DISCUSSION PAPER

August 2011 ■ RFF DP 11-39

Does Eco-Certification Boost Regulatory Compliance in Developing Countries?

ISO 14001 in Mexico

Allen Blackman

1616 P St. NW
Washington, DC 20036
202-328-5000 www.rff.org

RESOURCES FOR THE FUTURE

Does Eco-Certification Boost Regulatory Compliance in Developing Countries? ISO 14001 in Mexico

Allen Blackman

Abstract

Private sector initiatives certifying that producers of goods and services adhere to defined environmental process standards are increasingly popular worldwide. According to proponents, they can circumvent chronic barriers to effective public sector environmental regulation in developing countries. But eco-certification programs will have limited effects on producers' environmental performance if, as one would expect, they select for those already meeting certification standards. Rigorous evaluations of the environmental effects of eco-certification in developing countries that control for selection bias are rare. We use plant-level data on more than 80,000 Mexican facilities to determine whether ISO 14001 series certification of environmental management systems boosts regulatory compliance. We use propensity score matching to control for nonrandom selection into the program. We find that plants recently fined by environmental regulators were more likely to be certified, all other things equal, but that certified plants were subsequently fined just as often as similar uncertified plants. These results suggest that in Mexico, the ISO 14001 program attracts dirty plants under pressure from regulators—not just relatively clean ones—but does not have a large, lasting impact on their regulatory compliance.

Key Words: voluntary environmental regulation, duration analysis, propensity score matching, Mexico

JEL Classification Numbers: Q56, Q58, O13, O54, C41

Contents

Does Eco-Certification Boost Regulatory Compliance in Developing Countries? ISO 14001 in Mexico

Allen Blackman*

1. Introduction

Private sector initiatives certifying that producers of goods and services adhere to defined environmental process or performance standards are increasingly popular. In the past two decades alone, more than 300 initiatives have been launched in a wide range of countries and economic sectors (Ecolabel Index 2011). These include Forest Stewardship Council certification for managed forests, Rainforest Alliance certification for coffee farms, and International Organization for Standardization (ISO) 14001 series certification for manufacturing plants and other businesses. In theory, programs like these generate environmental benefits by widening the availability and improving the accuracy of information about individual producers' environmental performance, thereby enabling consumers, capital markets, communities, and regulators to more easily reward clean producers and sanction dirty ones. For example, armed with better information, consumers can buy or boycott products, and lenders can extend or withhold credit.

According to proponents, eco-certification holds special promise as an environmental management tool for developing countries (Dingwerth 2008; World Bank 2000). Although acute environmental problems are now common in these countries, conventional command-and-control regulation generally has failed to adequately address them. The reasons are well known: environmental regulatory agencies are undermanned and underfunded, written regulations are riddled with gaps, and the political will for strict enforcement is lacking. As a result, regulatory noncompliance is widespread (Russell and Vaughan 2003; Eskeland and Jimenez 1992). In

* Resources for the Future, 1616 P Street NW, Washington, DC 20036, (202) 328-5073, blackman@rff.org; and Environment for Development, Center for Central America,Turriabla, Costa Rica.

Partial funding for this research was provided by the Swedish Research Council Formas. I'm grateful to Santiago Guerrero, Bidisha Lahiri, Marisol Rivera, and Carlos Muñoz for help obtaining, cleaning, and assembling the data used for this study, and to Bidisha Lahiri, Billy Pizer, Juha Siikamaki, Carlos Chavez, Dietrich Earnhardt, Bill Lanen, and seminar participants at the University of Michigan, Kansas University, University of California–Santa Barbara, University of Wisconsin–Madison, and Triangle Research Environmental Economics Seminar for helpful comments and suggestions.

theory, eco-certification can sidestep these constraints by creating a private sector system of incentives, monitoring, and enforcement.

Yet credible empirical evaluations of the environmental effects of eco-certification in developing countries are rare. One reason is that requisite producer-level data are scarce. A second reason is that to be credible, an evaluation—whether of a developing or industrialized country program—must control for the nonrandom selection of certain types of producers into certification. Relatively clean producers often have strong incentives to obtain certification: the costs are relatively low because few additional investments are required to meet the certification criteria, and the benefits, such as enhanced sales, can be significant. An evaluation that failed to control for such selection would conflate the effect on environmental performance of certification with the effect of certified producers' preexisting characteristics. A recent review of the literature evaluating the environmental effects of eco-certification in five economic sectors where it is prevalent found only two published studies that controlled for selection bias, neither of which focused on a developing country program (Blackman and Rivera In Press).

The much larger literature on eco-certification and on (closely related) voluntary environmental programs (VEPs) in industrialized countries is certainly relevant.[1] But findings from this literature may not generalize to developing countries, for at least two reasons. First, the socioeconomic context in developing countries is clearly different. Second, the aims both of eco-certification and VEPs in developing countries are different. In industrialized countries, these policies are used to encourage firms to overcomply with mandatory regulations—that is, to perform above and beyond legal requirements. In developing countries, by contrast, the hope is that they will help remedy noncompliance with mandatory regulation (Blackman 2008, 2010). Hence, additional research is needed to determine whether eco-certification can improve environmental performance in developing countries.

To help fill that gap, this paper examines the ISO 14001 series voluntary certification program for environmental management systems (EMSs) in Mexico. To our knowledge, it is the first econometric analysis of the environmental effects of ISO 14001 certification in a developing country and only the second econometric analysis of the environmental effects of a VEP in a

[1] Like eco-certification programs, VEPs invite firms to meet defined environmental process or performance criteria. However, VEPs do not always issue a formal certification. In addition, they are often administered by state regulatory authorities, not private entities. See Koehler (2008), Lyon and Maxwell (2002), and Khanna (2001) for reviews.

developing country. It builds on two recent studies that use some of the same data.[2] These data were constructed by merging plant-level registries compiled by the Mexican Ministry of Economics, the Mexican Federal Environmental Attorney General's Office (*Procuraduría Federal de Protección al Ambiente*, PROFEPA, the enforcement branch of the Ministry of the Environment), and two private sector companies.

The focus of our empirical analysis is determining whether ISO 14001 certification boosts environmental regulatory compliance. To control for nonrandom selection into certification—including nonrandom selection of already clean plants free-riding on unrelated environmental management investments—we use a two-stage strategy. In the first stage, we estimate a duration model of ISO 14001 certification that includes as independent variables lagged PROFEPA fine dummies measuring past regulatory compliance. In the second stage, we use propensity score matching to assess the effect of the ISO 14001 certification on subsequent regulatory compliance; that is, we compare the average annual incidence of PROFEPA fines for certified plants and matched uncertified plants over an outcome period that begins after certification. We use the duration model estimated in the first stage to generate propensity scores used to match certified and uncertified plants. The combination of a first-stage duration model and a second-stage propensity score matching analysis has two attractive features. The duration model accounts for the timing of the dependent variable (certification) and the main independent variables of interest (lagged fine dummies) and controls for right censoring. The propensity score matching model provides a convenient means of testing restrictions on the composition of the control group of uncertified plants.

Our first-stage duration analysis indicates that plants recently fined by PROFEPA were more likely to obtain ISO 14001 certification, all other things equal. However, the second-stage matching analysis shows that after certification, participants were fined just as often as matched nonparticipants. These results suggest that although the program attracted relatively dirty plants, not just already clean ones, it did not have a large lasting impact on their regulatory compliance.

