management programs that merely rely only on an "honor system" to ensure that participants actually implement their requirements—a vulnerability to which many have attributed the studies' adverse findings (e.g., Gunningham, 1995; King & Lenox, 2000; Lenox & Nash, 2003; Rivera & de Leon, 2004).

A few studies have begun examining voluntary management programs with more robust verification and enforcement mechanisms. For example, Lenox & Nash (2003) found some evidence that a credible threat of expulsion from the forestry trade association enabled its voluntary environmental management program to attract a disproportionate number of participants with *ex ante* superior environmental performance. However, this enforcement mechanism is not robust, since they found no evidence that such a threat by the chemical industry association enabled its program to disproportionately attract superior performers. Evaluations of the ISO 9000 Quality Management System Standard, one of the few voluntary management programs where certification requires periodic verification by an independent auditor, have primarily measured performance using financial indicators (e.g., Corbett, Montes-Sancho, & Kirsch, 2005; Easton & Jarrell, 1998; Heras, Dick, & Casadesús, 2002; Terziovski, Samson, & Dow, 1997), which are several steps removed from the quality-assurance objectives of the standard. The two studies that investigated the effect of ISO 9000 on waste reduction, a performance indicator more directly related the standard's process quality-assurance objectives, yielded mixed results (King & Lenox, 2001; Terlaak & King, Forthcoming).

Thus it remains an open question whether a voluntary management program with a robust verification mechanism can legitimately distinguish adopters via a selection effect and/or a treatment effect. I address this question by examining certification to the ISO 14001 Environmental Management System Standard, which like ISO 9000 requires periodic verification of conformance by an independent auditor. I focus on ISO 14001 because this standard offers a unique combination: its adoption by tens of thousands of plants around the world indicates that many are placing their faith in its signaling or improvement potential, and data are available to assess its ultimate objective of distinguishing adopters based on superior environmental performance.

The ISO 14001 Environmental Management System Standard

The International Organization for Standardization's ISO 14001 Environmental Management System Standard is an international management standard that provides a comprehensive framework for conducting environmental management activities. Established in 1996, ISO 14001 requires organizations to: develop an environmental policy with a commitment to continuous improvement; identify all of its environmental aspects and then prioritize them based on the significance of their environmental impacts; establish environmental objectives and targets; develop work procedures to control environmental aspects; train employees on these procedures; demonstrate a commitment to comply with environmental laws and regulations; conduct self-assessment audits; and periodically review the management system. ISO 14001's requirements are largely based on the best management practices of the multinational corporations such as IBM that helped draft the standard. ISO 14001 was designed to be sufficiently flexible so that any type of organization could adopt the standard. By the end of 2004, over 90,564 organizations across 127 countries had adopted ISO 14001, including more than 4,759 in the US (International Organization for Standardization, 2004).

The standard was created under the premise that organizations that create or strengthen their environmental management system in accordance with ISO 14001 will benefit by reducing their operating costs and environmental impact, enhancing their corporate image, and experiencing fewer and less severe accidents and regulatory violations. Like other voluntary management programs, ISO 14001 contains no performance requirements.[1] Independent third-party auditors, who must meet ISO auditing requirements, are meant to bring "rigor and discipline" to ISO 14001 adoption (NAPA, 2001: 12). ISO authorizes one organization in each country to ensure the credibility of the auditing process by accrediting certifiers and providing training to ensure that audits are performed consistently and competently. Auditing seeks to verify that an organization's objectives, targets, and procedures are consistent with its commitments to

[1] The nearest the standard comes to discussing environmental performance is its requirement that an organization's environmental policy must include a commitment to continual improvement. However, this refers to improving the environmental management system, not environmental performance.

continual improvement and pollution prevention. Once an auditor is satisfied that an organization has fully met the standard's requirements, the auditor certifies the organization to ISO 14001. After its initial certification, an adopter must demonstrate that its environmental management system continues to meet ISO 14001's requirements in annual surveillance assessments and full re-assessments every three years (IAF, 2003).

Prior Evaluations of ISO 14001[2]

ISO 14001 has gained a favorable reputation among some due to a plethora of anecdotal evidence that describes how adopting the standard has led companies to reduce their energy use, generate less waste, experience fewer accidents and spills, and improve regulatory compliance (e.g., Burglund, 1999; Chin & Pun, 1999; Fielding, 1999; Smith & Feldman, 2003; Toffel, 2000). On the other hand, environmental regulators and activists have focused on the standard's lack of performance requirements as a significant impediment to claims that certification necessarily implies a significant and praiseworthy achievement (Courville, 2003: 288; Yiridoe, Clark, Marett, Gordon, & Duinker, 2003: 450). In particular, ISO 14001's emphasis on documentation has been the subject of much criticism, and it is possible that many facilities implementing such procedural initiatives are distracted from other tasks that might elicit better results.

Despite widespread anecdotal evidence, little systematic research has rigorously evaluated the performance implications associated with ISO 14001 certification (Delmas, 2004; Melnyk et al., 2002; Redinger & Levine, 1998; Rondinelli & Vastag, 2000). Some studies that have examined the effects of ISO 14001 have focused only on adopters (e.g., Szymanski & Tiwari, 2004), while others have compared adopters to non-adopters only before or after adoption (e.g., Hillary & Thorsen, 1999; Matthews, 2001). A recent study that compared the environmental performance of adopters to non-adopters over time concluded that ISO 14001 adopters reduced facility-wide pollution more than non-adopters (Potoski & Prakash, Forthcoming). However that study did not attribute the performance difference to a selection or treatment effect, and empirical evidence that calls into question the study's identifying assumption may

[2] For a comprehensive literature review of prior evaluations of ISO 14001, see Toffel (2005).

have resulted in the effect being overestimated.[3] Thus, while that study provides an important first step in examining whether ISO 14001 is associated with performance improvement, several methodological concerns suggest caution in interpreting its results.

My analysis improves upon the prior evaluations of ISO 14001 in several ways. I compare adopters to non-adopters over time in a manner that clearly distinguishes a selection effect from a treatment effect. I develop quasi-control groups through propensity score matching, which avoids controversial assumptions often required to support the choice of a valid instrumental variable. In addition, I not only assess facility-wide emissions and their associated health hazard, I also investigate the relationship between ISO 14001 and pollution intensity by controlling for facility size and production changes.

THEORY

There are several reasons why ISO 14001 may legitimately distinguish participants as possessing superior management practices, both by disproportionately attracting superior facilities (a selection effect) and by leading facilities to enhance their performance (a treatment effect). This distinction between selection and treatment effects is important to buyers and regulators contemplating *when* they should consider participants as superior to non-participants. If participants' superior performance derives only from a treatment effect, buyers and regulators would need to allow a time lag to pass after adoption to enable participants to develop superior performance. If ISO 14001 distinguishes participants via a selection effect, no such delay would be needed because adopters would already be exhibiting superior *ex ante* performance.

[3] In particular, the sampling strategy and the instrumental variable choice suggest caution in interpreting the results. In constructing their sample, the authors compared average emissions during 1996-1997 to 2000-2001, but coded facilities as "adopters" if they adopted anytime by 2001. Given that (1) there has been an overall decline in many facilities' toxic emissions since the early 1990s, and (2) every year since 1996 has seen a growing number of facilities adopt ISO 14001, the adopters in their sample were likely skewed toward the end of the sample period, when emissions of *all* facilities were lower. As such, the research design precludes identifying whether the performance differences between adopters and non-adopters during this period were the result of a selection effect, a treatment effect, or both. In addition, the study uses environmental regulatory compliance as an instrument for adoption, which is valid only under the identifying assumption that a facility's compliance record does not directly affect emissions. Since others have shown that being cited with a compliance violation directly leads to behavioral changes that reduce pollution and improve worker safety (Earnhart, 2004; Gray & Scholz, 1993; Kniesner & Leeth, 2004), this assumption is questionable. If the identifying assumption is invalid, estimations based on this instrument would result in an upward bias of (and thus overstating) the effect of adoption on performance.

Why Adopters May Already be Superior Performers: Signaling

The comprehensiveness of a firm's environmental management practices and its environmental performance are largely unobservable to outsiders. Nonetheless, a growing number of companies including Ford Motor Company, Toyota, Johnson Controls, and Bristol-Myers Squibb are including environmental management as a key criterion in selecting suppliers, and are using the ISO 14001 standard as an indicator of superior environmental management practices. More generally, empirical research suggests that ISO 14001 certification is substituting for firms' own monitoring of their suppliers' practices, especially when suppliers' production processes are particularly costly for buyers to observe (Christmann & Taylor, 2001; King, Lenox, & Terlaak, Forthcoming).

ISO 14001 certification can be a legitimate indicator of a supplier's possessing superior management practices if the standard disproportionately attracts facilities that already exhibit superior environmental performance. Spence's (1973) classic signaling model describes how information asymmetry problems between uninformed buyers and informed suppliers can be resolved. In his model, employers would like to hire highly productive employees but can only observe job applicants' educational attainment. The model's key insight is that for educational attainment to be a credible signal for ability, the cost of signaling (pursuing education) must be cheaper for more productive individuals.

For ISO 14001 adoption to provide a credible signal of costly-to-observe environmental management practices, the difference in the cost of sending the signal (adopting the standard) must be sufficiently large to make it worthwhile (profitable) to facilities with superior environmental management systems but not to others. Is this plausible? The most relevant costs to consider are the internal "soft" costs such as preparing documentation and conducting training, since these represent the largest proportion of total costs required to implement ISO 14001.[4] The very fact that the ISO 14001 standard was based on best management practices implies that organizations with comprehensive environmental management

[4] In contrast, fees associated with hiring a third-party certifier and actual registration are relatively small, typically ranging from $5,000 to $20,000 per facility (Dahlström, Howes, Leinster, & Skea, 2003; Kolk, 2000; Prakash, 2000), and are sometimes offset by government subsidies. More importantly, these fees are typically based on a facility's operational complexity and size, not on the status of its EMS before it began implementing ISO 14001.

systems should incur lower adjustment costs to become certified. Indeed, empirical evidence suggests that environmental management system adoption costs are lower among organizations with more management system experience (e.g., those that had already implemented Total Quality Management and Just-in-Time inventory systems) and that had already implemented pollution prevention practices (Darnall & Edwards, Forthcoming).

In contrast, if poor environmental performance is due to the absence of a systematic environmental management system, then such firms seeking to adopt ISO 14001 would have to invest significantly more to build an ISO 14001-compliant environmental management system from scratch. This supports the plausibility of the fundamental assumption underlying a signaling story: that adoption is cheaper and may only be worthwhile for those with superior environmental management practices. Given that these management practices are positively correlated with environmental performance (King et al., Forthcoming), companies with better environmental performance should be more likely to adopt ISO 14001.

Hypothesis 1. ***Facilities with better environmental performance are more likely to adopt the ISO 14001 standard.***

The Potential of ISO 14001 to Improve Environmental Performance

The process of developing an environmental management system sophisticated enough to meet the ISO 14001 standard can require and facilitate organizational learning. As describes below, this process often entails developing new routines and skills and creates new knowledge networks that can introduce innovative improvement opportunities.

