Third Party Certification and the Effectiveness of Voluntary Pollution Abatement Programs: Evidence from Responsible Care*

Martina Vidovic
Department of Economics, Rollins College
Email: mvidovic@rollins.edu

Michael S. Delgado
Department of Agricultural Economics, Purdue University
Email: delgado2@purdue.edu

Neha Khanna (corresponding author)
Department of Economics and Environmental Studies Program, Binghamton University
Email: nkhanna@binghamton.edu

November 6, 2015

Abstract

We ask whether the introduction of mandatory third party certification under the Responsible Care (RC) program from 2005 onwards yielded lower emissions from RC plants compared to non-RC plants in the United States chemical industry. We use facility level panel data from 935 plants between 1996 and 2010, and estimate the causal impact of third party certification on facility emissions. We address endogenous selection into RC via instrumental variables, and explore the incidence of essential heterogeneity in the treatment effect via the marginal treatment effect. Results indicate that firms selected into the RC program at least partly in response to unobservable gain, and we do not find evidence that third party certification led to a reduction in facility emissions compared to non-treated facilities outside the RC program.

Keywords: Certification, essential heterogeneity, instrumental variables, marginal treatment effect, self-regulation

JEL codes: Q53, Q58, L60

* We thank Andrew King for providing information on membership in the Chemical Manufacturers' Association from 1988-2001. We are indebted to participants at the 2014 ASSA conference, 2013 AERE annual summer conference, 2013 AAEA annual conference, 2013 SEA annual conference, seminar participants at Cornell University, as well as an anonymous Associate Editor at this *Journal* for helpful comments.

1. Introduction

Industry self-regulation via voluntary pollution abatement has become popular not only with industry groups but also with environmental policymakers because it gives them a relatively easy to use lever that does not require an act of Congress. There is a substantial academic debate on the effectiveness of such programs, with some authors arguing that these programs are quite effective in reducing pollution (e.g., Khanna and Damon 1999, Bi and Khanna 2012) while others argue, and with equal conviction, that these programs are ineffective at best (Gamper-Rabindran 2006, Vidovic and Khanna 2007, 2012, Carrión-Flores et al. 2013) and counter-productive at worst (King and Lennox 2000, Gamper-Rabindran and Finger 2013). The Achilles' heel of this debate as it relates to the United States (US) is that it relies on relatively old data from the 1990s and on programs that are either no longer in existence (for example, the 33/50 Program which ended in 1995) or on early versions of programs that were changed substantively in later years. For example, the American Chemistry Council's (ACC) Responsible Care (RC) program has been analyzed only through 2001 (Gamper-Rabindran and Finger 2013), after which major structural changes were incorporated.

We contribute to the literature on the effectiveness of voluntary pollution abatement in the US by using the recent structural changes in the RC program to assess whether independent third party certification, which was made mandatory from 2005 onwards, leads to a reduction in facility emissions relative to non-participating, non-certified facilities in the US chemical industry. Indeed, the impetus for introducing mandatory third party certification into the RC program was the growing concern that traditional voluntary abatement programs do not lead to improvements in environmental outcomes. Our research is the first to investigate the structural redesign of third party certification into RC.

To understand the impact of third party certification on facility emissions, we estimate average treatment effects using panel data techniques. We believe that firms likely selected into the RC program and so the sample of facilities that were subject to third party certification is not random (see also Gamper-Rabindran and Finger 2013). To address the possibility of endogenous treatment, we deploy instrumental variables to recover a consistent estimate of the treatment effect.

Recent work has drawn an important distinction between the classic notion of endogenous selection (selection on the level) and endogenous selection on participation gain (Heckman and Vytlacil 1999, 2001, 2005 and Heckman et al. 2006). If firms for which RC standards are relatively easy to adhere to are more likely to join the program, the model is one of classic selection. Alternatively, if through participation in RC a firm expects a substantial improvement in reputation with consumers, investors, or regulators, the model is one of selection on participation gain. While the voluntary abatement literature has considered both sources of endogeneity previously, no attempt has been made to empirically differentiate between the two.