The remainder of the paper is organized as follows. Section 2 provides background on ISO 14001 certification. Section 3 briefly reviews the relevant literature on ISO 14001

[2] Blackman et al. (2010) analyzes Mexico's flagship VEP, the Clean Industry Program, and Blackman and Guerrero (2010) analyzes the drivers—but not the environmental impacts—of ISO 14001 certification in Mexico.

certification. Section 4 discusses our data. Sections 5 and 6 present our models of certification and effects. The last section sums up and discusses policy implications.

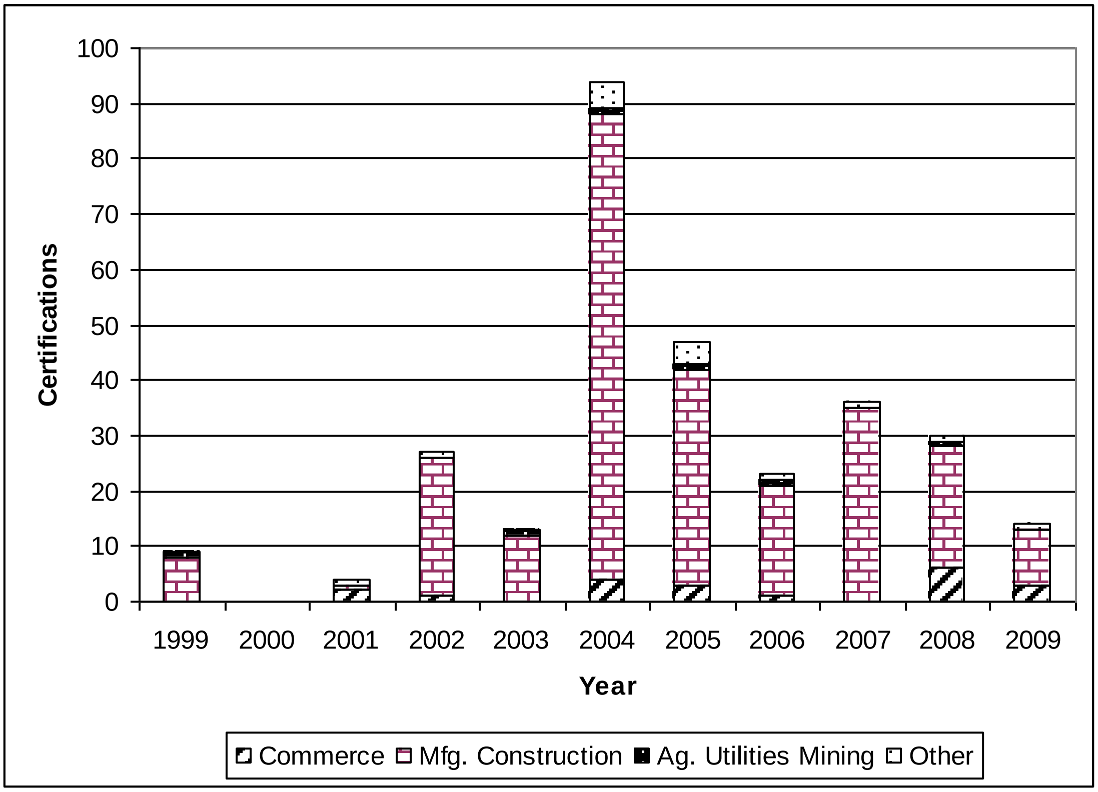
2. Background

A nongovernmental organization, ISO is the world's largest developer and publisher of international standards. It issued its 14001 series certification for EMSs in 1996 and revised it in 2004. By December 2007, more than 150,000 facilities in 148 countries had been certified (ISO 2009). To obtain ISO 14001 certification, which is valid for three years, plants must meet five criteria that together comprise a "plan-do-check-act" cycle. They must define an environmental management strategy, make a concrete plan to implement it (plan), implement the policy and document the results (do), conduct periodic internal performance audits (check), and take corrective action to promote continual improvement (act). An independent third-party auditor approved by ISO must verify that these criteria have been met. Plants need not meet hard performance targets to be certified: in general, certification requirements focus on process rather than performance. A supposed requirement for ISO 14001 certification is compliance with all applicable local environmental regulations. However, the stringency with which plants are held to this (and other) ISO 14001 criteria varies across countries and within them (Nel and Wessels 2010). The costs of certification can be quite substantial. They include the costs of third-party audits, creating a new EMS or modifying an existing one, and continually implementing that EMS. In the United States, audit costs range from \$239 to \$1,372 per employee, and implementation costs range from \$29 to \$88 per employee (Darnall and Edwards 2006).

In Mexico, ISO 14001 certification grew from zero plants in 1998 to 595 by the end of 2009 (*Contacto* 2005). However the majority these 595 plants were state-owned electricity-generating facilities (owned by *Comisión Federal Electricidad*) and petroleum refineries and distribution terminals (owned by *Petróleos Mexicanos*). We do not include these plants in our analysis because our data on plant characteristics (described in Section 4) are restricted to private sector facilities. In addition, the drivers of ISO 14001 certification and regulatory compliance are likely different for public and private sector plants. Figure 1 presents data on ISO 14001 certification by sector and year between 1999 and 2009 for the sample of 298 plants included in our empirical analysis.[3] It shows that most certifications in this period were in the manufacturing sector and occurred in 2004 and 2005.

[3] The selection of the sample is discussed in Section 4.

Figure 1. Within-Sample New ISO 14001 Certifications in Mexico, by Year and Sector (n=298)



3. Literature

This section briefly reviews the literature on ISO 14001 certification in developing and industrialized countries. It focuses on findings about (i) the effects of past environmental performance on the probability of certification and (ii) the effects of certification on environmental performance.

3.1. Developing-Country Studies

As noted above, to our knowledge, the present paper is the first econometric analysis of the environmental effect of ISO 14001 certification in a developing country. However, three other econometric studies have examined the drivers of ISO 14001 certification in developing countries: Christmann and Taylor (2001), Montiel and Husted (2009), and Blackman and Guerrero (2010). All three find that large plants trading in international markets are more likely to be certified. Only one study, Blackman and Guerrero (2010), examines the effect of past environmental performance on the probability of certification. Using a shorter version of the panel data used for the present study (that extends through 2005), they find plants that were recently fined were more likely to be certified.

3.2. Industrialized-Country Studies

3.2.1. Drivers of ISO 14001 Certification

Several papers examine the drivers of ISO 14001 certification in industrialized countries. However, only a handful—all focused on the United States—test for the effect on certification of past environmental performance or regulatory pressure. Most, but not all, find a positive correlation. For example, Darnall (2003) finds that plants in violation of hazardous waste and atomic energy regulations are more likely to be certified; King et al. (2005) find that plants with toxic emissions higher than the average for their size and sector are more likely to be certified; and Arimura et al. (2008) find that Japanese manufacturing plants subject to environmental performance standards and input taxes are more likely to be certified. However, Potoski and Prakash (2005a) find an inverted-U relationship between compliance with Clean Air Act regulations and certification: plants never in compliance and those always in compliance are more likely to be certified than those that are sometimes in compliance.

As for nonregulatory drivers of certification, virtually all econometric studies of ISO 14001 adoption find that firm size is positively correlated with certification (Arimura et al. 2008; King et al. 2005; Nakamura et al. 2001; Nishitani 2009; Potoski and Prakash 2005a). And virtually all studies conclude that sales to foreign buyers also are positively correlated with certification (Arimura et al. 2008; Bansal and Hunter 2003; King et al. 2005; Nishitani 2009).