ISO 14001 requires organizations to develop a comprehensive inventory of all the ways in which their activities may impact the environment (their "environmental aspects"), to rank their significance in terms of potential environmental impact, and to develop procedures to control aspects with highly significant environmental impacts. Meeting these requirements is often the first time organizations pursue such a comprehensive approach and explicitly prioritize environmental management efforts based on the significance of environmental impacts. As such, this process may enable organizations to better target their management efforts to improve their environmental performance. The standard also requires periodic management review of the environmental management system, which encourages companies to reconsider their priorities when their activities or processes change.

Several aspects of implementing an ISO 14001-compliant environmental management system and operating according to the standard may spur new sources of innovation. First, the standard requires organizations to train all employees whose work may create significant impacts on the environment. Such training is typically provided to more than half the facility's employees, and some companies train over 95% of their employees (Corbett & Luca, 2002). This training often enables employees to better identify pollution prevention opportunities and empowers them to offer recommendations (Darnall, Gallagher, Andrews, & Amaral, 2000; Rondinelli & Vastag, 2000; Toffel, 2000). Second, cross-functional teams are commonly used to implement ISO 14001. Such teams can foster systems thinking and shared objectives to more efficiently transfer information and tacit know-how within organizations (Kogut & Zander, 1992), including new ideas to prevent waste and pollution across various production process stages (King, 1995, 1999; Russo & Fouts, 1997). Third, adopters often begin looking to other organizations, such as competitors, buyers, suppliers, and external consultants, for ideas to reduce their environmental impacts (Toffel, Hill, & McElhaney, 2003). ISO 14001 certification also can put adopters on the "radar screen" of NGOs looking for companies with whom to collaborate, facilitating preferential access to the expertise these NGOs possess (Rondinelli & London, 2003).

Finally, ISO 14001 can help organizations maintain a higher priority for their environmental management tasks. For example, the standard requires adopters to conduct periodic internal audits to ensure that employees are complying with documented environmental procedures. It also requires the organization to conduct initiatives to achieve the documented objectives and targets, and senior management to periodically review the environmental management system. Adopters often embed their ISO 14001 certification in their corporate or brand image, which creates additional incentives to ensure they continue to meet the standard's requirements throughout their annual external surveillance audits.

Hypothesis 2. ***Facilities that adopt ISO 14001 will improve their environmental performance more than non-adopters.***

Institutional scholars have suggested that organizational motivations to adopt formal programs vary over time. In particular, while early adopters may adopt these programs to improve organizational practices, facilities adopting in subsequent years may be motivated to maintain legitimacy (Tolbert & Zucker, 1983). This trend has been found among adopters of quality management programs, financial reporting standards, and long-term incentive plans (Mezias, 1990; Westphal, Gulati, & Shortell, 1997; Westphal & Zajac, 1994). This can lead early adopters to make substantive changes to implement the program, but later adopters to only symbolically adopt the program. For example, early adopters of corporate affirmative action programs were more likely to implement substantive affirmative action offices, while later adopters were more likely to take a more symbolic approach of implementing rules (Edelman, 1992). Because this temporal pattern of motivations has been suggested in the case of ISO 14001 (Jiang & Bansal, 2003), early adopters of ISO 14001 may exert more effort to implement robust environmental management systems than late adopters. As such, later adopters may do "just enough" to meet the standard's requirements to obtain certification. In this case, early adopters will derive more performance improvement from adopting ISO 14001 than later adopters.

Hypothesis 3. ***Early adopters of ISO 14001 will improve their environmental performance more than late-adopters.***

SAMPLE AND MEASURES

To test these hypotheses, I assembled a comprehensive dataset using 1991-2003 data obtained from several publicly available government databases, several commercial databases, journal articles, and reports. I also filed several Freedom of Information Act requests of the United States Environmental Protection Agency (US EPA). Specific data sources are described in the measures section below.

Sample

My sample includes manufacturing facilities within the United States that have reported emissions of toxic chemicals to the US EPA Toxic Release Inventory (TRI) program. This program includes facilities that manufacture, import, process, or use any of the listed substances in amounts greater than threshold quantities (typically 10,000 or 25,000 pounds) and have at least 10 full-time employees (US EPA, 1999). I focus on the five industries with the most ISO 14001 adopters as of 2001: chemicals, fabricated metal products, industrial machinery and equipment, electrical and electronic equipment, and transportation equipment (SIC Codes 28 and 34-37) (McGraw Hill, 2001). Facility details were obtained from the US EPA's TRI website and the US EPA's Risk-Screening Environmental Indicators 2.1 model.[5] To facilitate a closer comparison between ISO 14001 adopters and non-adopters, I restricted the sample to sub-industries (4-digit SIC Codes) that included at least one adopter and one non-adopter.

Measures

Environmental performance. I measure environmental performance using several metrics based on toxic emissions data reported annually to the US EPA TRI program from 1991 to 2003, the latest year for which data are available. A rare source of uniformly reported, legally mandated facility-level environmental performance data, the TRI dataset has been widely employed in the management literature to measure companies' environmental performance (Gerde & Logsdon, 2001; Toffel & Marshall, 2004). The first measure, *pounds of emissions*, mimics how the US EPA and the media rank the "dirtiness" of TRI reporters. I gather data on the air emissions of the "core group" of TRI chemicals—those that have been required every year and with a consistent reporting threshold—and create annual facility totals. Because these totals were highly skewed, I log the annual sums (after adding 1 to accommodate zero values). I also estimate the health hazard posed by these emissions. To accommodate the enormous variation in the toxicity of TRI chemicals, I employ chemical-specific toxicity weights pertaining to inhalation exposure from the US EPA's Risk-Screening Environmental Indicators model (US EPA, 2002), as recommended by Toffel & Marshall (2004). To calculate *health hazard*, I multiplied the pounds of each core chemical emitted to air by its inhalation toxicity factor, took the sum of these products, and log the sum (after adding 1).

[5] US EPA provides TRI data at http://www.epa.gov/tri and the Risk-Screening Environmental Indicators model at http://www.epa.gov/opptintr/rsei/

Production and facility size. While there is a strong conceptual link between emissions and facility production volumes, the latter are typically considered proprietary and are difficult to acquire. As a proxy for production, I create a *production index* based on two variables: (1) facility employment, obtained from Dun & Bradstreet for a single year for each facility in my sample, which I refer to as that facility's baseline year, and (2) annual facility production ratios, which is the ratio of production volume in the current year to that of the prior year, obtained from the TRI dataset (King & Lenox, 2000). I calculate the production index in four steps: (1) I set production index equal to facility employment for the baseline year; (2) for each year following the baseline year, I calculate the production index by multiplying the prior year's production index by the current year's production ratio; (3) for each year preceding the baseline year, I calculate the production index by dividing the subsequent year's production index by the subsequent year's production ratio; and (4) I log the result.

ISO 14001 certification. I obtained the identity of ISO 14001 adopters and their certification year from the *ISO 14001 Registered Company Directory North America* (QSU, 2002b) and various state environmental regulator websites. Based on this data, I create several dummy variables. *Certification year* is coded one for adopters only in the year they became certified. *Post-certification* is coded one for adopters starting the year they became certified. *Early adopters' post-certification* is coded one starting the certification year for adopters initially certified in 1996-1999. *Late adopters' post-certification* is coded one starting the certification year for adopters initially certified in 2000-2002.

EMPIRICAL ANALYSIS

Selection Analysis

The selection analysis seeks to discern whether adopters' and non-adopters' environmental performance differed prior to adoption (Hypothesis 1). I begin by assessing whether facility-wide pounds of emissions is associated with ISO 14001 adoption. I then examine whether pollution intensity affects facilities' propensity to adopt the standard by including the production index to control for facility size and production changes. I repeat these two steps by substituting health hazards for pounds of emissions. I employ a probit model where the dependent variable is a dummy coded one the year an adopter becomes certified. The key variables to detect a selection effect are lagged values of each outcome measure.[6] In each specification, I include industry dummies (3-digit SIC Codes) and year dummies. The year dummies control for events in a given year that might impact the emissions of all facilities, such as changes in federal regulations, federal enforcement priorities, or the introduction of new technologies. I include 1991-2003 data, but drop adopters after their certification year to avoid confounding the selection analysis with any potential effects of certification.

Table 2 presents the probit results. I report standard errors clustered by facility to account for non-independence among observations from the same facility. Columns 1 and 2 report the models that employ lagged facility-wide performance metrics. Each of these models suggest that "dirtier" facilities adopt ISO 14001: facilities that emitted more pounds of toxic chemicals and that imposed greater health hazards on their communities were more likely to adopt ISO 14001 ($p<0.001$). The probit coefficients can be interpreted as follows: compared to a baseline propensity to adoption evaluated at the mean of all variables, a one standard deviation increase in emissions (health hazard) corresponds to a 28% (32%) increase in the probability of adoption.

The results of the pollution intensity models (Columns 3 and 4) reveal an important distinction, however. After controlling for facility size, these results indicate that facilities with lower pollution intensity were more likely to adopt ($p<0.001$), both in terms of pounds of emissions (Column 3) and health hazard (Column 4). Holding facility size (and all else) constant, a one standard deviation *decrease* in emissions (health hazard) corresponds to a 15% (12%) increase in the probability of adoption. In summary, these results suggest that less pollution-intensive facilities are more likely to adopt ISO 14001, but that adopters are larger and therefore emit more facility-wide pollution prior to adoption, compared to other TRI-reporting facilities in their industry.

[6] For each performance metric, I use the average of the 1- and 2-year lags because this prevents many observations from dropping out of the sample, which contains many missing values.

Treatment Analysis

I employ a difference-in-differences approach to evaluate whether the adoption of ISO 14001 influences environmental performance (Hypotheses 2 and 3). This approach uses a control group's performance during the post-treatment period as the counterfactual for how the treatment group would have performed if it had not received the treatment. Unbiased estimates from a difference-in-differences approach requires two assumptions about the similarity between the treatment and control groups: (1) that both groups would respond similarly to treatment, and (2) that the control group's performance trend serves as the counterfactual of the treatment group (i.e., if the adopters had not adopted, their performance would have mimicked the control group's). The selection analysis conducted above showed that adopters' and non-adopters' environmental performance differed during the pre-adoption period. The two groups are likely to differ in other ways as well. For example, larger facilities may have more resources to devote to the adoption of both ISO 14001 and pollution control technologies that affect environmental performance. To eliminate these potential sources of bias, I used propensity score matching to identify a quasi-control group of non-adopters with similar adoption determinants and pre-period outcome trends to the adopters.

Propensity score matching. Matching is a widely used approach to construct a quasi-control group based on similar characteristics as the treatment group (Heckman, Ichimura, & Todd, 1998). Matching on the propensity score—the probability of receiving the treatment conditional on covariates—is as valid as matching on a series of individual covariates (Rosenbaum & Rubin, 1983). The identifying assumption is that the assignment to the treatment group is associated only with observable "pre-period" variables, and that all remaining variation across the groups is random. This assumption is often referred to as the "ignorable treatment assignment" or "selection on observables."