In the case of classic sample selection, our strategy is a straightforward application of an instrumental variables regression, which yields a consistent estimate of the average treatment effect (ATE), which is homogeneous (described in Section 4). When there is selection on participation gain, there is heterogeneity even in the average (over observables) effect of treatment, such that the estimated ATE from the first instrumental variables estimator cannot fully describe the effect of treatment. This kind of heterogeneity is referred to as *essential heterogeneity* by Heckman et al. (2006), because traditional parameters are unable to measure the treatment effect. To explore whether the third party certification effect is heterogeneous, we focus on identification and estimation of the *marginal treatment effect* (MTE), following the seminal work of Björkland and Moffitt (1987) and more recently by Heckman et al. (2001), Heckman and Vytlacil (1999, 2001, 2005), and Heckman et al. (2006). In our case, the MTE is the average treatment effect for facilities that belong to a firm that chose to select into RC along some marginal value of participation (described in detail below).[1] By focusing on the MTE, we can recover the pointwise average treatment effect over the range of support of the propensity of self-selection for facilities that selected into treatment, through which we can understand heterogeneity in the treatment effect. We use the tools developed by Heckman et al. (2010), and recently applied by Carneiro et al. (2011) to empirically test whether selection is on levels or gains.

Our empirical approach is in contrast to previous empirical papers in the voluntary abatement literature that have ignored differences between classic selection and selection on unobserved gain. Hence, our focus on the MTE constitutes a novel and significant methodological improvement in the literature. More importantly, our consideration of the three models – i.e., a baseline panel difference-in-differences model without instruments, a panel model with instruments but an assumed homogeneous treatment effect, and the instrumental variables model that identifies the MTE – allows us to shed light on the nature of firm selection into the RC program and to assess the extent to which the third party certification requirement is capable of inducing participating facilities to reduce emissions relative to other, non-certified facilities under a variety of possible self-selection mechanisms.

We have an unbalanced panel of 12,999 observations from 935 facilities belonging to 325 firms in the US chemical manufacturing industry between 1996 and 2010. We have constructed our dataset to mitigate potential problems associated with missing data on facility membership in RC for three years as well as potential biases in the effect of treatment on emissions that could result from facilities not reporting emissions, being traded between firms, or entering and exiting the program just before or after the treatment takes place.

[1] The decision to participate in RC was made at the firm level, and our emissions data are measured at a facility level. More detailed discussion of the structure of RC and our data follows in subsequent sections.

Overall we do not find evidence that third party certification leads to a decrease in facility emissions. Nonetheless, we find that selection into RC is in part based on unobserved gains. Our estimate of the MTE indicates that firms most likely to participate in RC do not significantly change emissions under mandatory third party certification relative to other non-certified facilities in the chemical industry. This result is also found by our standard instrumental variables estimator. The MTE analysis indicates that some facilities whose parent firms are more reluctant to participate may increase their emissions level if subject to mandatory third party certification.[2] Our results provide gloomy indication that the introduction of third party certification into the RC program in 2005 is not sufficient for improving the reputation of the program, nor do our results generally suggest that third party certification is an effective means of inducing effective self-regulation by facilities.

2. Self-Regulation and Third Party Certification

Over the past three decades, voluntary approaches to environmental management have become equally popular among policymakers, industry groups and non-governmental organizations. The US EPA's Partnership Programs website alone lists over 40 programs with more than 13,000 participants (http://www.epa.gov/partners/programs/index.htm). The growing reliance on self-regulatory approaches begs the question whether voluntary programs are able to elicit meaningful changes in environmental performance and whether the signals they send accurately reflect the behavior of their participants. Prior research evaluating the effectiveness of voluntary pollution abatement programs found that participation in such programs was either not associated with promoting superior environmental performance among their participants (Rivera et al. 2006,

[2] This result echoes the findings of Gamper-Rabindran and Finger (2013), that RC participants had higher emissions than non-participants in the years 1988-2001.