3.2.2. ISO 14001 Certification Environmental Impacts

The evidence on whether ISO 14001 certification improves plants' environmental performance is mixed: results depend on the measure of environmental performance and the location and type of plants studied. On one hand, Arimua et al. (2008) find that in the Japanese manufacturing sector, ISO 14001 certification spurs reductions in natural resource use and waste emissions. Using cross-sectional survey data from seven countries that are members of the Organization for Economic Co-operation and Development, Darnall and Kim (in press) reach similar conclusions. And Potoski and Prakash (2005a, 2005b) find that ISO 14001 certification causes U.S. air polluters to cut their emissions by more than uncertified plants and to spend less time out of compliance with Clean Air Act regulations. On the other hand, however, using detailed panel data for a sample of Canadian pulp and paper plants, Barla (2007) finds that ISO 14001 certification is correlated with increases in some types of water pollution. And using U.S. panel data, King et al. (2005) find that although adoption of an EMS spurs reductions in toxic emissions, ISO 14001 certification of the EMS does not have an additional impact.

4. Data

Official Mexican plant-level census data are not publicly available. We constructed our own plant-level data set from four sources. The first is a list of all 595 Mexican facilities that obtained ISO 14001 certification through December 2009, along with the plant's location and the year (but not date) of certification. This list was compiled from an annual registry published by *Contacto* magazine.[4]

The second data source is the July 2004 System of Mexican Business Information (*Sistema de Información Empresarial Mexicano*, SIEM). The Mexican Ministry of Economics compiles and maintains SIEM and uses it to promote Mexican commerce. The database is constantly updated to include new entrants and omit plants that have exited the market. It is not time specific; that is, it does not indicate when plants entered SIEM or whether their characteristics subsequently changed. SIEM contains basic information on geographic location, sector, gross sales, accounting capital, and whether the facility exports, imports, and is a government supplier. Our raw SIEM data consist of 528,618 records. However, to limit our subsample of uncertified facilities to those types of plants that had a proven history of ISO 14001 certification, we dropped all plants in sectors (defined by North American Industrial Classification codes) that were not also represented in the ISO 14001 database. This step eliminated approximately 80 percent of the plants in the SIEM data.

Our third data source is a 2009 registry of fines levied by PROFEPA, which is responsible for monitoring and enforcement of most environmental regulations. This registry contains records of every PROFEPA fine between January 1992 (the year PROFEPA was created) and December 2009, including the amount and date of the fine. It has data on 44,008 plants.

[4] Unlike lists of Mexican ISO 14001 plants available on Internet sites, *Contacto* includes plant location data needed to merge our four data sets.

Our final data source is a 2007 list of 3,850 Mexican *maquiladoras* obtained from a commercial registry (Mexico's Maquila Online Directory 2007). Designed to take advantage of relatively inexpensive Mexican labor, *maquiladoras* are (mostly foreign-owned) assembly plants that import inputs and export outputs without paying tariffs or duties.

Because the four databases do not have a common numerical code identifying individual plants, we merged them by nonnumerical identifiers—plant name, state, and *municipio* (county).[5] These nonnumeric data did not uniquely identify plants. For example, in the SIEM data, multiple records have the same plant name, state, and *municipio.* To avoid incorrectly matching records, we dropped all records that were not uniquely identified. This resulted in a loss of 5 to 20 percent of the records in each data set. The end result was a sample of 80,611 plants, of which 298 were ISO 14001 certified and 80,313 were not.

5. First-Stage Certification Model

This section describes our first stage model focusing first on the independent variables used to explain ISO 14001 certification, then on our empirical strategy, and finally on our results.

5.1. Independent Variables

5.1.1. Time-Varying Independent Variables: Fines

Of the potential drivers of ISO 14001 certification, we are particularly interested in past environmental performance and the consequent threat of regulatory sanctions since, as discussed above, we need to control for the possibility that already clean plants free-riding on unrelated investments in pollution control self-select into certification. As a proxy, we use four dichotomous dummy variables indicating how recently the plant has been fined. They are the only time-varying independent variables in our certification model. FINE_1YR identifies plants fined in the year before the current year, FINE_2YR identifies plants fined two years before, FINE_3YR identifies plants fined three years before, and FINE_3YR_PLUS identifies plants fined more than three years before. We observe only the year, not the date, of ISO certification. Hence, we do not include a fine dummy variable for the current year to avoid conflating cases where a fine precedes certification with cases where it follows certification.

[5] Using finer geographical identifiers (e.g., city) proved impractical because of a lack of uniformity across the databases.

To motivate these variables, Table 1 present summary statistics on PROFEPA fines between 1992 and 2009. In our entire sample of 80,611 plants, 2 percent were fined. Of the 1,468 plants that were fined, the average number of fines was two per plant, and the average fine was 135,000 pesos (approximately US$13,500).

ISO 14001 certified plants were fined far more often uncertified plants: 34 percent of certified plants were fined versus only 2 percent for uncertified plants. Hence, there appears to be a simple correlation between fines and ISO 14001 certification. However, this correlation does not necessarily imply causation, for at least two reasons. First, it may have been generated by underlying differences in plant characteristics. For example, it could simply reflect a tendency for large plants to be fined and also to obtain ISO 14001 certification. Second, it does not take into account the temporal relationship between these events. For example, it lumps together cases in which a fine was followed by certification 1 year later and cases in which a fine was followed by certification 10 years later, even though the former are more likely to represent actual causation. As discussed below, our first-stage model of certification addresses both of these issues: it controls for a variety of underlying plant characteristics and takes into account the temporal relationship between fines and certification.

Table 1. PROFEPA Fines, 1992–2009

Sample ➔ ↓		***All***	***ISO 14001 certified***	***Uncertified***
All plants		(n=80,611)	(n=298)	(n=80,313)
	Fined	1.82%	33.56%	1.70%
Fined plants		(n=1,468)	(n=100)	(n=1,368)
	Total no. fines	2,881	209	2,672
	Average no. fines/plant	1.96	2.09	1.95
	Average size fine (pesos)	134,702.80	90,961.41	137,900.30

5.1.2. Time-Invariant Independent Variables

Table 2 lists the time-invariant independent variables in the first-stage certification model and presents sample means for all plants in the sample, for certified plants, and for uncertified plants. Because our main focus is the effect of PROFEPA fines on certification, these variables mainly serve as controls.[6] A few notes about the sector and location fixed effects are in order. We include five sector fixed effects dummies drawn from the 17 sector categories in the SIEM data set. We omit dummies for the remaining 12 sectors either because all observations in these sectors were dropped from the regression sample (see Section 4) or because these dummies were perfectly or near perfectly correlated with the dependent variable; that is, they correspond to sectors in which either zero or a handful of the plants in our sample were certified.[7] We include sixteen location fixed effects dummies that identify the state where each plant is located. Dummies for the 16 other states in Mexico were dropped because they were perfectly or nearly perfectly correlated with the dependent variable.[8]

[6] See Blackman and Guerrero (2010) for a more in-depth discussion of these variables and their expected influence on the probability of certification.

[7] The model includes dummies for Sector 2 (commercial wholesale), Sector 3 (commercial retail), Sector 4 (construction), Sector 6 (industrial manufacturing), and Sector 15 (real estate services). We dropped dummies for Sector 5 (electricity, water, and gas), Sector 11 (waste management, remediation), Sector 12 (entertainment, culture, and sports), Sector 13 (health and social assistance), and Sector 14 (educational services) because all observations in these sectors were eliminated from the regression sample. We dropped dummies for Sector 1 (agriculture, livestock, forestry, fishing, and hunting), Sector 7 (information and mass media), Sector 8 (mining), Sector 9 (other services except government), Sector 10 (temporary lodging, food and beverage preparation), Sector 16 (professional, scientific, and technical services), and Sector 17 (transport and mail) because they were perfectly or near-perfectly correlated with our dependent variable.