When used to evaluate job training programs, propensity score matching methods have performed well in replicating the results of randomized experiments under three conditions: (1) the same data sources are used for participants and non-participants; (2) an extensive set of covariates are employed in the program-participation model that is used to estimate propensity scores; and (3) participants are matched to non-participants in the same local labor market (Smith & Todd, 2005). In addition, Heckman, Ichimura, & Todd (1997) note that substantial bias can result if: (4) controls are included whose propensity scores are off the support of the participants' propensity scores; (5) the distributions of the participants and non-participants' propensity scores differ; or (6) unobservable factors influence both participation and outcomes.

I address these six potential sources of bias as follows. First, I use the identical data sources for all facilities (participants and non-participants). Second, I gather an extensive set of adoption covariates based on a comprehensive literature review. Third, I ensure that participants and non-participants operate within the same markets by matching participants to non-participants within the same industry. I address the fourth and fifth concerns by implementing nearest neighbor matching with a "caliper" restriction to preclude matching beyond a fixed threshold, and by excluding observations whose propensity scores are off the common support. The sixth concern addresses selection on unobservables. In the context of ISO 14001, it is possible that facilities with an "environmentalist" culture (which I do not observe in my data) may have managers who are both more likely to insist upon strong environmental performance and be more prone to adopt ISO 14001. To the extent that such unobserved factors are stable over time during the sample period, I address this concern by including facility fixed effects.[7]

I implement propensity score matching in three steps. First, I generate propensity scores by estimating a logit model for adoption status during 1996-2003, omitting adopters after their adoption year. I include many potential adoption determinants identified through an extensive literature review on the adoption of ISO 14001 and other environmental management programs.[8] Table 1 describes these covariates and provides summary statistics. I also included lagged outcome levels (average of prior two years) and lagged 4-year trends to increase the likelihood that the matched control group's pre-period performance trends would be similar to the adoption group's (Barbera & Lyona, 1996; Dehejia & Wahba,

[7] Instrumental variable models have been employed in some program evaluations that examined the performance implications of voluntary management programs (e.g., Khanna & Damon, 1999; Potoski & Prakash, Forthcoming; Rivera & de Leon, 2004; Welch, Mazur, & Bretschneider, 2000). I failed to identify a credible instrumental variable—one that is correlated with the adoption decision but has no independent influence on performance—that would enable me to use this alternative method.
[8] Details on this comprehensive literature review are provided in the Appendix. Also, see Toffel (2005).

1999; Eichler & Lechner, 2002).[9] Predicted values from the logit model constitute propensity score estimations.

I created three matched samples for the treatment analysis. Sample A is a baseline sample that includes the 860 adopters and 11217 non-adopters that are in those sub-industries (4-digit SIC Codes) that include at least one adopter, and that include at least one pre-1996 and one post-1996 observation. I implement nearest-neighbor matching based on the propensity scores to construct other matched samples (Leuven & Sianesi, 2003). Sample B is formed by identifying, for each adopter during its certification year, up to five non-adopters with the closest propensity scores that are within a fixed "caliper limit" of 0.001. The latter restriction results in fewer than five matches when fewer close matches are available. This process yielded a matched sample of 509 adopters and 1808 non-adopters, for an average of 3.6 matches per adopter.

I assessed the similarity of the non-adopters and adopters in the matched control group in several ways. Since the pseudo-R^2 statistic indicates how well the regressors explain the participation probability, the matching process should result in a substantial reduction in the pseudo-R^2 value if the logit model is re-estimated on just the matched sample (Sianesi, 2004). Indeed, the pseudo-R^2 declines from 0.17 (Sample A) to 0.08 (Sample B). I also used t-tests to compare the means of the adoption covariates. While the adopters and non-adopters in Sample A had statistically significant differences for 11 of the 16 covariates, only two covariates differed significantly at the 10% level between the matched adopters and non-adopters in Sample B.[10]

I constructed a second nearest-neighbor matched sample by first imposing additional restrictions on potential matches. To construct Sample C, I limited the potential candidates for an adopter's five nearest neighbors to non-adopters that shared its industry (2-digit SIC Code), size (employment) range, and prior

---

[9] The logit results of the full adoption model are provided in Table A-1 of the Appendix. Briefly, adoption is more likely in facilities that provide evidence of a formal environmental management system, participated in the US EPA 33/50 program, had already adopted the ISO 9000 Quality Management System standard, had released chemicals in their effluent to publicly-owned wastewater treatment facilities, were located in states with higher compliance costs, and were larger (more employees).

[10] Table A-2 in the Appendix presents the following: t-tests of equal means across adopters and non-adopters, standardized bias before and after matching, and t-tests that compare the average performance trends of adopters and non-adopters in the years before the match.

4-year emissions trend range.[11] Because of these more restrictive pre-conditions, I used a wider caliper of 0.2 to exclude nearest neighbors that were "too distant". This resulted in a matched sample of 465 adopters and 1622 non-adopters, for an average of 3.5 matches per adopter. T-tests indicate that this matching technique also balanced the covariates: the mean values of all but one covariate were indistinguishable across the matched non-adopters and adopters at the 10% level. After re-estimating the logit model on this matched sample, the pseudo-R^2 value of 0.10 represents a substantial reduction from the original 0.17 value, providing further evidence of a successful matching process. Nearest-neighbor matching substantially reduced standardized bias (Rosenbaum & Rubin, 1985) from an average across all covariates of -15% (Sample A) to 0.5% (Sample B) and -1% (Sample C) after matching. In addition, kernel density plots confirmed that the adopters and non-adopters in the matched groups had very similar distributions of the covariates.

Finally, I assessed whether non-adopters and adopters experienced similar performance trends prior to the match year. If the two groups experienced similar pre-match performance trends, this would substantially bolster the plausibility of the difference-in-difference identifying assumption that in the absence of treatment, the adopters' subsequent performance would have mimicked the control group's. T-test results confirmed that the two groups' pre-adoption performance trends during the 4-years prior to the match year were statistically indistinguishable, across each of the three comparison groups.[12] As a whole, these findings bolster the plausibility of the identifying assumption of the difference-in-differences method: that the two matched control groups serve as valid counterfactuals for the treatment group.

Model specifications. For Sample A, which includes all facilities in the sub-industries (4-digit SIC Codes) that include at least one adopter and one non-adopter, I estimate the following equation to estimate the effect of ISO 14001 adoption:

$$y_{it} = \beta_1 D_{it} + \beta_{2t} \lambda_t + \beta_3 PI_{it} + \alpha_i + \varepsilon_{it} \qquad (1)$$

[11] Size categories were created for facilities with 0-49, 50-124, 125-499, 500-999, 1000-9999, and 10,000+ employees. Emission trends were calculated as the average of 1 and 2 year lagged emissions levels minus the average of 3 and 4 year lags, divided by the sum of these two averages. Emission trends were divided into four categories based approximately on quintiles: –1 to –0.2, –0.2 to 0, 0 to 0.2, and 0.2 to 1 (upper bounds not inclusive, except the latter category).

[12] See the bottom three rows of Table A-2 in the Appendix.

where y_{it} is the outcome variable for facility i in year t, D_{it} is the post-certification dummy coded 1 for adopters starting in the certification year, λ_t represents year fixed effects, and α_i is the facility fixed-effect. The production index, PI_{it}, is initially omitted to estimate facility-wide emissions and subsequently included to estimate pollution intensity.

For the nearest-neighbor matched samples (Samples B and C), I employ a more flexible specification:

$$y_{it} = \beta_1 D_{it} + \beta_{2t}\ \tau_{gt} + \beta_{3g}\ \tau_{gt} \times g_i + \beta_4\ PI_{it} + \alpha_i + \varepsilon_{it} \qquad (2)$$

where g_i is a complete set "matched group" fixed effects. τ_{gt} is a "counter" for each matched group coded sequentially: "0" in the 4th year before group *g*'s match year, "1" in the 3rd year before group *g*'s match year, through "8" in the 4th year after group *g*'s match year. The other variables are defined as in the previous equation. For Sample B, each "matched group" includes all adopters whose propensity scores are within the caliper limit, and all of their respective matches. For Sample C, each "matched group" includes all adopters that share the same industry, size category, and pre-period outcome trend category, and whose propensity scores are within the caliper limit—as well as all of their respective matches. Including interactions of the matched group dummies and counters allows each matched group to have its own temporal trend. Again, the production index (PI_{it}) is omitted when estimating facility-wide emissions, and included when estimating pollution intensity models.

To detect whether treatment effects varied between early adopters and late adopters, I substitute the post-certification term (D_{it}) with the *early adopters' post-certification* and *late adopters' post-certification* terms. As described earlier, the former is a post-certification dummy for early adopters (1996-1999) and the latter is a post-certification dummy for late adopters (2000-2002). To assess whether the treatment effect differs between early and late adopters, I conduct a Wald test to determine whether the coefficients on these variables are significantly different.

Treatment Analysis Results

Table 3 presents the results across three comparison groups. Prevailing practice among those employing difference-in-difference models is to report OLS standard errors (Bertrand, Duflo, & Mullainathan, 2004). However, failing to account for potential serial correlation risks significantly underestimating standard errors. As such, I test for serial correlation (Drukker, 2003; Wooldridge, 2002: 282-283) and find evidence of serial correlation in each of the 12 models. As such, I report robust standard errors clustered by facility to accommodate an arbitrary autocorrelation process.

The coefficient on the post-certification dummy (β_1) is negative for all 12 models, which indicates that on average adopters subsequently improved their environmental performance compared to each of the three groups of non-adopters across all four performance metrics. However, as described below, the differences between the groups were not always statistically significant.

Column 1 of Table 3 displays that subsequent to certification, adopters' facility-wide emissions declined 39% more than non-adopters' in Sample A during the same period, a statistically significant difference ($p<0.01$). Column 2 indicates that adopters in Sample B subsequently reduced their emissions 30% more than their matched non-adopters did ($p=0.055$). Column 3 indicates that adopters in Sample C subsequently reduced their emissions 11% more than their matched non-adopters did, but this difference is not statistically significant.

Columns 4-6 provide the results of health hazards associated with the facility-wide toxic air emissions. Focusing on the matched groups, adopters in Sample B subsequently reduced the health hazard associated with their facility-wide emissions 45% more than the matched non-adopters ($p=0.054$). Adopters in Sample C reduced the health hazard associated with their facility-wide emissions 31% more than non-adopters, but this is only marginally significant ($p=0.07$ for one-tailed test).

The results also provide evidence that adopters subsequently reduced their pollution intensity compared to non-adopters. Measured in pounds, adopters subsequently reduced their pollution intensity 24% more (Sample B) or 17% more (Sample C) than the matched non-adopters, although the latter difference was only marginally significant. As columns 11-12 indicate, adopters reduced their pollution intensity in terms of health hazard 45% more (Sample B) or 42% more (Sample C) than their matched non-adopters. Each of these differences is statistically significant ($p<0.05$).