Gamper-Rabindran 2006, Vidovic and Khanna 2007, 2012) or led to worse environmental outcomes (King and Lennox 2000, Gamper-Rabindran and Finger 2013). On the other hand, Khanna and Damon (1999), Innes and Sam (2008), Sam et al. (2009), Bui and Kapon (2012) and Bi and Khanna (2012) argue that such programs are effective in reducing pollution. Some authors have begun to caution that program design characteristics and lack of performance requirements may be responsible for the failure of voluntary approaches to make a difference (Darnall and Carmin 2005, Potoski and Prakash 2005, Rivera et al. 2006). Weak performance standards and the absence of effective enforcement may provide an incentive for firms to free ride and to serve their own interests at the expense of other participants.

Evidence regarding the ineffectiveness of US voluntary programs in achieving environmental protection is primarily based on the 33/50 Program (Gamper-Rabindran 2006, Vidovic and Khanna 2007, 2012), the Sustainable Slopes program (Rivera et al. 2006), and the early years of the RC program (King and Lennox 2000, Gamper-Rabindran and Finger 2013); all are programs that relied on self-monitoring and assurance from participants that they adhered to the program requirements. It is not clear from the results whether participants failed to adopt superior environmental protection practices or whether the programs failed to elicit improvement from the participants. At least in the case of RC, King and Lenox (2000) argue that voluntary programs designed by industry associations lack appropriate implementation, monitoring, and reporting procedures that would initiate superior environmental performance by participants.

Among voluntary programs that award a label or recognition if certain standards are met, third-party oversight has emerged as a way of providing credibility to the certification system. For example, to ensure integrity and sustainability, the EPA integrated third party certification in its Water Sense and Energy Star programs. The forest product label from the Forest Stewardship

Council and the sustainable seafood label from the Marine Stewardship Council use third party certification to recognize sustainable management of forests and fisheries. Similarly, the International Organization for Standardization instituted third party audits of the ISO 9000 Quality Management System Standard and ISO 14001 Environmental Management System Standard. Recently, the ACC incorporated third party certification in its signature RC program.

Studies that assess whether third party certification improves environmental performance via voluntary approaches mainly focus on one program, ISO 14001. Several early works found that ISO 14001 certified firms reduced waste and resource use significantly more than non-registrants (Rao and Hamner 1999, Montabon et al. 2000, Melnyk et al. 2002). Unfortunately these studies suffer from potential methodological and sample issues and their results should be interpreted with caution.[3] King et al. (2005) found a weak negative effect of ISO 14001 on emissions; certification provides stakeholders with information about the ongoing efforts to improve the performance of an environmental management system but it is not correlated with reductions in emissions. Russo (2009) found that being an early adopter is associated with lower emissions and that emissions fall the longer a facility operates under ISO 14001 certification. The two most systematic studies that compare the environmental performance of adopters and non-adopters over time are Potoski and Prakash (2005) and Toffel (2006). Both studies, using different methodologies, find that ISO 14001 certified facilities reduced their emissions more than non-certified facilities. These authors therefore suggest that programs with enforcement mechanisms based on third-party audits could potentially improve compliance with underlying program

[3] For example, Rao and Hamner (1999) use information collected from a questionnaire administered to ISO 14001 registrants and there is no information on non-registrants in their dataset. Montabon et al. (2000) and Melnyk et al. (2002) construct independent and dependent variables from answers to a survey where the respondents were likely the same people who made decisions regarding their firm's participation in ISO 14001 and provided opinions on its impact on the firm's performance.

commitments even in the absence of public disclosure of the audit information.

We add to the existing literature on the effectiveness of voluntary management programs by examining whether the introduction of independent third party certification from 2005 onwards yielded lower emissions from RC facilities compared to statistically equivalent non-RC facilities in the US chemical industry. The advantage of studying the RC program in this context is that the mandatory certification under RC was modeled on the certification under ISO 14001. Our analysis sheds light on whether third party oversight of voluntary abatement programs makes them a more effective instrument in the US policymaker's environmental toolbox.