[8] The model includes dummies for Aguascalientes, Baja California, Chihuahua, Coahuila, Distrito Federal, Guanajuato, Jalisco, Mexico, Nuevo Leon, Puebla, Queretaro, San Luis Potosi, Tamulipas, Tlaxcala, Veracruz, Yucatan. We dropped dummies for Baja California Sur, Campeche, Chiapas, Colima, Durango, Guerrero, Hidalgo, Michoacan, Morelos, Nayarit, Oaxaca, Quintana Roo, Sinaloa, Sonora, Tabasco, and Verecruz.

Table 2. Time-Invariant Independent Variables and Sample Means (%)

Variable	*Explanation* (*all are 0/1 dummy variables*)	*All* (*n=80,611*)	*IS0 14001 certified* (*n=298*)	*Uncertified* (*n=80,313*)
EXPORT	exporter	8.990	73.154	8.752
IMPORT	importer	12.915	78.523	12.672
GSUPPLIER	government supplier	11.552	14.765	11.540
MAQUILA	*maquiladora*	0.803	33.893	0.680
SA_0_3M	gross revenue 0–3 million pesos	85.540	23.826	85.769
SA_3M_12M	gross revenue 3–12 million pesos	7.366	7.047	7.367
SA_12M_PLUS	gross revenue 12 million pesos +	7.093	69.128	6.863
CAP_0_900	accounting capital 0–900K pesos	81.138	26.175	81.342
CAP_901_5M	acc. capital 900K–5 million pesos	9.620	7.718	9.627
CAP_5M_PLUS	acc. capital 5 million pesos +	9.242	66.107	9.031
SECTOR_*	5 sector fixed effects	--	--	--
STATE_*	16 state fixed effects	--	--	--

5.2. Empirical Strategy: Duration Analysis

To analyze ISO 14001 certification, we use a duration model, which estimates a hazard rate, h, interpreted as the conditional probability that a plant obtains ISO 14001 certification at time t, given that it has not already been certified, and given the characteristics of the plant at time t, including its history of fines.[9] More formally,

$$h(t, \mathbf{X_t}, \boldsymbol{\beta}) = f(t, \mathbf{X_t}, \boldsymbol{\beta})/(1- F(t, \mathbf{X_t}, \boldsymbol{\beta})) \qquad (1)$$

where $F(t, \mathbf{X_t}, \boldsymbol{\beta})$ is a cumulative distribution function that gives the probability that the plant was certified prior to time t, $f(t, \mathbf{X_t}, \boldsymbol{\beta})$ is its density function, $\mathbf{X_t}$ is a vector of explanatory variables related to the characteristics of the plant (some of which change over time), and $\boldsymbol{\beta}$ is a vector of parameters to be estimated. Following convention, the hazard rate is broken down into two components. The first is a baseline hazard, $h_0(t)$, that is a function solely of time (not of any explanatory variables) and that is assumed to be constant across all plants. It picks up any effects not captured by explanatory variables, such as the diffusion of knowledge about ISO 14001 certification or changes in macroeconomic conditions. The second component of the hazard rate

[9] For an introduction to the use of duration models in economics, see Kiefer (1988).

is a function of the vector of explanatory variables, $\mathbf{X_t}$. Combining these two components, the hazard rate is written

$$h(t) = h_0(t)\exp(\mathbf{X_t}'\boldsymbol{\beta}). \quad (2)$$

We estimate $\boldsymbol{\beta}$ using maximum likelihood. We use a Cox (1975) proportional hazard model because it does not require parametric assumptions about the density function.[10] We use years as our temporal unit of analysis. Although we know the day on which plants were fined, we know only the year in which plants were ISO 14001 certified. Finally, note that although plants can and do obtain ISO 14001 certification more than once, our duration model seeks to explain the decision to obtain certification for the first time. As is standard practice with duration models, observations (here, plants) are dropped from the regression sample after their first "failure" (here, certification).

A duration framework is appropriate for analyzing the effect of fines on ISO 14001 certification, for two reasons. First, it explicitly accounts for the temporal relationship between these phenomena, which (as discussed above) helps determine whether fines actually cause certification. Second, it avoids the problem of right censoring that would arise in a cross-sectional dichotomous choice (probit or logit) model if some plants that were not certified in December 2009 (when our panel ends) subsequently joined the program. A duration model circumvents this problem by estimating the conditional probability of certification in each period.

[10] There are two broad approaches to specifying duration models. One is to make parametric assumptions about the time dependence of the probability density function, $f(t, \mathbf{X_t}, \boldsymbol{\beta})$. Common assumptions include exponential, Weibull, and log-logistic distributions. Each assumption implies a different shape for the baseline hazard function, $h_0(t)$. For example, an exponential probability density function generates a flat hazard function, $h_0(t)$, which implies that the probability of obtaining ISO 14001 certification (apart from the influences of regulatory activity and plant characteristics) stays the same over time. A second general approach is to use a Cox (1975) proportional hazard model, which does not require a parametric assumption about the density function. This feature accounts for the popularity of the Cox model among economists.

5.3. Results

5.3.1. Time-Varying Independent Variable: Fines

Because the hazard function given by equation (2) is nonlinear, the estimated coefficients do not have a simple interpretation (technically, they can be interpreted as the effect on the log hazard rate of a unit change in the explanatory variable at time t). Exponentiated coefficients, however, can be interpreted as the hazard ratio—that is, the ratio of the hazard rate given an increase in an explanatory variable at time t (a unit increase in a continuous variable or a change from 0 to 1 of a dichotomous dummy variable) relative to the baseline hazard rate at time t. A hazard ratio greater than unity indicates that an increase in the explanatory variable increases the hazard rate relative to the baseline. For example, a hazard ratio of 2 means that an increase in the explanatory variable doubles the hazard rate relative to the baseline.

Three of our fine variables—FINE_1YR, FINE_2YR, and FINE_3YR—are significant at the 5 percent, 10 percent, and 1 percent level, respectively, and one—FINE_3YR_PLUS—is not significant (Table 3, Model A). The hazard rates for FINE_1Y, FINE_2YR, and FINE_3YR indicate that a fine assessed one year prior to the current year increases the probability that a plant will obtain ISO 14001 certification in the current year by a factor of 1.8, a fine assessed two years earlier increases this probability by a factor of 1.5, and fine assessed three years earlier, by a factor of 2.3. However, we are not able to reject the hypothesis that these three coefficients are equal. Therefore, the appropriate interpretation is that a fine within three years of the current year increases the probability of certification by roughly a factor of two. The lack of significance of FINE_3YR_PLUS indicates that a fine assessed more than three years prior to the current period does not affect the probability of certification.

Table 3. Cox Proportional Hazard Models of ISO 14001 Certification in Mexico, 1999–2009: Hazard Ratio (s.e.)

Variable	*Model A Full sample*	*Model B Large mfg. plants only*
FINE_1YR	1.781** (0.421)	1.692* (0.471)
FINE_2YR	1.501* (0.364)	1.272 (0.376)
FINE_3YR	2.250*** (0.468)	1.971*** (0.493)
FINE_3YR_PLUS	1.280 (0.232)	1.026 (0.230)
EXPORT	1.902*** (0.423)	1.766** (0.485)
IMPORT	2.573*** (0.622)	3.051*** (1.083)
GSUPPLIER	1.063 (0.197)	0.928 (0.213
MAQUILA	3.246*** (0.567)	2.3668*** (0.490)
SA_3M_12M	0.946 (0.265)	--
SA_12M_PLUS	2.205*** (0.497)	--
CAP_901K_5M	1.044 (0.275)	--
CAP_5M_PLUS	2.689*** (0.558)	--
Sector fixed effects	yes	--
State fixed effects	yes	yes
Total plants	80,611	2,278
ISO 14001 plants	298	184
Log likelihood	-2461.892	-1297.0627

Model A: full sample.
Model B: sample restricted to large manufacturing plants.
***Significant at 1% level.
**Significant at 5% level.
*Significant at 10% level.