As a robustness test, I re-estimated the treatment effects using the alternative environmental performance metrics and empirical specifications from an influential evaluation of an industry-sponsored voluntary management program (King & Lenox, 2000). I used its facility fixed effects specification and estimated adopters' subsequent "absolute improvement", a metric akin to the facility's percent reduction in emissions.[13] The results indicate that subsequent to certification, adopters improved more than non-adopters both in terms of pounds ($p<0.001$) and health hazard, though the latter is only marginally significant ($p<0.11$). I also estimated adopters' subsequent "relative improvement", which compares trends in a facility's emissions relative to others within the same industry (4-digit SIC Code) controlling for facility size.[14] The results provide strong evidence that adopters subsequently improved their relative emissions more than non-adopters, both in terms of pounds and health hazard ($p<0.001$). Overall, these alternative models bolster the results of the main treatment effects analysis by providing additional evidence that ISO 14001 adopters subsequently improved their performance compared to non-adopters.

The models that compared the subsequent performance of early versus late adopters revealed several important distinctions. In nearly all cases, early adopters—but not late adopters—subsequently outperformed non-adopters across all three comparison groups.[15] This finding reveals an important nuance in interpreting the limited statistical significance of some of the average treatment effects: earlier adopters' substantial improvements compared to their matched non-adopters was offset by later adopters' similar subsequent performance to their matches.

DISCUSSION AND CONCLUSIONS

This study evaluated whether participating in a voluntary management program that requires periodic third-party verification serves as a credible signal of superior environmental management practices. I examined facilities' annual pounds of toxic emissions because the media and US EPA often

[13] "Absolute improvement" is the difference in a facility's emissions in the focal and subsequent year, divided by its average emissions in those two years. I regressed the two improvement variables on "relative emissions", production index, and facility fixed effects using the sample of facilities in sub-industries (4-digit SIC Codes) that included at least one adopter and one non-adopter. "Relative emissions" is the difference between a facility's actual emissions and its predicted emissions based on its sector (4-digit SIC) and size in a given year. Table A-3 in the Appendix presents the results.
[14] "Relative improvement" is the difference between a facility's relative emissions in the focal and subsequent year.
[15] Table A-4 in the Appendix presents the results.

use this metric to rank the “dirtiest” companies. Prior to adoption, eventual adopters were emitting more facility-wide pounds of toxic chemicals than non-adopters. Eventual adopters’ emissions were also posing a greater health hazard on communities surrounding their plants, compared to the health hazards posed by non-adopters’ facility-wide emissions. However, after controlling for facility size, I found that eventual adopters were less pollution intensive, both in terms of pounds of emissions and their associated health hazards.

After becoming certified to ISO 14001, I found evidence that adopters subsequently improved their environmental performance more than a similar set of non-adopters did, across each of the four of the performance metrics I examined: facility-wide pounds of toxic emissions, facility-wide health hazard, pollution intensity in terms of pounds of toxic emissions, and pollution intensity in terms of health hazard.

Taken as a whole, these results demonstrate that a voluntary management program with a robust verification mechanism can indeed distinguish organizations based on their difficult-to-observe management practices. With ISO 14001, this differentiation appears to occur both prior to adoption through a positive selection effect (for pollution intensity) and after certification through subsequent reductions in total emissions and their associated health hazards, as well as in pollution intensity. These findings represent an important departure from prior studies that found no evidence that superior performers disproportionately adopted voluntary management programs with weak or no verification mechanisms. This suggests that third-party certification may be a critical element to ensure that voluntary management programs legitimately distinguish adopters from non-adopters, and thus can be used to resolve information asymmetries regarding difficult-to-observe management practices.

Interesting disparities were revealed when early adopters were compared to late adopters. Across several performance metrics, only early adopters significantly outperformed non-adopters. While this finding may be attributed in part to limited statistical power to assess the performance of more recent adopters (i.e., there are fewer post-adoption years in the dataset), this result is similar to another study that found some evidence that only early adopters of the ISO 9000 quality standard outperformed non-adopters across several financial indicators (Benner & Veloso, 2005). The different performance implications of early versus late adopters may suggest that the former may be motivated to adopt for technical efficiency while the latter are motivated to maintain legitimacy, as suggested by Jiang & Bansal (2003). This finding has important implications for those seeking to use ISO 14001 as a screening mechanism, and suggests the need for more research to explore whether and how adoption motives may affect subsequent performance.

Implications

Firms. This study has important implications for the hundreds of thousands of firms that are relying on voluntary management programs to signal superior management practices to interested buyers, regulators, and local communities. The evidence that ISO 14001 distinguishes adopters as less pollution-intensive may encourage firms concerned about their suppliers' environmental management practices and performance to use ISO 14001 to screen suppliers, a practice some firms have already begun implementing (Fielding, 2000; Sissell, 2000; Strachan, Sinclair, & Lal, 2003). Because adopters subsequently reduced their pollution intensity compared to non-adopters, firms waiting for evidence that ISO 14001 adoption is associated with environmental performance improvement may be encouraged to adopt ISO 14001.

Policymakers. Many policymakers are considering using voluntary management programs to improve the efficiency of achieving environmental, labor, and financial regulatory objectives. Because "priority schemes for [regulatory] inspections are very unsophisticated" (Wasserman, 1987: 20), regulators could redeploy their scarce resources from adopters of voluntary management programs that credibly indicate superior environmental performance levels or trends. Until now, the absence of a demonstrable link between ISO 14001 certification and superior environmental performance has resulted in few environmental agencies shifting their enforcement scrutiny away from certified firms (Hillary & Thorsen, 1999; US EPA, 2003).

My finding evidence that adopters were less pollution intensive at the time of adoption and that adoption is associated with further subsequent reductions in pollution intensity suggests that regulators should seriously consider using ISO 14001 adoption as an indicator of superior performance. The results of this study also have implications for the design of other voluntary management programs. For example, it paints an encouraging picture for the newly revised Responsible Care program, a voluntary management program sponsored by the chemical industry. After persistent skepticism of the program's legitimacy, which was substantially eroded as evidence mounted that participants were performing worse than non-participants (King & Lenox, 2000; Lenox & Nash, 2003), the industry association announced a complete overhaul "intended to improve public perceptions of the industry, and member companies' perceptions of Responsible Care itself" (Chemical Week, 2002: 33). The revised program will require Responsible Care members to obtain third-party certification, a dramatic departure from its fundamental weakness that it had been allowing—and in the US actually *required*—any company that is a member of the national chemical industry association to "join" Responsible Care, regardless of whether they were implementing any of Responsible Care's management codes.

However, in addition to strengthening its verification requirements, the substantial changes to the Responsible Care program may be discarding its most promising feature: its nearly 100 prescriptive management practices governing environmental, health, safety, and community issues. Combining prescriptive management practices with mandatory periodic third-party certification may be the best combination to ensure that voluntary management programs improve participants' performance. Indeed, this is the model adopted by other recent industry-specific voluntary management programs launched by NGOs, such as the Forest Stewardship Council's forest certification program. This approach merits evaluation.

As an alternative to more prescriptive industry-specific management practices, voluntary management programs can also ensure performance improvement among its participants by requiring such improvements as a condition for ongoing participation. The few government-initiated voluntary programs that have actually been shown to elicit performance improvement—such as the US EPA's 33/50 and Indonesia's PROPER-PROKASIH programs (Blackman, Afsah, & Ratunanda, 2004; Khanna & Damon, 1999)—actually require performance improvement as a condition of participation.

Limitations and Future Research

It is important to note several limitations of the data and methods employed in this study. TRI data are self-reported to US EPA by facilities, and are only occasionally verified by some state environmental agencies and by US EPA. In addition, these data are often based on estimates rather than actual measurements. In addition, program evaluation elicits serious challenges in constructing plausible comparison groups. I addressed this by comparing adopters to multiple control groups. The matching methods I employed assume that after one conditions on observable variables (used in generating the propensity score), participation in the program is random. While I gathered comprehensive data based on an extensive literature review, scholars in the future may identify additional factors that influence the decision to adopt ISO 14001. Additionally, although I used fixed effects to reduce the effect of time-invariant unobserved factors, I cannot rule out that time varying, facility-specific effects may affect both adoption and performance. Finally, evaluations of other independently monitored voluntary management programs should be conducted to confirm whether this verification mechanism is a sufficient feature to ensure that voluntary management programs legitimately distinguish adopters and thus resolve information asymmetries.

Voluntary management programs present a plethora of future research opportunities. Many other performance metrics could be employed to assess whether voluntary environmental management programs reduce environmental impacts. For example, participants may be more aggressively improving energy efficiency to reduce greenhouse gas emissions, increasing their use of recycled materials, or enhancing the recyclability of their own products. Participants might also improve training and operational processes and bolster their regulatory knowledge (Dahlström et al., 2003), which can expedite responding to regulators' requests for information, reduce compliance costs, and enhance regulatory relations (Gupta & Piero, 2003). Furthermore, future research could address the extent to which ISO 14001 adopters exhibit superior environmental regulatory compliance, and whether and how operating according to ISO 14001 enables facilities to improve their regulatory compliance.

The vast majority of empirical evaluations of voluntary management programs have focused on those in the United States. Further research is needed in other national contexts, where results may be more profound. Indeed, effects of these programs in the US context may merely represent a "lower bound" of their potential to elicit improved environmental performance. For example, the negative publicity regarding TRI emissions—coupled with a growing awareness in the US of cost-saving pollution prevention opportunities—may have already exhausted most profitable avenues of reducing TRI emissions. Such opportunities may be more abundant in other countries, particularly in developing countries where local regulators, communities, and the media typically exert less pressure for environmental performance improvement. Dasgupta et al.'s (2000) finding that Mexican manufacturers who implemented an ISO 14001-style environmental management system reported better compliance suggests promising results in such domains.

While most studies that have evaluated the extent to which voluntary programs are achieving their ultimate objectives have focused on environmental programs, many research opportunities exist in other domains. Codes of conduct, industry-initiatives, government voluntary programs, and international standards that govern occupational health and safety, human rights, quality management, and other management processes continue to proliferate. The need is greater than ever to discover which programs are legitimately differentiating participants, and which program features are critical to ensure their credibility. Absent such knowledge, many of the millions of hours and dollars spent implementing these tools may be wasted. The methods presented this paper could be employed to evaluate these issues.

REFERENCES

Alchian, A. A. & Demsetz, H. 1972. Production, information costs, and economic organization. ***The American Economic Review***, 62(5): 777-795.

Barbera, B. M. & Lyona, J. D. 1996. Detecting abnormal operating performance: The empirical power and specification of test statistics. ***Journal of Financial Economics***, 41(3): 359-399.

Benner, M. & Veloso, F. 2005. ***Process management practices and performance: Competition and the moderating effect of technological capabilities***. Working Paper, Department of Engineering and Public Policy, Carnegie Mellon University.

Bertrand, M., Duflo, E., & Mullainathan, S. 2004. How much should we trust difference-in-difference estimates. ***Quarterly Journal of Economics***, 119: 249-275.

Blackman, A., Afsah, S., & Ratunanda, D. 2004. How Do Public Disclosure Pollution Control Programs Work? Evidence from Indonesia. ***Human Ecology Review***, 11(3).

Burglund, M. 1999. Does an environmental management system really improve the environment? ***Pollution Control News***, 19(7): 5.

CAFOD. 2004. ***Clean Up Your Computer: Working***

conditions in the electronics sector. London: Catholic Agency for Overseas Development.