3. The Potential for Responsible Care to Improve Environmental Outcomes

In 1988, the ACC (then known as the Chemical Manufacturers' Association) adopted the RC initiative to promote continuous Environmental, Health, Safety and Security (EHS&S) performance improvement for all of its members. The industry association implemented the program in order to improve public perception about the safety of the chemical industry and in anticipation of more stringent regulatory interventions following the chemical disaster at the Union Carbide plant in Bhopal, India, and the subsequent leak from the Union Carbide's pesticide plant in Institute, West Virginia, in the mid-1980s. Participation in Responsible Care was made a condition for membership in ACC.

Throughout the 1980s and 1990s, the program was structured around a set of codes of EHS&S management practices. In 1996 a voluntary peer-review process called Management System Verification was added to the program. The process served to verify that appropriate systems were implemented to assure ongoing compliance with the firm's EHS&S performance goals and external regulations. The system was not an audit of a firm and did not identify non-compliance with regulations or the level of emissions at a facility.

In 2002 the ACC announced substantial changes to RC, recognizing that US regulation of the chemical industry had caught up with RC requirements. In that year 75 percent of the original RC activities were covered by laws and regulations compared to only 13 percent in 1988 (Phillips 2006). Stakeholders lost support for the program and firms began to differentiate themselves from it. As part of its change, the program implemented the Responsible Care Management System (RCMS), a management system approach built on the "Plan-Do-Check-Act" philosophy to improve firm performance in key areas of community awareness and emergency response, security, distribution, employee health and safety, pollution prevention, and process and product safety (ACC 2013). To enhance transparency, it adopted mandatory third-party certification of those management systems. Under independent oversight, every RC firm must certify that it has a management system in place and demonstrate progress toward improved performance. To obtain certification, firms undergo headquarter and facility audits conducted by independent, accredited auditing firms (ACC 2013). Third party certification was officially launched in 2005 and all ACC members were required to complete third party audits by the end of 2007.

The ACC requires that certification is renewed every three years, and firms can choose to demonstrate conformance either to the RCMS or the RC 14001 technical specification which combines RC and ISO 14001 certification. Recognition and popularity of ISO 14001 with stakeholders worldwide prompted firms to seek an approach that would avoid duplicating the RC and ISO 14001 audit processes. The RC 14001 technical specification integrates elements of both RC third party certification and ISO 14001 allowing a single certification process to fulfill both program requirements (Phillips 2006). In order to obtain the RC 14001 certification, the organization must conform to the ISO 14001 with respect to environment as well as to health, safety, and security requirements within the scope of RC.[4]

Unlike performance standards that set the level of environmental protection or state requirements for improved environmental performance, certified management standards such as RC 14001 only require firms to establish processes and management systems to ensure that environmental goals are developed, assessed, and met. However, certification may still provide information to stakeholders by conveying that an environmental management system exists and whether it leads to improvement. Voluntary programs without third party oversight are often criticized for the potential for some participants to shirk. If certification is costly and stakeholders are willing to pay for superior performance, certification may provide a credible signal. According to the ACC, the third-party auditing system is part of the drive to increase credibility and public confidence in RC. Although the ACC always mandated that all firms must adopt RC as a condition for membership, critics questioned the credibility of the program because ACC membership is voluntary and the ACC has never expelled a member for non-compliance (Prakash 2000). The certification system is likely to formalize managerial commitment to achieving environmental performance goals (Rondinelli and Vastag 2000), provide accountability, and reduce opportunities for opportunistic behavior (King and Lenox 2000). In addition, the ACC requires public disclosure of environmental information by all members on its website and with governmental agencies. Therefore, we anticipate that following the implementation of third party certification, RC participants improved their environmental performance compared to non-participants in the US chemical industry.

4. Empirical Methodology

[4] Unfortunately, the data made available by the ACC on its website do not allow us to distinguish between firms that are certified under RCMS vs. RC14001.