A potential concern about our analysis is that fine variables could, in principle, be endogenous if they are correlated with unobserved plant characteristics that affect certification.[11] Although such endogeneity cannot be ruled out, it is unlikely to be driving the observed correlation between fines and certification. The reason is that endogeneity would be unlikely to generate the temporal response function implied by the results for our fine variables—namely, an effect of fines on the probability of certification that diminishes the more distant in time was the fine (i.e., after three years). Instead, endogeneity would generate a response that did not change over time. Hence, our results suggest a causal relationship between fines and ISO 14001 certification.

5.3.2. Time-Invariant Independent Variables

Estimated hazard ratios for EXPORT, IMPORT, and MAQUILA indicate that plants selling their goods in overseas markets were 1.9 times more likely to obtain ISO certification, those importing foreign inputs were 2.6 times more likely, and those that were *maquiladoras* were 3.2 times more likely, all other things equal (Table 3, Model A). The hazard ratio for GSUPPLIER is not significant. Estimated hazard ratios for the sales and capital dummies suggest that larger plants were more likely to join the program. Specifically, plants with more than 12 million pesos in sales were 2.2 times more likely to be certified, and plants with more than 5 million in capital were 2.7 times more likely to be certified.[12] For the sector fixed effects (not included in Table 3), estimated hazard ratios indicate that, compared with plants in reference sectors, those in Sector 6 (industrial manufacturing) were more likely to be certified, and those in Sector 3 (commercial retail) and Sector 4 (construction) were less likely.[13] Finally, for the state fixed effects (also not included in Table 3), our results indicate that, compared with plants in the reference states, those in Colima, Puebla, and Tamaulipas were more likely to be certified, and those in the Federal District and Jalisco were less likely.[14]

---

[11] For example, aside from our sector dummies, our covariates do not include a precise measure of the complexity of the production process, so complexity is partly unobserved. It could be that complex plants are more likely to be fined because they have a higher potential for violating environmental regulations and are also more likely to obtain ISO 14001 certification because they tend to employ educated and sophisticated managers. If this were actually true, then fines would be endogenous.

[12] The reference groups for these dummies are plants with less than 3 million pesos in sales and those with less than 900,000 pesos in capital.

[13] The reference group comprises plants in the seven sectors (among the 12 represented in our regression sample) in which ISO certification was limited (see footnote 7 for a list).

[14] The reference group comprises plants in 16 states in which ISO certification is limited (see footnote 8 for a list).

6. Second-Stage Impact Model

This section describes our second-stage model of the influence of ISO 14001 certification on regulatory compliance, focusing first on our empirical strategy, then on our results, and finally, on robustness checks.

6.1. Empirical Strategy: Propensity Score Matching

As noted above, in evaluating the effect of ISO 14011 certification on regulatory compliance, a key challenge is controlling for the nonrandom selection into certification of plants with characteristics likely to affect compliance (e.g., size, ties to overseas markets, lagged fines). To do that, we use a matching estimator. That is, following Rosenbaum and Rubin (1983) and more recently List et al. (2003) and Dehejia and Wahba (2002), we construct a matched control sample of uncertified plants with observable characteristics that are very similar to those of certified plants, and we measure program impact as the difference between compliance rates for ISO 14001–certified plants and for this matched control sample. Hence, compliance rates for the matched uncertified plants proxy for the unobserved counterfactual: what compliance rates for certified plants would have been had they not been certified.[15] More specifically, our impact measure—the average treatment effect on the treated (ATT)—is the difference between the average number of fines incurred per year by participants and by a matched sample of nonparticipants:

$$\frac{\sum_{i} F(i, J(i)+1, T)}{N_i} - \frac{\sum_{i'} F(i', J(i)+1, T)}{N_{i'}} \qquad (3)$$

[15] This approach depends on two identifying assumptions. The first assumption is that conditional only on agents' observed characteristics, the certification decision is ignorable for purposes of measuring outcomes. That is, we are able to control for all confounding variables that simultaneously affect the certification decision and the outcome variable. This first assumption is untestable. The second assumption, "common support" or "overlap," is that the distribution of observed characteristics for uncertified plants is similar to that for certified plants, such that plants with similar characteristics have a positive probability of being certified and of being uncertified.

where

i and i'	are a matched pair of ISO 14001–certified and uncertified plants;
F(J(i)+1,T)	is the average number of fines levied on facility i each year from year J(i)+1 through year T;
J(i)	is the year during which participant i obtains ISO 14001 certification;
T	is 2009, the last year of our fines panel; and
N_i and $N_{i'}$	are the number of certified and uncertified plants in the matching analysis sample.

For each matched pair, the first year of the period over which the annual average is calculated—the "outcome period"—is the first year after the participant obtained ISO 14001 certification, and the last year is 2009, the last year of our fines panel. The one-year lag ensures that all fines in the outcome period occurred after the plant was certified.

Creating a large set of matched pairs of plants with the exact same observed characteristics is challenging when, as in our case, these characteristics are numerous. However, Rosenbaum and Rubin (1983) demonstrate that we need to match plants only on the basis of their propensity score—that is, their likelihood of participation as predicted by a regression model. We use the duration model of program participation presented above to generate propensity scores for each certified and uncertified plant in our sample for each year.[16] We match ISO 14001–certified plants with uncertified plants using propensity scores for the year the former obtained certification. For each year of our study period, we prohibit matches between plants certified in that year and plants certified in later years.[17]

---

[16] Other program evaluations in which the timing of program participation plays an important role have relied on duration models to generate propensity scores. See, for example, Brodaty et al. (2001), Sianesi (2004), and Pizer et al. (2008).

[17] Take, for example, a plant certified in 2000. In principle, it could be matched to another plant that was not certified in 2000 but that was certified in 2001. In essence, such matches amount to pairing treated plants with other treated plants.

Various methods are available to match participants and nonparticipants based on propensity scores (Caliendo and Kopeinig 2008; Morgan and Harding 2006). To ensure robustness, we report results from five: nearest neighbor 1-to-1 matching, wherein each certified plant is matched to the uncertified plant with the closest propensity score; nearest neighbor 1-to-4 matching, wherein each certified plant is matched to the four uncertified plants with the closest propensity scores and the counterfactual outcome is the average across these four; nearest neighbor 1-to-8 matching; nearest neighbor 1-to-16 matching; and kernel matching, wherein a weighted average of all nonparticipating plants is used to construct the counterfactual outcome. For all five models we enforce a common support and allow matching with replacement.

We performed balance tests for the five matching estimators. In general, the balance is quite good. In the case of two of the five matching estimators (nearest neighbor 1-1 and 1-4), for all 33 covariates, the mean for the subsample of certified plants is not significantly different from the mean for the subsample of matched uncertified plants. For the other three matching estimators, the means for the two subsamples are significantly different only for one or two covariates. Table 4 reports median standardized bias (MSB)—Rosenbaum and Rubin's (1983) balance statistic—across all covariates for each matching estimator.[18] The highest MSB is 4.8 percent for the nearest neighbor 1-8 estimator, and the lowest is 3.2 percent for the kernel estimator (Table 4, Model I). Although a clear threshold for acceptable balance does not exist, according to Caliendo and Kopeinig (2008), standardized bias below 3 to 5 percent is generally viewed as sufficient. These encouraging balance statistics are likely due to the fact that, although our regression model includes 33 explanatory variables, our data set includes 269 uncertified plants for each certified plant, making it easier to find close matches for each certified plant.