Chemical Week. 2002. Details of a Revamped Responsible Care Take Shape. (Nov. 20): 33.

Chin, K. S. & Pun, K.-F. 1999. Factors influencing ISO 14000 implementation in the printed circuit board manufacturing industry in Hong Kong. ***Journal of Environmental Planning & Management***, 42(1): 123-134.

Christmann, P. & Taylor, G. 2001. Globalization and the environment: determinants of firm self-regulation in China. ***Journal of International Business Studies***, 32(3): 439-458.

Corbett, C. J. & Luca, A. 2002. ***Global Survey on ISO 9000 and ISO 14000: Summary of Findings***. Unpublished manuscript, The Anderson School, University of California at Los Angeles.

Corbett, C. J., Montes-Sancho, M. J., & Kirsch, D. A. 2005. The Financial Impact of ISO 9000 Certification in the United States: An Empirical Analysis. ***Management Science***, 51(7): 1046-1059.

Courville, S. 2003. Social accountability audits: Challenging or defending democratic governance? ***Law & Policy***, 25(3): 269-298.

Dahlström, K., Howes, C., Leinster, P., & Skea, J. 2003. Environmental management systems and company performance: Assessing the case for extending risk-based regulation. ***European Environment***, 13: 187–203.

Darnall, N. & Edwards, D. J. Forthcoming. Predicting the cost of environmental management system adoption: The role of capabilities, resources, and ownership structure. ***Strategic Management Journal***.

Darnall, N., Gallagher, D. R., Andrews, R. N. L., & Amaral, D. 2000. Environmental management systems: Opportunities for improved environmental and business strategy? ***Environmental Quality Management***, 9(3): 1-10.

Dasgupta, S., Hettige, H., & Wheeler, D. 2000. What improves environmental compliance? Evidence from Mexican industry. ***Journal of Environmental Economics and Management***, 39: 39-66.

Dehejia, R. H. & Wahba, S. 1999. Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. ***Journal of the American Statistical Association***, 94(448): 1053-1062.

Delmas, M. 2004. ***The take-off period in the diffusion of international standards: The case of ISO 14001***. Unpublished manuscript, Donald Bren School of Environmental Science & Management.

Drukker, D. M. 2003. Testing for serial correlation in linear panel-data models. ***Stata Journal***, 3(2): 168-177.

Earnhart, D. 2004. Regulatory factors shaping environmental performance at publicly-owned treatment plants. ***Journal of Environmental Economics and Management***, 48(1): 655-681.

Easton, G. S. & Jarrell, S. L. 1998. The effects of Total Quality Management on corporate performance: An empirical investigation. ***Journal of Business Ethics***, 71(2): 253-307.

Edelman, L. B. 1992. Legal ambiguity and symbolic structures: Organizational mediation of civil rights law. ***American Journal of Sociology***, 97(6): 1531-1576.

Eichler, M. & Lechner, M. 2002. An evaluation of public employment programmes in the East German state of Sachsen-Anhalt. ***Labour Economics***, 9: 143–186.

Fielding, S. 1999. Going for the green: ISO 14001 delivers profits. ***Industrial Management***, (March-April): 31.

Fielding, S. 2000. ISO 14001 delivers profits. ***Automotive Engineering International***, (July): 140.

Gerde, V. W. & Logsdon, J. M. 2001. Measuring environmental performance: use of the toxics release inventory (TRI) and other US environmental databases. ***Business Strategy and the Environment***, 10(5): 269-285.

Gray, W. B. & Scholz, J. T. 1993. Does regulatory enforcement work? A panel analysis of OSHA enforcement. ***Law & Society Review***, 27(1): 177-214.

Gunningham, N. 1995. Environment, self-regulation, and the chemical industry: Assessing Responsible Care. ***Law & Policy***, 17(1): 57-109.

Gupta, M. & Piero, T. 2003. Environmental management is good business. ***Industrial Management***, 45(5): 14-19.

Hall, B. & Kerr, M. L. 1991. ***1991-1992 Green Index: A State-By-State Guide to the Nation's Environmental Health***. Washington DC: Island Press.

Heckman, J. J., Ichimura, H., & Todd, P. 1998. Matching as an econometric evaluation estimator. ***The Review of Economic Studies***, 65(2): 261-294.

Heckman, J. J., Ichimura, H., & Todd, P. E. 1997. Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. ***The Review of Economic Studies***, 64(4): 605-654.

Heras, I., Dick, G. P. M., & Casadesús, M. 2002. ISO 9000 registration's impact on sales and profitability: A longitudinal analysis of performance before and after accreditation. ***International Journal of Quality and Reliability Management***, 19(6): 774-791.

Hillary, R. & Thorsen, N. 1999. Regulatory and self-regulatory measures as routes to promote cleaner production. ***Journal of Cleaner Production***, 7(1): 1-11.

IAF. 2003. ***IAF Guidance on the Application of ISO/IEC Guide 66: General Requirements for Bodies Operating Assessment and Certification/registration of Environmental Management Systems (EMS) (Issue 3, IAF GD 6:2003)***. Cherrybrook, Australia: International Accreditation Forum, Inc.

International Organization for Standardization. 2004. ***The ISO Survey of ISO 9000 and ISO 14000 Certificates: Twelfth cycle: up to and including 31 December 2003.*** Geneva: International Organization for Standardization (ISO).

Jiang, R. J. & Bansal, P. 2003. Seeing the need for ISO 14001. ***Journal of Management Studies***, 40(4): 1047-1067.

Jin, G. Z. & Leslie, P. 2003. The effect of information on product quality: Evidence from restaurant hygiene grade cards. ***Quarterly Journal of Economics***, 118(2): 409-451.

Khanna, M. & Damon, L. 1999. EPA's voluntary 33/50 program: Impact on toxic releases and economic performance of firms. ***Journal of Environmental Economics and Management***, 37: 1-25.

King, A. 1995. Innovation from differentiation: Pollution control departments and innovation in the printed circuit industry. ***IEEE Transactions on Engineering Management***, 42(3): 270-277.

King, A. 1999. Retrieving and transferring embodied data: Implications for the management of interdependence within organizations. ***Management Science***, 45(7): 918-935.

King, A. & Lenox, M. 2000. Industry self-regulation without sanctions: The chemical industry's responsible care program. ***Academy of Management Journal***, 43(4): 698-716.

King, A. & Lenox, M. 2001. Lean and green? An empirical examination of the relationship between lean production and environmental performance. ***Production and Operations Management***, 10(3): 244-256.

King, A. A., Lenox, M. J., & Terlaak, A. Forthcoming. Strategic use of decentralized institutions: Exploring certification with the ISO 14001 management standard. ***Academy of Management Journal***.

Kniesner, T. J. & Leeth, J. D. 2004. Data mining mining data: MSHA enforcement efforts, underground coal mine safety, and new health policy implications. ***Journal of Risk and Uncertainty***, 29(2): 83-111.

Kogut, B. & Zander, U. 1992. Knowledge of the firm, combinative capabilities, and the replication of technology. ***Organization Science***, 3(3): 383-397.

Kolk, A. 2000. ***The economics of environmental management***. New York: Financial Times.

Krueger, A. B. & Mas, A. 2004. Strikes, scabs and tread separations: Labor strife and the production of defective Bridgestone/Firestone tires. ***Journal of Political Economy***, 112(2): 253-289.

Lenox, M. J. & Nash, J. 2003. Industry self-regulation and adverse selection: A comparison across four trade association programs. ***Business Strategy and the Environment***, 12: 343-356.

Leuven, E. & Sianesi, B. 2003. PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing, Version 1.2.3. http://ideas.repec.org/c/boc/bocode/s432001.html.

Levinson, A. 1996. Environmental regulations and manufacturers' location choices: Evidence from the Census of Manufactures. ***Journal of Public Economics***, 62: 5-29.

Matthews, D. H. 2001. ***Assessment and Design of Industrial Environmental Management Systems***. PhD Dissertation, Carnegie Mellon University, Pittsburgh.

McGraw Hill. 2001. ***Market Report - ISO 14001 Registered Company Directory North America. Vol 2, No 1***. Washington DC.

McNeil, D. G., Jr. 2004. At last, a company takes PETA seriously. ***New York Times***, (July 25): 4-4.

Medina-Ross, V. 2002. ***Environmental Management in the Mexican Chemical Sector: Drivers, Obstacles and Implications***. Unpublished manuscript, Norwich, UK.

Melnyk, S. A., Sroufe, R. P., Calantone, R. L., & Montabon, F. L. 2002. Assessing the effectiveness of US voluntary programmes: An empirical study. ***International Journal of Production Research***, 40(8): 1853-1878.

Mezias, S. J. 1990. An institutional model of organizational practice: financial reporting at the Fortune 200. ***Administrative Science Quarterly***, 35(431-457).

Mukamel, D. B. & Mushlin, A. I. 1998. Quality of care information makes a difference: An analysis of market share and price changes after publication of the New

York State Cardiage Surgery Mortality Reports. ***Medical Care***, 36(7): 945-954.

Naimon, J., Shastri, K., & Sten, M. 1997. Do environmental management programs improve environmental performance trends? A study of Standard & Poors 500 companies. ***Environmental Quality Management***, 7(1): 81-90.

NAPA. 2001. ***Third-Party Auditing of Environmental Management Systems: U.S. Registration Practices for ISO 14001***. Washington, DC: National Academy of Public Administration (NAPA).

NEETF. 2001. ***Going Green... Upstream: The Promise of Supplier Environmental Management***. Washington DC: The National Environmental Education & Training Foundation (NEETF).

O'Rourke, D. 2003. Outsourcing regulation: Analyzing nongovernmental systems of labor standards and monitoring. ***Policy Studies Journal***, 31(1): 1-29.

Potoski, M. & Prakash, A. Forthcoming. Covenants with Weak Swords: ISO 14001 and Facilities' Environmental Performance. ***Journal of Policy Analysis and Management***.

Prakash, A. 2000. ***Greening the firm: The politics of corporate environmentalism***. New York: Cambridge University Press.

QSU. 2002a. ***ISO 9000 Registered Company Directory North America***. Fairfax, VA: QSU Publishing.

QSU. 2002b. ***ISO 14001 Registered Company Directory North America, Volume 2, Number 2***. Fairfax, VA: QSU Publishing Company,.

Redinger, C. F. & Levine, S. P. 1998. Development and evaluation of the Michigan Occupational Health and Safety Management System Assessment Instrument: A universal OHSMS performance measurement tool. ***American Industrial Hygiene Association Journal***, 59(8): 572-581.

Rivera, J. & de Leon, P. 2004. Is greener whiter? The Sustainable Slopes Program and the voluntary environmental performance of western ski areas. ***Policy Studies Journal***, 32(3).

Rivera, J., de Leon, P., & Koerber, C. 2005. ***Is greener whiter yet? The Sustainable Slopes program after five years***. Working Paper, George Washington University.

Rondinelli, D. & Vastag, G. 2000. Panacea, common sense, or just a label? The value of ISO 14001 environmental management systems. ***European Management Journal***, 18(5): 499-510.