Overview

The hypothesis we test is whether third party certification causes a facility to lower its emissions of TRI air releases relative to other facilities that are not certified under RC. Under some common assumptions (described below) the appropriate parameter to assess this hypothesis is the average treatment effect (ATE), or the expected effect of third party certification on facility emissions for any randomly drawn facility subjected to third party certification. We suspect, however, that these assumptions might not hold – given the structure of the mandatory third party certification requirement under RC and the nature of our data, firms that self-selected into the RC program also self-selected into third party certification so that treatment by third party certification is non-random (see also Gamper-Rabindran and Finger 2013). Depending on the nature of selection into treatment, the resulting heterogeneity in the effect of third party certification on facility emissions (e.g., Heckman and Vytlacil 2005, Heckman et al. 2006) may render the ATE insufficient for understanding the program impact.

Heckman and his co-authors have developed an empirical framework that unifies notions of conditional exogeneity and different forms of selection (e.g., Heckman and Vytlacil 1999, 2001, 2005 and Heckman et al. 2006), and allows researchers to empirically test whether selection into treatment exists, and if so, the nature of the selection (Heckman et al. 2010, Carneiro et al. 2011). This framework provides a clear lens through which to understand why firms participated in the RC program, and for determining the causal impact of third party certification on facility emissions even in the presence of treatment parameter heterogeneity. We begin by deriving a general regression model of facility emissions that nests several empirical models that differ in their assumption regarding selection. We describe a testing procedure based on the general design that informs us of the existence and nature of selection (Heckman et al. 2010), from which we derive our preferred treatment effect estimates.

Framework of Potential Outcomes

Define Y_{it} to be the level of total air emissions reported to the Toxic Releases Inventory (TRI) for facility $i = 1,2,\dots,N$ in time $t = 1,2,\dots,T$. Let Y_{it}^1 and Y_{it}^0 denote two potential outcomes such that Y_{it}^1 indicates the level of air emissions if the facility is treated by third party certification and Y_{it}^0 indicates the level of air emissions if the facility is not subject to third party certification. Under a linear in parameters restriction, the models

$$Y_{it}^1 = X_{it}\beta^1 + c_i + \tau_t + U_{it}^1 \tag{1}$$

$$Y_{it}^0 = X_{it}\beta^0 + c_i + \tau_t + U_{it}^0$$

define each potential outcome as a function of time-varying covariates X_{it}, a vector of state-specific coefficients β^l for $l = \{0, 1\}$, a time-constant unobservable facility specific effect c_i, an unobservable year-specific factor τ_t, and time-varying unobservables U_{it}^l that correspond to each state.

The fundamental problem of causal inference (Holland 1986) is that Y_{it}^1 and Y_{it}^0 are never observed for the same facility in the same time – we either observe facility emissions if the facility is treated under third party certification, or we observe facility emissions if the facility is not treated. That is, we observe $Y_{it} = D_{it}Y_{it}^1 + (1 - D_{it})Y_{it}^0 = Y_{it}^0 + (Y_{it}^1 - Y_{it}^0)D_{it}$ where D_{it} is an indicator for treatment by third party certification, and $\Delta_{it} \equiv Y_{it}^1 - Y_{it}^0$ defines the effect of third party certification on facility emissions.

Model (1) implies the regression

$$Y_{it} = X_{it}\beta^0 + [X_{it}(\beta^1 - \beta^0) + (U_{it}^1 - U_{it}^0)]D_{it} + c_i + \tau_t + U_{it}^0 \tag{2}$$

in which the treatment effect is $\Delta_{it} \equiv Y_{it}^1 - Y_{it}^0 = X_{it}(\beta^1 - \beta^0) + (U_{it}^1 - U_{it}^0)$ and U_{it}^0 defines the usual regression error. In the general case, Δ_{it} varies over facilities and years even after controlling for observable covariates X_{it} through $U_{it}^1 - U_{it}^0$, which is the unobservable change in facility emissions from treatment by third party certification. This general design is termed *essential heterogeneity* by Heckman et al. (2006) because the treatment effect cannot be generally summarized by a single parameter (such as the ATE).[5] The expected treatment effect, known as the marginal treatment effect (MTE), is