[18] For a given covariate, the standardized bias is the difference of means in the treatment (here ISO 14001–certified) and control (here uncertified) subsamples as a percentage of the square root of the average sample variance in both subsamples. We report the median standardized bias (MSB) for all covariates. MSB before matching is 24.365 percent.

Table 4. Average Treatment Effect on Treated (ATT) Estimates and Median Standardized Bias (MSB) by Propensity Score Matching Method and Model[a]

Method	*Model I Main model*		*Model II Exact matching on fines*		*Model III Large mfg. plants only*	
	ATT[b]	***MSB***[c]	***ATT***[b]	***MSB***[c]	***ATT***[b]	***MSB***[d]
Nearest neighbor 1-1	0.001 (0.008)	3.469	0.008 (0.014)	6.034	0.009 (0.011)	3.220
Nearest neighbor 1-4	0.001 (0.006)	3.880	-0.001 (0.007)	4.018	0.007 (0.008)	4.671
Nearest neighbor 1-8	-0.001 (0.005)	4.778	-0.009 (0.006)	4.837	0.001 (0.007)	4.834
Nearest neighbor 1-16	-0.003 (0.005)	4.373	-0.007 (0.005)	4.380	-0.002 (0.006)	2.820
Kernel	-0.001 (0.005)	3.187	-0.006 (0.004)	2.695	-0.004 (0.006)	2.645

Model I: treatment and control plants matched on hazard rate for the year of certification.
Model II: same as I but with exact matching on 0/1 dummy = 1 if plant ever fined.
Model III: same as I but with sample restricted to large manufacturing plants.
[a]Average annual incidence of fines among treatment group (ISO-certified plants) for full sample (Models I and II) = 0.023 fines/year and for restricted sample (Model III) = 0.024 fines/year.
[b]Standard errors (in parentheses) computed using bootstrap with 500 repetitions, except kernel estimator, which uses 250 repetitions.
[c]For a given covariate, the standardized bias is the difference of means in the ISO 14001–certified and uncertified subsamples as a percentage of the square root of the average sample variance in both groups. We report the median standardized bias (MSB) for all covariates. Models I–II: MSB before matching is 24.365%; Model III: MSB before matching is 17.147%.

Calculating standard errors for ATT is not straightforward because they should, in principle, account for the fact that propensity scores are estimated and for the imputation of the common support (Heckman et al. 1998). Therefore, we bootstrap standard errors using 500 replications for the nearest neighbor estimators and 250 for the kernel estimator (which is exceptionally computationally intensive).

6.2. Results

Our results are not consistent with the hypothesis that ISO 14001 certification boosts regulatory compliance (Table 4, Model I). For all five matching estimators, ATT is not significantly different than zero. That is, there is not a statistically significant difference between the average annual incidence of fines for certified plants and for matched uncertified plants.

Although these results suggest that ISO 14001 certification did not have a lasting effect on regulatory compliance, a caveat is in order. The limited number of observations and variability in our outcome data limit the power of our hypothesis test. As a result, we cannot infer that ISO 14001 certification had absolutely no negative effect on the rate at which participants were subsequently fined, only that it did not have a *large* negative effect. Specifically, we can reject the hypothesis that the program reduced this fine rate by more than 42 percent (ATT ≤ 42% of certified plants' fine rate).[19]

6.3. Robustness

This subsection discusses three reasons that our ATT estimates could be biased downwards and presents robustness checks that address two of these reasons. The first is inadequate matching on key plant characteristics that might drive our outcome variable. For example, if the lagged fine variables in the duration analysis are particularly important determinants of future PROFEPA fines, and if certified plants that were not previously fined are matched to uncertified plants that were, then our ATT estimates could be biased. To address this concern, we estimate an "exact matching" model in which plants that were fined at some point during our panel are matched to other plants that were fined at some point, and vice versa. The results are virtually identical to those presented above (Table 4, Model II). We also estimated models with exact matching on dummy variables indicating whether the plant exports, is a *maquiladora,* has more than 30 million pesos in sales, and is in the manufacturing sector. Again, the results, are qualitatively identical to those for Model I (*see Table A1, at end*).

A second reason that our ATT estimates could be biased downward is if PROFEPA monitoring were more stringent for ISO 14001–certified plants than for matched uncertified plants. This might happen if (i) PROFEPA targets the *types* of plants that tend to obtain ISO

---

19 To see this, first note that the t-score used to test hypotheses about ATT is $t = \left(\overline{X}_T - \overline{X}_C\right)/\sigma_{(\overline{X}_T - \overline{X}_C)}$ where $\overline{X}_T$ and $\overline{X}_C$ are the mean outcomes for the treatment and control groups, the numerator therefore is the ATT, and $\sigma_{(\overline{X}_T - \overline{X}_C)}$ is the standard error of ATT. If we know $\overline{X}_T$, $\sigma_{(\overline{X}_T - \overline{X}_C)}$, and the t-score needed to reject the null hypothesis that ISO 14001 certification did not reduce the rate at which participants were subsequently fined (ATT ≥ 0), which we will denote t*, then we can calculate $\overline{X}_C^*$, the mean outcome for the control group needed to reject this null hypothesis. That is, $\overline{X}_C^* = \overline{X}_T + t^*\sigma_{(\overline{X}_T - \overline{X}_C)}$. The one-sided t-score needed to reject the null at 5 percent given 297 degrees of freedom is 1.650. As noted above, $\overline{X}_T$ is 0.023. $\sigma_{(\overline{X}_T - \overline{X}_C)}$ depends on the matching model used. As a result, each matching model is associated with a different $\overline{X}_C^*$. The average $\sigma_{(\overline{X}_T - \overline{X}_C)}$ for all five matching models is 0.006. Thus, to reject the null hypothesis, $\overline{X}_C$ would need to be 0.033, which is 42 percent [(0.033-0.023)/0.023)] higher than $\overline{X}_T$. Since $\overline{X}_C$ serves as our counterfactual (i.e., what the mean fine rate for participants would have been had they not participated), an equivalent interpretation is that ISO 14001 certification would need to reduce the mean fine rate by at least 42 percent. Hence, our results allow us to reject any hypothesis that ISO 14001 certification reduced participants' incidence of fines by more than 42 percent.

14001 certification; and (ii) our matching routines pair a significant number of targeted certified plants with nontargeted uncertified plants. The first condition holds: PROFEPA has an explicit, written policy of targeting large facilities in particularly dirty sectors, including manufacturing and mining (DOF 1990, 1992; Quezada 2005). Our duration analysis shows that these same types of plants tend to get ISO 14001 certified. Specifically, we find that plants that were large and in the manufacturing sector (Sector 6) were particularly likely to get certified. To test whether differential PROFEPA monitoring across plants drives our results, we reestimated our duration and impact models with a subsample restricted to the large manufacturing plants that PROFEPA targets: plants in Sector 6 with more than 12 million pesos in sales. The 2,278 plants meeting these criteria accounted for 184 of the 298 ISO certifications in the full sample of 80,611 plants. The results of the duration and matching models using this restricted subsample are qualitatively identical to those from the full sample (Table 3, Model B, and Table 4, Model III). This exercise also provides reassurance that our duration results are robust to multicollinearity among our fines, size, and sector regressors resulting from PROFEPA's targeting policy. By construction, this subsample eliminates such multicollinearity.