Rondinelli, D. A. & London, T. 2003. How corporations and environmental groups cooperate: Assessing cross-sector alliances and collaborations. ***Academy of Management Executive***, 17(1): 61-76.

Rosenbaum, P. R. & Rubin, D. B. 1983. The central role of the propensity score in observational studies for causal effects. ***Biometrika***, 70(1): 41-55.

Rosenbaum, P. R. & Rubin, D. B. 1985. Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. ***The American Statistician***, 39(1): 33-38.

Russo, M. V. & Fouts, P. A. 1997. A resource-based perspective on corporate environmental performance and profitability. ***Academy of Management Journal***, 40: 534-559.

Sianesi, B. 2004. An Evaluation of the Active Labour Market Programmes in Sweden. ***The Review of Economics and Statistics***, 86(1): 133-155.

Sissell, K. 2000. Autos and electronics drive certification. ***Chemical Week***, 162(14): 42-43.

Smith, G. & Feldman, D. 2003. ***Company Codes of Conduct and International Standards: An Analytical Comparison - Part I of II***. Washington, DC: World Bank.

Smith, J. A. & Todd, P. E. 2005. Does matching overcome LaLonde's critique of nonexperimental estimators? ***Journal of Econometrics***, 125(1-2): 305-353.

Spence, M. 1973. Job market signaling. ***The Quarterly Journal of Economics***, 87(3): 355-374.

Strachan, P. A., Sinclair, I. M., & Lal, D. 2003. Managing ISO 14001 implementation in the United Kingdom Continental Shelf (UKCS). ***Corporate Social Responsibility and Environmental Management***, 10(1): 50-63.

Strom, S. 1996. A sweetheart becomes suspect; Looking behind those Kathie Lee labels. ***The New York Times***, (June 27).

Szymanski, M. & Tiwari, P. 2004. ISO 14001 and the reduction of toxic emissions. ***Policy Reform***, 7(1): 31–42.

Terlaak, A. & King, A. Forthcoming. The effect of certification with the ISO 9000 Quality Management Standard: A signaling approach. ***Journal of Economic Behavior and Organization***.

Terziovski, M., Samson, D., & Dow, D. 1997. The business value of quality management systems certification - Evidence from Australia and New Zealand. ***Journal of Operations Management***, 15(1): 1-18.

Toffel, M. 2000. Anticipating greener supply chain demands: One Singapore company's journey to ISO 14001. In R. Hillary (Ed.), ***ISO 14001: Case Studies and Practical Experiences***. Sheffield, UK: Greenleaf Publishing.

Toffel, M. W. 2005. ***Voluntary environmental management initiatives: Smoke signals or smoke screens?*** Phd Dissertation, Haas School of Business, University of California, Berkeley.

Toffel, M. W., Hill, N., & McElhaney, K. A. 2003. BMW Group's sustainability management system: Preliminary results, ongoing challenges, and the UN Global Compact. ***Corporate Environmental Strategy***, 10(3): 51-62.

Toffel, M. W. & Marshall, J. D. 2004. Improving environmental performance assessment: A comparative analysis of weighting methods used to evaluate chemical release inventories. ***Journal of Industrial Ecology***, 8(1-2): 143-172.

Tolbert, P. S. & Zucker, L. G. 1983. Institutional Sources of Change in the Formal Structure of Organizations: The Diffusion of Civil Service Reform, 1880-1935. ***Administrative Science Quarterly***, 28(1): 22-39.

US EPA. 1995. ***Sector Notebook - Profile of the Electronics and Computer Industry (EPA/310-R-95-002)***. Washington, DC: US Environmental Protection Agency.

US EPA. 1999. ***EPCRA Section 313 Release Reporting Requirements "blue brochure"***. Washington, DC: US Environmental Protection Agency.

US EPA. 2002. ***User's Manual for RSEI Version 2.1 [1988-2000 TRI Data]***. Washington, DC: US Environmental Protection Agency.

US EPA. 2003. ***Guidance on the Use of Environmental Management Systems in Enforcement Settlements as Injunctive Relief and Supplemental Environmental Projects***. Washington DC: US Environmental Protection Agency.

Walton, S. V., Handfield, R. B., & Melnyk, S. A. 1998. The green supply chain: Integrating suppliers into environmental management process. ***International Journal of Purchasing and Materials Management***, 34(2): 2-11.

Wasserman, C. 1987. ***Environmental compliance and enforcement: Theory, practice, and the challenge to environmental economists (EPA Report 230-10-87-030)***. Paper presented at the Association of Environmental and Resource Economists, Newark, Delaware.

Welch, E. W., Mazur, A., & Bretschneider, S. 2000. Voluntary behavior by electric utilities: levels of adoption and contribution of the climate challenge program to the reduction of carbon dioxide. ***Journal of Policy Analysis and Management***, 19: 407–425.

Westphal, J. D., Gulati, R., & Shortell, S. M. 1997. Customization or conformity? An institutional and network perspective on the content and consequences of TQM adoption. ***Administrative Science Quarterly***, 42(2): 366-394.

Westphal, J. D. & Zajac, E. J. 1994. Substance and symbolism in CEO's long-term incentive plans. ***Administrative Science Quarterly***, 39: 367-390.

Williamson, O. E. 1985. ***The Economic Institutions of Capitalism: Firms, Markets, Relational Contracting***. New York: Free Press.

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

Yiridoe, E. K., Clark, J. S., Marett, G. E., Gordon, R., & Duinker, P. 2003. ISO 14001 EMS standard registration decisions among Canadian organizations. ***Agribusiness***, 19(4): 439-457.

Table 1. Variable Definitions and Descriptive Statistics

Selection and treatment analysis	Description and source	N	Mean	SD	Min*	Max
Certification year	Dummy coded 1 in the year when a facility became ISO 14001 certified. ISO (2002b) and state websites.	207078	0.005	0.068	950	
Post-certification	Dummy coded 1 in the year a facility becomes ISO 14001 certified and thereafter. ISO (2002b) and state websites.	207078	0.014	0.119	2953	
Facility-wide emissions	Log of 1 plus the number of pounds of toxic chemical emissions reported to the US EPA Toxic Release Inventory	158760	5.10	4.61	0	17.57
Facility-wide health hazard	For each TRI chemical, the pounds a facility releases to air in a given year is multiplied by the chemical-specific toxicity weights pertaining to inhalation exposure, and then these products are summed across all of a facility's annual releases, log transformed after adding 1.	159107	8.09	6.52	0	24.39
Production index	In 1997, log of facility employment. In other years, facility employment IS iteratively adjusted by the facility's annual production ratios (see text for more details). US EPA Toxic Release Inventory database (production ratios) and Dun & Bradstreet (employment).	79896	4.97	1.39	-1.16	12.15

Propensity score estimation	Description and source	N=45162 Mean	SD	Min*	Max
Certification year	As defined above.	0.01	0.12	*618*	
Lagged environmental management system evidence	Dummy coded one in years when a facility's submissions to the US EPA Toxic Release Inventory indicated source reduction methods included internal pollution prevention audits, participative team management, or employee recommendations under a formal company program. Lagged one year. US EPA TRI database.	0.18	0.39	*8319*	
US EPA 33/50 participant	Dummy coded one for facilities that are members of companies that participated in this US EPA program (which ended in 1995). US EPA via a Freedom of Information Act Request.	0.29	0.45	*12950*	
Lagged ISO 9000 certified	Dummy coded one starting the year a facility was certified to the ISO 9000 Quality Management System Standard. Lagged one year. QSU (2002a).	0.04	0.19	*1673*	
Lagged RCRA violations	Number of times during the prior two years the facility was cited for violations of hazardous waste regulations pursuant to the US Resource Conservation and Recovery Act, the federal environmental statute that is the basis of far more environmental inspections and enforcement actions than any other for the industries in the sample (US EPA, 1995). US EPA RCRIS database via a Freedom of Information Act request.	0.41	1.16	0	13
Lagged RCRA inspections	Number of times during the prior two years the facility was inspected for violations of hazardous waste regulations pursuant to the US Resource Conservation and Recovery Act. US EPA RCRIS database.	0.37	0.74	0	12
Lagged enforcement actions (dummy)	Number of years during the prior two years that the US EPA brought an enforcement action against a facility for serious violations of any federal environmental regulation. US EPA's ICIS database via a Freedom of Information Act request.	0.02	0.10	0	1
Lagged waste to POTW (dummy)	Dummy coded one in years when a facility releases effluent with toxic chemicals to publicly-owned treatment facility (POTW). Lagged one year. US EPA TRI database.	0.32	0.47	*14542*	
State compliance cost	A state's relative compliance costs based on pollution abatement operating adjusted for differences in industry composition and other factors. Levinson (1996).	-0.06	0.13	-0.57	0.273
State environmental policy comprehensiveness	An index that measures the extent to which each state has implemented environmental policies to address 50 pollution, waste, and land use issues. Hall & Kerr (1991).	22.98	6.38	5	38
Percent college graduates in 2000	Percentage of those 25 years and older living within the facility's Census Tract who attended college. US Census Bureau 2000 Decennial Survey.	0.19	0.13	0	1
Per capita income in 1999	Log median per capita income for 1999 of individuals living within the facility's Census Tract. US Census Bureau 2000 Decennial Survey.	9.82	0.35	6.89	11.88
Employment in 1997	Log facility employment in 1997. Dun & Bradstreet.	4.91	1.31	2.30	10.17
Lagged facility-wide emissions	Average emissions (defined above) during prior two years.	5.86	4.18	0	16.06
Lagged facility-wide health hazard	Average health hazard (defined above) during prior two years.	9.32	5.96	0	23.98
Lagged emission hazardousness per pound	Average health hazardousness per pound of toxic air emission during prior two years. Emission hazardousness is calculated as the log of 1 plus the following ratio: *facility-wide health hazard* divided by *facility-wide emissions* (each term in this ratio as defined above except not log transformed).	3.80	3.38	0	13.82
Lagged production index	Average production index (defined above) during prior two years	4.92	1.33	0.29	10.99
Trend in facility-wide emissions	Lagged 4-year outcome variables temporal trends were calculated as the average of 1 and 2 year lagged values minus the average of 3 and 4 year lags, divided by the sum of these two averages.	-0.07	0.35	-1	1
Trend in facility-wide health hazard		-0.07	0.35	-1	1
Trend in emission hazardousness per pound		0.004	0.09	-1	1

* Instead of minimum and maximum, italicized values in the last column displays the number of observations coded 1 for dummy variables. N represents facility-year observations. For selection and treatment analyses, the sample includes observations from 16,896 facilities during 1991-2003. For propensity score estimation, the sample includes 45162 facility-year observations from 7764 facilities during 1996-2003.