$$MTE(x, u^D) = E[\Delta_{it} | X_{it} = x, U_{it}^D = u^D] = x(\beta^1 - \beta^0) + E[U_{it}^1 - U_{it}^0 | U_{it}^D = u^D] \tag{3}$$

where U_{it}^D defines a vector of facility-time specific unobservables (defined in detail below), which after averaging over X_{it} yields

$$MTE(u^D) = \bar{X}(\beta^1 - \beta^0) + E[U_{it}^1 - U_{it}^0 | U_{it}^D = u^D] \tag{4}$$

which remains a function of the unobservable. The MTE is the average effect of treatment for facilities that are indifferent to participation given a particular value of the unobservables, u^D. This is clearly different from the case where $U_{it}^1 = U_{it}^0$, in which the MTE is equivalent to the ATE:

$$ATE(x) = E[\Delta_{it} | X_{it} = x] = x(\beta^1 - \beta^0) \tag{5}$$

which, averaging over X_{it} yields the ATE parameter

$$ATE = \bar{X}(\beta^1 - \beta^0). \tag{6}$$

When $U_{it}^1 = U_{it}^0$, ATE is constant (conditional on X_{it}). In this latter model, if D_{it} is correlated with U_{it}^0, we get the classic sample selection model and standard instrumental variables methods can be applied to estimate the parameters. A stronger assumption that D_{it} is conditionally uncorrelated with U_{it}^0 allows direct estimation via non-instrumental variable panel methods.

We believe it is unlikely that firm participation in the RC program is random. This means that the most restrictive version of (2) – that is, $U_{it}^1 = U_{it}^0$ and D_{it} is uncorrelated with U_{it}^0

[5] This selection mechanism is also referred to as *selection on the gain*, because the economic agent chooses participation in part on the expected gains from participation (e.g., Heckman et al. 2006). The model is also one of correlated random coefficients (Heckman et al. 2010).

conditional on (X_{it}, c_i, τ_t) – is not consistent with our empirical problem. It is more likely that firms self-selected into the RC program (see Gamper-Rabindran and Finger 2013) for unobservable reasons that are not likely to be captured by c_i or τ_t. The question is, is this a problem with essential heterogeneity?

Indeed, since ACC membership is voluntary, participation in RC is also voluntary – a firm would participate only if the benefits outweigh the costs.[6] For example, a firm is likely to participate if adhering to the RC standards is relatively easy, given the firm's current technology or required set of inputs. A firm with older technology and/or a firm with production that is more dependent on relatively dirtier inputs with few options for substitutability would find it more difficult to maintain RC standards and may be less apt to participate. This case is well-known in the voluntary abatement literature (e.g., Khanna and Damon 1999 and Vidovic and Khanna 2007), and describes the classic sample-selection problem; this corresponds to the version of (2) such that $U_{it}^1 = U_{it}^0$ and $D_{it}$is correlated with U_{it}^0.

An alternative possibility, and one that need not be mutually exclusive from the sample-selection problem, is the general version of (2) that includes selection on the participation gain. The literature has also postulated a scenario in which a firm might join a voluntary abatement program in order to improve its environmental reputation and/or corporate social governance image with consumers, investors, and regulators (e.g., Arora and Gangopadhyay 1995; Arora and Cason 1996; Henriques and Sadorsky 1996; Segerson and Miceli 1998; Khanna, Quimio, and Bojilova 1998; Khanna and Damon 1999; Maxwell, Lyon and Hackett 2000; Vidovic and Khanna

[6] The voluntary abatement literature has produced a variety of reasons why participation in a voluntary abatement program may be beneficial despite the fact that abatement is costly. These include, but are not limited to, product differentiation to attract high paying consumers with preferences for environmentally friendly products, optimal (cost-minimizing) strategy in a dynamic game with a regulator, and a desire to influence mandatory standards such that the firm owns the best available technology.