To digress briefly, PROFEPA's targeting policy also explains an apparent inconsistency in our arguments. Section 1 states that environmental regulatory pressure in developing countries is usually weak, a generalization that holds for Mexico during our study period (Brizzi and Ahmed 2001; Gilbreath 2003; OECD 2003). But in Section 5.3 we find that regulatory pressure drove ISO 14001 certification in Mexico. The reason both statements are valid is that although PROFEPA monitoring and enforcement may have been weak for the average facility in Mexico, for the types of plants getting ISO 14001 certified, it was significantly stronger. Our fines data reflect this differential monitoring and enforcement. The incidence of fines among all plants during our study period was 2 percent, but the incidence among large manufacturing plants (more than 12 million pesos in capital) in the manufacturing sector was 23 percent. A complementary explanation is that ISO 14001 certification relies on carrots as well as regulatory sticks to entice participants, including economic incentives created by buyers. As a result, even relatively weak regulatory pressures might trigger a decision to obtain certification.

Finally, our ATT estimates could be biased downward if regulatory monitoring that leads to fines was more stringent for certified plants than for uncertified plants during the outcome period, even *within* the categories of plants that PROFEPA targets. If this were the case, then improvements in regulatory compliance due to certification might not be reflected in a lower incidence of PROFEPA fines, simply because certified plants were inspected more frequently or closely and were therefore fined more than similar uncertified plants. Unfortunately, the hard data on PROFEPA inspections needed to test and control for the stringency of monitoring are not available. Prima facie arguments cut both ways. On one hand, it could be that ISO 14001 certification somehow improves PROFEPA's ability to monitor a plant's compliance, perhaps because PROFEPA has access to results of third-party audits or internal recordkeeping on ISO 14001 planning and implementation. On the other hand, however, PROFEPA may have rewarded plants that obtained ISO 14001 certification with less stringent monitoring just as participants in voluntary programs of the U.S. Environmental Protection Agency are sometimes rewarded with relaxed oversight (Toffel and Short 2008; Stafford 2007).

7. Conclusion

We have used data on some 80,000 industrial facilities and other businesses to analyze ISO 14001 certification in Mexico. The first stage of our analysis focused on identifying the drivers of certification, and the second stage focused on determining whether it improves regulatory compliance. In the first-stage analysis, we used a duration model because it explicitly accounts for the timing of the dependent variable (certification) and the main independent variable of interest (regulatory activity) and because it controls for right censoring. Our results suggest that regulatory fines do motivate certification: a fine roughly doubles the probability of certification for three years after it is assessed. Hence, in Mexico, the ISO 14001 program does more than certify already clean plants free-riding on unrelated investments in pollution control; it has attracted dirty plants under pressure from regulators.

In the second-stage analysis of program impacts, we used post-certification annual average incidence of fines to measure regulatory compliance and we used propensity score matching to control for bias created by self-selection into the program of plants with characteristics likely to affect compliance. All five of our matching estimators indicate that after obtaining ISO 14001 certification, plants were fined no more or less often than matched uncertified plants. These results are robust to variations in the criteria for matching, including exact matching on plant's history of fines, and to variations in the composition of our regression sample. Hence, our second-stage results suggest that the ISO 14001 certification did not have a large lasting effect on the average rate of compliance. It is important to reiterate that variability in our outcome data precludes observing modest effects on the average rate of compliance. Nevertheless, we believe our results are credible and shed important light on the environmental benefits of an increasingly popular but little-studied regulatory policy in developing countries.

Are our first- and second-stage results contradictory? If ISO 14001 participants included dirty plants under pressure from regulators, and if certified plants implemented all the measures required to obtain certification, then why did certified plants on average not have better compliance records than uncertified plants? The most likely explanation is simply that the effect of ISO 14001 certification on plants' compliance was limited or temporary or both. This explanation is consistent with findings from studies of the environmental consequences of ISO 14001 certification in industrialized countries (Barla 2007; King et al. 2005) and more broadly with findings of studies of VEPs in industrialized countries (Koehler 2008; Khanna 2001; Lyon and Maxwell 2008; Pizer et al. 2008). Our findings raise questions about the ability of eco-certification to help boost regulatory compliance in developing countries.

References

Arimura. T., A. Hibiki, and H. Katayama. 2008. Is a Voluntary Approach an Effective Environmental Policy Instrument? A Case for Environmental Management Systems. *Journal of Environmental Economics and Management* 55: 281–95.

Bansal, P., and T. Hunter. 2003. Strategic Explanations for the Early Adoption of ISO 14001. *Journal of Business Ethics* 46(3): 289–99.

Barla, P. 2007. ISO 14001 Certification and Environmental Performance in Quebec's Pulp and Paper Industry. *Journal of Environmental Economics and Management* 55(3): 291–306.

Blackman, A. 2008. Can Voluntary Environmental Regulation Work in Developing Countries? Lessons from Case Studies. *Policy Studies Journal* 36(1): 119–41.

———. 2010. Alternative Pollution Control Policies in Developing Countries. *Review of Environmental Economics and Polic*y 4(2): 234–53.

Blackman, A., and S. Guerrero. 2010. What Drives Voluntary Eco-Certification in Mexico? Discussion Paper 10-26. Washington, DC: Resources for the Future.

Blackman, A. and J. Rivera. In Press. Producer-level Benefits of Sustainability Certification. *Conservation Biology*.

Blackman, A., B. Lahiri, B. Pizer, M. Rivera Planter, and C. Muñoz Piña. 2010. Voluntary Regulation in Developing Countries: Mexico's Clean Industry Program. *Journal of Environmental Economics and Management* 60(3): 182–92.

Brizzi, A., and K. Ahmed. 2001. Sustainable Future. In M. Guigale, O. Lafourcade, and V.H. Nguyen (eds.), *Mexico: A Comprehensive Development Agenda for the New Era.* Washington, DC: World Bank, Chapter 4.

Brodaty, T., B. Crepon, and D. Fourgere. 2001. Using Matching Estimators to Evaluate Alternative Youth Employment Programs: Evidence from France 1986–1988. In M. Lechner and F. Pfeiffer (eds.), *Econometric Evaluation of Labour Market Policies*. Physica-Verlag, 85–123.

Caliendo, M., and S. Kopeinig. 2008. Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys* 32: 31–72.

Christmann, P., and G. Taylor. 2001. Globalization and the Environment: Determinants of Firm Self-Regulation in China. *Journal of International Business Studies* 32(3): 439–58.

Contacto de Unión Empresarial. 2005. Edición Especial. Mexico, D.F.

Cox, D. 1975. Partial Likelihood. *Biometrika* 62: 269–76.

Darnall, N. 2003. Why Firms Certify to ISO 14001: An Institutional and Resource Based View. Academy of Management Proceedings One: A1-A6.

Darnall, N., and D. Edwards. 2006. Predicting the Cost of Environmental Management System Adoption: The Role of Capabilities, Resources, and Ownership Structure. *Strategic Management Journal* 27: 301-320.

Darnall, N., and Y. Kim. In press. Which Types of Environmental Management Systems Are Related to Greater Environmental Improvements? *Public Administration Review*.

Dehejia, R.H., and S. Wahba. 2002. Propensity Score-Matching Methods for Nonexperimental Causal Studies. *The Review of Economics and Statistics* 84(1): 151.

Diario Oficial de la Federación (DOF). 1990. Listado de Actividades Altamente Riesgosas por los Efectos que Puedan Generar en el Equilibrio Ecológico y en el Medio Ambiente. March 28.

———. 1992. Segundo Listado de Actividades Altamente Riesgosas por los Efectos que Puedan Generar en el Equilibrio Ecológico y en el Medio Ambiente. May 4.