Table 2. Selection Results: Probit Models

Dependent variable: Became ISO 14001 certified this year (dummy)

Model:	(1) **Facility-wide emissions**	(2) **Facility-wide health hazard**	(3) **Pollution intensity - emissions**	(4) **Pollution intensity - health hazard**
Log pounds of emissions (2-year average lag)	0.022 [0.003]***		-0.013 [0.004]***	
Log health hazard (2-year average lag)		0.017 [0.002]***		-0.007 [0.003]**
Log production index (2-year average lag)			0.188 [0.014]***	0.185 [0.014]***
Year dummies	Y	Y	Y	Y
Sub-industry dummies (3-digit SIC Code)	Y	Y	Y	Y
Observations (facility-years)	89786	89786	4824	48240
Facilities	15183	15183	8633	8633
Adopters	935	935	785	785
Pseudo R^2	0.13	0.13	0.18	0.18
Wald χ^2	1034.26***	1039.78***	1002.62***	1002.11***
Log likelihood intercept only	-5198.05	-5198.05	-4011.41	-4011.41
Log likelihood full model	-4513.46	-4506.81	-3274.29	-3275.91
Marginal effect of one standard deviation increase in the *pounds of emission* or *health hazard* variable, compared to the baseline probability of adoption evaluated at mean of all variables	28%	32%	-15%	-12%

This table displays probit coefficients with robust standard errors clustered by facility in brackets. *** p<0.01, ** p<0.05, * p<0.10. The sample includes 1996-2003, and sub-industries (4-digit SIC Codes) that have at least one facility that adopted ISO 14001 during this period within SIC Codes 28 or 34-37; adopters are excluded from the sample after their certification year. All four specifications include dummies for years and industries (3-digit SIC Codes).

Table 3. Difference-in-Differences Results

Model:	**Facility-wide emissions**			**Facility-wide health hazard**			**Pollution intensity–emissions**			**Pollution intensity–health hazard**		
Dependent variable	Pounds of emissions			Health hazard			Pounds of emissions			Health hazard		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample	A	B	C	A	B	C	A	B	C	A	B	C
Post-certification (dummy)	-0.39	-0.30	-0.11	-0.06	-0.45	-0.31	-0.32	-0.24	-0.17	-0.01	-0.45	-0.42
	[0.11]***	[0.16]*	[0.14]	[0.17]	[0.23]*	[0.21]+	[0.12]***	[0.14]*	[0.13]+	[0.18]	[0.20]**	[0.19]**
Production index							Y	Y	Y	Y	Y	Y
Facility fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year dummies or counters	Year	Counter	Counter	Year	Counter	Counter	Year	Counter	Counter	Year	Counter	Counter
Matched group dummies × counter		Y	Y		Y	Y		Y	Y		Y	Y
N	129835	22655	13299	130270	22664	13304	66588	19924	11602	66588	19927	11602
Facilities	12077	1979	1770	12118	1980	1770	7627	1979	1764	7627	1980	1764
Adopters	860	412	380	860	412	380	522	412	374	522	412	374
Adjusted R^2	0.69	0.71	0.78	0.67	0.66	0.81	0.76	0.77	0.81	0.72	0.70	0.84

Sample A includes all non-adopters in those sub-industries (the 4-digit SIC Codes within SIC Codes 28 and 34-37) that include at least one adopter, as well as all adopters in these industries. These specifications include year dummies, facility fixed effects, production index (for pollution intensity models only), and employ all observations during 1991-2003.

Sample B is a matched sample consisting of those non-adopters whose propensity scores in a given year were among the 5-nearest neighbors for each adopter in its adoption year, excluding those non-adopters whose distance from the adopter exceeds a caliper limit. Sample C is another 5-nearest neighbor matched sample, but potential matches were restricted to non-adopters with propensity scores in the adoption year that also shared the adopter's industry (2-digit SIC code), size (employment) range category, and emissions pre-trend category, and a wider caliper restriction was used. For Samples B and C, the specification includes a counter for each matched group to denote how many years until (or after) the match year, an interaction between the counter and a matched group dummy, facility fixed effects, and production index (for pollution intensity models only), and employ observations within +/-4 years of the match year during 1991-2003.

Because first-order autocorrelation is indicated in all specifications (the F-statistic for the test of the null hypothesis of no first-order autocorrelation is rejected at p<0.001), standard errors clustered by facility are employed to accommodate an arbitrary autocorrelation process and heteroscedasticity.
*** p<0.01, ** p<0.05, * p<0.10 for a two-tailed test. + p<0.10 for a one-tailed test

APPENDIX

Adoption Determinants

I conducted an extensive literature review to identify potential adoption determinants that could be included in the model to estimate ISO 14001 propensity scores. This review drew on Bansal & Hunter (2003), Baron (2003), Carraro, Katsoulacos, & Xepapadeas (1996), Christmann & Taylor (2001), Delmas (2000; 2002), Delmas & Toffel, 2004 (2004), Florida & Davison (2001), Hamilton (1999), Helland (1998), Henriques & Sadorsky (1996), Khanna & Anton (2002), Khanna & Damon (1999); King et al., (Forthcoming), Lawrence & Morell (1995), Majumdar & Marcus (2001), Maxwell, Lyon, & Hackett (2000), Nash, Ehrenfeld, MacDonagh-Dumler, & Thorens (2000), Raines (2002), Rugman & Verbeke (1998), Vidovic & Khanna, (2003), and Welch, Mazur, & Bretschneider (2000).

Table A-1. Results of Logit Model to Estimate Propensity Scores

Dependent variable: Became ISO 14001 certified this year (Certification year dummy)

	Logit coefficient	SE	dF/dx
Evidence of environmental management system, 1 year ago	0.38	[0.11]***	0.0021 #
EPA 33/50 participant	0.17	[0.10]*	0.0009 #
ISO 9000 certified at least 1 year ago	0.28	[0.17]	0.0015 #
RCRA violations, average of 1 and 2 years ago	-0.04	[0.04]	-0.0002
RCRA inspections, average of 1 and 2 years ago	0.08	[0.07]	0.0004
Any enforcement actions, average of 1 and 2 years ago	-0.17	[0.40]	-0.0008
Any waste transfers to POTW, 1 year ago	0.37	[0.10]***	0.0019 #
Compliance cost per state	2.25	[0.47]***	0.0107
State environmental policy comprehensiveness	-0.02	[0.01]	-0.0001
Percent college graduates in 2000	0.45	[0.57]	0.0021
Per capita income in 1999 (log)	-0.20	[0.21]	-0.0009
Employment in 1997 (log)	0.34	[0.11]***	0.0016
Pounds emitted, average of 1 & 2 years ago (log)	-0.03	[0.07]	-0.0001
Health hazard, average of 1 & 2 years ago (log)	-0.01	[0.06]	-0.0001
Emission hazardousness per pound, average of 1 & 2 years ago (log)	0.01	[0.06]	0.0000
Dummy for missing values of emission hazardousness per pound, average of 1 & 2 years ago	2.62	[0.39]***	0.0335 #
Production index, average of 1 & 2 years ago (log)	0.04	[0.11]	0.0002
Dummy for missing values of production index, average of 1 & 2 years ago (log)	0.48	[0.11]***	0.0027 #
Percent change in facility-wide emissions over prior 4 years	-0.02	[0.56]	-0.0001
Dummy for missing values of percent change in facility-wide emissions over prior 4 years	-1.40	[0.92]	-0.0047 #
Percent change in facility-wide health hazard over prior 4 years	1.54	[0.60]***	0.0073
Dummy for missing values of percent change in facility-wide health hazard over prior 4 years	0.26	[0.93]	0.0013 #
Percent change in hazardousness per pound of emission over prior 4 years	-0.02	[0.46]	-0.0001
Dummy for missing values of percent change in hazardousness per pound of emission over prior 4 years	-0.97	[0.22]***	-0.0040 #
Year dummies	Y		
EPA Region dummies	Y		
SIC 2-digit dummies	Y		
Observations (facility-years)	45162		
Facilities	7764		
Adopters	618		
Pseudo R^2	0.17		
Wald χ^2	863.66***		
Log likelihood intercept only	-3265.91		
Log likelihood full model	-2697.80		

This table presents the results of the logit model used to estimate propensity scores. A full set of dummies is included for years, EPA Regions, 2-digit SIC Codes, as well as a dummies to denote missing values of lagged emission hazardousness and each of the lagged 4-year outcome trends. Brackets contain robust standard errors clustered by facility: * p<0.10, ** p<0.05, *** p<0.01. The second column displays the change in the probability of adoption for an infinitesimal change (or unit change when denoted #) in each independent variable evaluated at the mean all variables. Sample includes facilities in sub-industries (4-digit SIC Codes) with at least one adopter and one non-adopter; adopters are omitted after their certification year.

Table A-2. Indicators of Covariate and Pre-Trend Balancing

	(A1)	(A2)	(A3)	(A4)	(B1)	(B2)	(B3)	(B4)	(B5)	(C1)	(C2)	(C3)	(C4)	(C5)
	Sub-sample A: Entire sample in 1996				**Sub-sample B: Five nearest neighbors within-year with caliper**					**Sample C: Five nearest neighbors within industry-year-size category-pretrend category, with caliper**				
	Mean, non-adopters	Mean, adopters	T-test p value	Std bias	Mean, non-adopters	Mean, adopters	T-test p value	Std bias	Pct bias reduction	Mean, non-adopters	Mean, adopters	T-test p value	Std bias	Pct bias reduction
Adoption Determinants														
Evidence of environmental management system	0.18	0.28	**0.00**	-17%	0.23	0.25	0.54	-4%	76%	0.23	0.22	0.54	3%	118%
EPA 33/50 participant	0.25	0.46	**0.00**	-31%	0.42	0.45	0.33	-5%	84%	0.39	0.41	0.61	-3%	90%
ISO 9000 certified by this year	0.02	0.02	0.32	-3%	0.07	0.08	0.51	-4%	-33%	0.07	0.06	0.81	1%	133%
RCRA violations, sum of 1 & 2 years ago	0.42	0.53	**0.04**	-6%	0.53	0.45	0.26	6%	200%	0.41	0.53	**0.08**	-10%	-67%
RCRA inspections, sum of 1 & 2 years ago	0.37	0.50	**0.00**	-12%	0.46	0.40	0.24	7%	158%	0.40	0.45	0.26	-6%	50%
Enforcement actions, sum of 1 & 2 years ago	0.01	0.01	0.40	3%	0.03	0.02	0.47	4%	-33%	0.02	0.03	0.85	-1%	133%
Any waste transfers to POTW 1 year ago	0.24	0.44	**0.00**	-30%	0.50	0.48	0.53	3%	110%	0.44	0.46	0.48	-4%	87%
State compliance cost	-0.07	-0.04	**0.00**	-17%	-0.05	-0.04	0.43	-4%	76%	-0.06	-0.05	0.40	-4%	76%
State environmental policy comprehensiveness	23.19	23.00	0.48	2%	22.41	23.08	**0.07**	-11%	650%	22.64	22.77	0.75	-2%	200%
Percent college graduates in 2000	0.19	0.19	0.65	-1%	0.19	0.19	0.95	0%	100%	0.20	0.19	0.70	2%	300%
Per capita income in 1999 (log)	9.82	9.84	0.31	-3%	9.83	9.84	0.60	-3%	0%	9.85	9.85	0.80	1%	133%
Employment in 1997 (log)	4.79	5.94	**0.00**	-65%	5.90	5.92	0.76	-1%	98%	5.71	5.72	0.57	-1%	98%
Pounds emitted, average of 1 & 2 years ago (log)	5.26	6.21	**0.00**	-15%	5.91	5.46	**0.09**	10%	167%	5.79	5.66	0.55	2%	113%
Health hazard, average of 1 & 2 years ago (log)	7.99	9.64	**0.00**	-18%	9.74	9.21	0.17	8%	144%	9.72	9.54	0.58	2%	111%
Emission hazardousness per pound, average of 1 & 2 years ago (log)	3.10	3.74	**0.00**	-13%	4.05	3.98	0.74	2%	115%	4.39	4.19	0.40	5%	138%
Production index, average of 1 & 2 years ago (log)	4.77	5.89	**0.00**	-60%	5.92	5.93	0.83	-1%	98%	5.74	5.73	0.64	1%	102%
Outcome Trends Prior to Match Year														
Facility-wide emissions trend over prior 4 years	-0.12	-0.11	0.38	-3%	-0.05	-0.04	0.67	-3%	0%	-0.07	-0.06	0.35	-2%	33%
Facility-wide health hazard trend over prior 4 years	-0.12	-0.10	0.34	-3%	-0.03	-0.03	0.94	0%	100%	-0.06	-0.05	0.60	-2%	33%
Unit hazardousness trend over prior 4 years	0.03	0.03	0.90	1%	0.07	0.05	0.50	5%	-400%	0.02	0.03	0.53	-5%	600%
			Mean:	-15%			Mean:	0.5%	90%			Mean:	-1%	131%
			Median:	-12%			Median:	0.0%	98%			Median:	-1%	111%