Dingwerth, K. 2008. Private Transnational Governance and the Developing World: A Comparative Perspective. *International Studies Quarterly* 52: 607–34.

Ecolabel Index (2011). www.ecolabelindex.org.

Eskeland, G., and E. Jimenez. 1992. Policy Instruments for Pollution Control in Developing Countries. *World Bank Research Observer* 7(2): 145–69.

Gilbreath, J. 2003. *Environment and Development in Mexico: Recommendations for Reconciliation.* Washington, DC: Center for Strategic and International Studies.

Heckman, J., H. Ichimura, J. Smith, and P. Todd. 1998. Characterizing Selection Bias Using Experimental Data. *Econometrica* 66: 1017–98.

International Organization for Standardization (ISO). 2009. The ISO Survey--2007. Available at http://www.iso.org/iso/survey2007.pdf.

Jiménez, O. 2007. Voluntary Agreements in Environmental Policy: An Empirical Evaluation for the Chilean Case. *Journal of Cleaner Production* 15: 620–37.

Khanna, M. 2001. Economic Analysis of Non-Mandatory Approaches to Environmental Protection. *Journal of Economic Surveys* 15(3): 291–324.

Kiefer, N. 1988. Economic Duration Data and Hazard Functions. *Journal of Economic Literature* 26: 646–79.

King, A., M. Lenox, and A. Terlaak. 2005. The Strategic Use of Decentralized Institutions: Exploring Certification with the ISO 14001 Management Standard. *Academy of Management Journal* 48(6): 1091–106.

Koehler, D. 2008. The Effectiveness of Voluntary Environmental Programs: A Policy at a Crossroads? *Policy Studies Journal* 35(4): 689–22.

List, J.A., D.L. Millimet, P.G. Fredriksson, and W.W. McHone. 2003. Effects of Environmental Regulations on Manufacturing Plant Births: Evidence from a Propensity Score Matching Estimator. *Review of Economics and Statistics* 85(4): 944–52.

Lyon, T., and J. Maxwell. 2002. Voluntary Approaches to Environmental Regulation: A Survey. In M. Frazini and A. Nicita (eds.), *Economic Institutions and Environmental Policy*. Aldershot and Hampshire: Ashgate Publishing.

———. 2008. Environmental Public Voluntary Programs Reconsidered. *Policy Studies Journal* 35(4): 723–50.

Mexico's Maquila Online Directory. 2007. Available at http://www.maquiladirectory.com/.

Montiel, I., and B. Husted. 2009. The Adoption of Voluntary Environmental Management Programs in Mexico: First Movers as Institutional Entrepreneurs. *Journal of Business Ethics* 88: 349–63.

Morgan, S., and D. Harding. 2006. Matching Estimators of Causal Effects: Prospects and Pitfalls in Theory and Practice. *Sociological Methods and Research* 35(1): 3–60.

Nel, J., and J. Wessels. 2010. How to Use Voluntary, Self-Regulatory and Alternative Environmental Compliance Tools: Some Lessons Learnt. *Potchefstroom Electronic Law Journal* 13(5): 48–78.

Nakamura, M., T. Takahashi, and I. Vertinsky. 2001. Why Japanese Firms Chose to Certify: A Study of Managerial Responses to Environmental Issues. *Journal of Environmental Economics and Management* 42(1): 23–52.

Nishitani, K. 2009. An Empirical Study of the Initial Adoption of ISO 14001 in Japanese Manufacturing. *Ecological Economics* 68: 669–79.

Organization for Economic Co-operation and Development (OECD). 2003. *OECD Environmental Performance Reviews: Mexico.* Paris: OECD Environment Directorate.

Pizer, W., R. Morgenstern, and J.-S. Shih. 2008. Evaluating Voluntary Climate Programs in the United States. Discussion Paper 08-13. Washington, DC: Resources for the Future.

Potoski, M., and A. Prakash. 2005a. Covenants with Weak Swords: ISO 14001 and Facilities' Environmental Performance. *Journal of Policy Analysis and Management* 24(4): 745–69.

———. 2005b. Green Clubs and Voluntary Governance: ISO 14001 and Firms' Regulatory Compliance. *American Journal of Political Science* 49(2): 235–48.

Quezada, J.E. 2005. Underattorney of Industrial Inspections, Procuraduría Federal de Protección al Ambiente. Telephone interview, Mexico City, October 4.

Rosenbaum, P., and D. Rubin. 1983. The Central Role of the Propensity Score in Observational Studies for Causal Effects. *Biometrika* 70: 41–55.

Russell, C., and W. Vaughan. 2003. The Choice of Pollution Control Policy Instruments in Developing Countries: Arguments, Evidence and Suggestions. In *International Yearbook of Environmental and Resource Economics*, vol. VII. Cheltenham, UK: Edward Elgar.

Sianesi, B. 2004. An Evaluation of the Active Labour Market Programmes in Sweden. *Review of Economics and Statistics* 86(1): 133–55.

Stafford, S. 2007. Should You Turn Yourself In? The Consequences of Environmental Self-Policing. *Journal of Policy Analysis and Management* 26(2): 305–26.

Toffel, M., and J. Short. 2008. Coming Clean and Cleaning Up: Is Voluntary Disclosure a Signal of Effective Self-Policing? Discussion Paper 08-98. Cambridge, MA: Harvard Business School.

World Bank. 2000. *Greening Industry: New Roles for Communities, Markets, and Governments.* New York: Oxford University Press.

Table A1. Average Treatment Effect on Treated (ATT) Estimates and Median Standardized Bias (MSB) by Propensity Score Matching Method and Model[a]

Method	*Model IV Exact matching on exports*		*Model V Exact matching on maquiladora*		*Model VI Exact matching on sales*		*Model VII Exact matching on sector*	
	ATT	***MSB***[b]	***ATT***	***MSB***[b]	***ATT***	***MSB***[b]	***ATT***	***MSB***[b]
Nearest neighbor 1-1	0.006 (0.014)	6.927	0.004 (0.008)	5.300	0.000 (0.013)	6.845	0.009 (0.009)	8.343
Nearest neighbor 1-4	-0.006 (0.011)	6.900	-0.005 (0.006)	4.814	-0.005 (0.008)	6.615	-0.004 (0.010)	6.969
Nearest neighbor 1-8	-0.000 (0.007)	6.121	-0.008 (0.006)	4.766	0.002 (0.007)	4.904	0.001 (0.007)	5.471
Nearest neighbor 1-16	-0.004 (0.006)	4.476	-0.006 (0.005)	4.361	-0.000 (0.005)	4.722	-0.001 (0.006)	4.025
Kernel	-0.003 (0.004)	3.881	-0.003 (0.004)	2.408	-0.002 (0.004)	3.740	-0.003 (0.004)	2.799

Model IV: same as I but with exact matching on 0/1 dummy = 1 if plant exports.
Model V: same as I but with exact matching on 0/1 dummy = 1 if plant is a maquiladora.
Model VI: same as I but with exact matching on 0/1 dummy = 1 if plant has over 30 million pesos in sales.
Model VII: same as I but with exact matching on 0/1 dummy = 1 if plant in manufacturing sector.
[a]For all models, the average annual incidence of fines among treatment group (ISO-certified plants) = 0.023 fines/year.
[c]For a given covariate, the standardized bias is the difference of means in the ISO 14001 certified and uncertified subsamples as a percentage of the square root of the average sample variance in both groups. We report the median standardized bias (MSB) for all covariates. For all models, MSB before matching is 24.365%.
***Significant at 1% level.