This table compares the mean covariates and pre-trends between the non-adopters and adopters across three samples. The *p* values are from t-tests that evaluate whether the group means are equal. Standardized bias is calculated as the difference between group means divided by the square root of the average variance across the two groups.

$$\text{Standardized Bias} = 100 * \frac{(\overline{X}_1 - \overline{X}_0)}{\sqrt{0.5 * (V_1(X) + V_0(X))}}$$

where $\overline{X}_1$ (V_1) represents the mean (variance) in the treatment group and $\overline{X}_0$ (V_0) is the analogue for the control group. The standardized difference after matching uses the same approach but employs the means and variances of the matched sample. The standardization allows comparisons between covariates, as well as before and after matching.

Sample A includes the 607 adopters and 8455 non-adopters in the sub-industries (4-digit SIC Codes) with at least one adopter and one non-adopter and that have propensity scores in 1996. The covariates and lagged performance values compared are those from the facilities' 1996 values. Sample B is a matched sample of those non-adopters whose propensity scores in a given year were among the 5-nearest neighbors for each adopter in its adoption year, excluding those nearest neighbors whose distance from the adopter exceeds a caliper limit. This sample includes 509 adopters and 1808 non-adopters. Sample C is another 5-nearest neighbor matched sample, but potential matches were restricted to non-adopters with propensity scores in the adoption year that also shared the adopter's industry (2-digit SIC code), size (employment) range category, and emissions pre-trend category, and a wider caliper restriction was used. This Sample includes 465 adopters and 1622 non-adopters. To compare adopters and non-adopter means within Sample B and Sample C, I first calculated mean covariate values for adopters and non-adopters within each of the matched groups. The table displays and compares the means of these group means because the difference-in-difference specifications that employ Sample B or Sample C compares adopters and non-adopters within these matched groups.

Table A-3. Robustness Test Results: Alternative Treatment Analysis Models

Performance metric	(1) **Absolute improvement, pounds**	(2) **Absolute improvement, health hazard**	(3) **Relative improvement, pounds**	(4) **Relative improvement, health hazard**
Robustness test for	*Facility-wide emissions*	*Facility-wide health hazard*	*Pollution intensity – pounds*	*Pollution intensity – health hazard*
Post-certification	0.44	0.04	0.43	0.55
Unadjusted SE	(0.03)***	(0.02)*	(0.08)***	(0.12)***
Cluster SE	[0.03]***	[0.02][a]	[0.10]***	[0.14]***
Relative emissions, pounds	0.17		0.64	
	[0.00]***		[0.01]***	
Relative emissions, health hazard		0.12		0.66
		[0.00]***		[0.01]***
Production index	-0.01	0.01	0.01	0.06
	[0.01]	[0.01]	[0.04]	[0.05]
Facility fixed effects	Y	Y	Y	Y
Observations	45428	47399	52090	56295
Facilities	5565	5941	7005	7602
Adopters	298	306	311	319
Adjusted R-squared	0.14	0.31	0.27	0.27

Standard errors clustered by facility in brackets. Unadjusted standard errors in parentheses.
* p<0.10, ** p<0.05, *** p<0.01.
[a] p=0.11

Table A-4. Difference-in-Differences Results: Early vs. Late Adopters

Model:	**Facility-wide emissions**			**Facility-wide health hazard**			**Pollution intensity – emissions**			**Pollution intensity – health hazard**		
Dependent variable	Pounds of emissions			Health hazard			Pounds of emissions			Health hazard		
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)	(11)	(12)
Sample	A	B	C	A	B	C	A	B	C	A	B	C
Post-certification, early adopters	-0.63	-0.61	-0.51	-0.36	-0.71	-0.47	-0.47	-0.54	-0.50	-0.08	-0.62	-0.36
	[0.20]***	[0.25]**	[0.21]**	[0.30]	[0.36]**	[0.33]+	[0.20]**	[0.22]**	[0.19]**	[0.30]	[0.31]**	[0.29]
Post-certification, late adopters	-0.19	0.05	0.34	0.17	-0.14	0.14	-0.19	-0.08	0.15	0.05	-0.27	-0.47
	[0.12]	[0.17]	[0.15]**	[0.19]	[0.27]	[0.24]	[0.14]	[0.15]	[0.15]	[0.20]	[0.25]	[0.24]*
Production index							Y	Y	Y	Y	Y	Y
Facility fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year or counter fixed effects	Year	Counter	Counter	Year	Counter	Counter	Year	Counter	Counter	Year	Counter	Counter
Matched group dummies × counter		Y	Y		Y	Y		Y	Y		Y	Y
N	129835	22655	13299	130270	22664	13304	66588	19924	11602	66588	19927	11602
Facilities	12077	1979	1770	12118	1980	1770	7627	1979	1764	7627	1980	1764
Adopters	860	412	380	860	412	380	522	412	374	522	412	374
Adjusted R-squared	0.69	0.71	0.78	0.67	0.66	0.81	0.76	0.77	0.81	0.72	0.70	0.84
Test of equal effects: early vs late adopters												
H_0: Early adopters = Late adopters												
F statistic	3.45*	4.52**	11.27***	2.27	1.62	0.72	1.35	5.33**	7.52***	0.12	0.78	0.08

Early Adopters refers to facilities that adopted 1996-1999.
Late Adopters refers to facilities that adopted 2000-2002.
See Table 3 for description of Samples A, B, and C.
Standard errors clustered by facility in brackets.
*** p<0.01, ** p<0.05, * p<0.10 for a two-tailed test. + p<0.10 for a one-tailed test

References for Appendix

Bansal, P. & Hunter, T. 2003. Strategic explanations for the early adoption of ISO 14001. ***Journal of Business Ethics***, 46(3): 289-299.

Baron, D. P. 2003. Private politics. ***Journal of Economics & Management Strategy***, 12(1): 31-66.

Carraro, C., Katsoulacos, Y., & Xepapadeas, A. (Eds.). 1996. ***Environmental policy and market structure***. Boston: Kluwer Academic Publishers.

Christmann, P. & Taylor, G. 2001. Globalization and the environment: determinants of firm self-regulation in China. ***Journal of International Business Studies***, 32(3): 439-458.

Delmas, M. & Toffel, M. W. 2004. Stakeholders and environmental management practices: An institutional framework. ***Business Strategy and the Environment***, 13(4): 209-222.

Delmas, M. A. 2000. Barriers and incentives to the adoption of ISO 14001 by firms in the United States. ***Duke Environmental Law & Policy Forum***, 11(1): 1-38.

Delmas, M. A. 2002. The diffusion of environmental management standards in Europe and the United States: An institutional perspective. ***Policy Sciences***, 35: 91-119.

Florida, R. & Davison, D. 2001. Gaining from green management: Environmental management systems inside and outside the factory. ***California Management Review***, 43(3): 64-84.

Hamilton, J. T. 1999. Exercising property rights to pollute: Do cancer risks and politics affect plant emission reductions? ***Journal of Risk and Uncertainty***, 18(2): 105-124.

Helland, E. 1998. The enforcement of pollution control laws: Inspections, violations, and self-reporting. ***The Review of Economics and Statistics***, 80(1): 141-153.

Henriques, I. & Sadorsky, P. 1996. The determinants of an environmentally responsive firm: an empirical approach. ***Journal of Environmental Economics and Management***, (May): 381.

Khanna, M. & Anton, W. Q. 2002. Corporate environmental management: Regulatory and market-based pressures. ***Land Economics***, 78(4).

Khanna, M. & Damon, L. 1999. EPA's voluntary 33/50 program: Impact on toxic releases and economic performance of firms. ***Journal of Environmental Economics and Management***, 37: 1-25.

King, A. A., Lenox, M. J., & Terlaak, A. Forthcoming. Strategic use of decentralized institutions: Exploring certification with the ISO 14001 management standard. ***Academy of Management Journal***.

Lawrence, A. T. & Morell, D. 1995. Leading-edge environmental management: Motivation, opportunity, resources and processes. In D. Collins, & M. Starik (Eds.), ***Special Research Volume of Research in Corporate Social Performance and Policy, Sustaining the Natural Environment: Empirical Studies on the Interface Between Nature and Organizations***: 99-126. Greenwich, CT: JAI Press.

Majumdar, S. K. & Marcus, A. A. 2001. Rules versus discretion: The productivity consequences of flexible regulation. ***Academy of Management Journal***, 44(1): 170-179.

Maxwell, J. W., Lyon, T. P., & Hackett, S. C. 2000. Self-regulation and social welfare: The political economy of corporate environmentalism. ***The Journal of Law & Economics***, 43(2): 583-619.

Nash, J., Ehrenfeld, J., MacDonagh-Dumler, J., & Thorens, P. 2000. ***ISO 14001 and EPA'S Region 1's Startrack Program: Assessing their Potential as Tools in Environmental Protection***. Washington DC: National Academy of Public Administration.

Raines, S. S. 2002. Implementing ISO 14001--An international survey assessing the benefits of certification. ***Corporate Environmental Strategy***, 9(4): 418-426.

Rugman, A. M. & Verbeke, A. 1998. Corporate strategies and environmental regulations: An organizing framework. ***Strategic Management Journal***, 19(4): 363-375.

Vidovic, M. & Khanna, N. 2003. ***Reducing Toxic Releases through Voluntary Efforts: Incentives for Participation in the 33/50 Program and Implications for the Distribution of Pollution***. Paper presented at the Western Economic Association International 78th Annual Conference, Denver, Colorado.

Welch, E. W., Mazur, A., & Bretschneider, S. 2000. Voluntary behavior by electric utilities: levels of adoption and contribution of the climate challenge program to the reduction of carbon dioxide. ***Journal of Policy Analysis and Management***, 19: 407–425.