2007; Innes and Sam 2008). A firm is likely aware of the potential reputational gain from participation, and participates if the net gain is positive. If RC provides the firm with an opportunity to improve its image, then the effect of treatment depends heterogeneously on these (unobservable) gains. In this case, $U_{it}^1 \neq U_{it}^0$, and the appropriate model is the general version in equation (2).

It is worth clarifying that, while a firm can choose whether or not to participate in the RC program, a participating firm cannot choose whether or not to be third party certified. Starting in 2005 third party certification was mandatory for any firm participating in the RC program. As we describe in Section 5, our dataset does not include facilities that switch RC participation status, so that RC participation remains constant for each facility throughout our sample period. This means that participation in the RC program is perfectly collinear with third party certification treatment (post 2005). Therefore, endogenous selection (of any kind) into the RC program translates directly into endogenous selection into third party certification.

Given the likely parent firm self-selection into treatment, we model the probability that a facility belongs to a firm that participated in RC via

$$D_{it}^* = Z_{it}\theta + v_i - V_{it} \tag{7}$$

where D_{it}^* is the unobserved gain from participation in RC, Z_{it} is a vector of facility and firm specific, time-varying covariates that influence the participation decision, v_i is a time-invariant facility specific effect, θ is a parameter vector, and V_{it} is an unobserved random variable. We observe D_{it} as a binary indicator for RC participation, and hence treatment by third party certification from 2005 onwards, via

$$D_{it} = 1[D_{it}^* \geq 0] = 1[Z_{it}\theta + v_i - V_{it} \geq 0] \tag{8}$$

where $1[\cdot]$ is the indicator function. As described by Heckman et al. (2006), the participation decision can be written as $P(Z_{it}, v_i) > U_{it}^D$ for $P(Z_{it}, v_i) = F_V(Z_{it}\theta + v_i)$ and $U_{it}^D = F_V(V_{it})$

where $F_V(\cdot)$ is the cumulative distribution function of V. The convenience of this transformation is that we can use the propensity score, $P(D_{it} = 1|Z_{it} = z, v_i)$, to assess the nature of selection and to estimate the MTE.

This decision rule framework is important for defining the margin over which an individual firm and all of its facilities are indifferent to participation, given its unobservable net benefit from participating in the RC program. For any $z \in Z$, and hereafter keeping the facility effect in $P(\cdot)$ implicit, define $D(z) = 1[P(z) > U_{it}^D]$ as an indication of RC participation for facility i in time t had Z_{it} been exogenously fixed at z, *ceteris paribus*. We can statistically exact such exogenous variation in Z_{it}, via instrumental variables that are plausibly excluded from X_{it}, and ascertain each facility's participation status over a range of z. As discussed in Heckman and Vytlacil (1999, 2001, 2005), Heckman et al. (2006) and Carneiro et al. (2011), the MTE can be evaluated at $U_{it}^D = p$ where p is a limit point in the support of the propensity score. Evaluating the MTE at a high value of p provides an estimate of the third party certification treatment effect for facilities that have unobservable gains that make them least likely to participate in the RC program, whereas evaluating the MTE at low values of p provides an estimate of the treatment effect for facilities that are most likely to self-select into the RC program. By equation (8), a facility with a high value of U_{it}^D requires a greater value of p, artificially adjusted via the instrumental variables, to be a participant.[7]

Parametric Estimation and Specification Testing

From equation (2), and following examples provided in Heckman et al. (2010) and Carneiro et al.

[7] An interesting implication is that the MTE is only identified over the overlapping range of support of $P(Z)$, which means that the larger the overlapping range of support of $P(Z)$ the longer the margin over which the MTE can be evaluated. This bears implications for the ability to identify the ATE, which requires full support over $(0,1)$ (Heckman and Vytlacil 1999).

(2011), we specify the model to be a cubic polynomial of the propensity score

$$Y_{it} = X_{it}\beta^0 + X_{it}(\beta^1 - \beta^0)P(Z_{it}) + \sum_{j=1}^{3} \eta_j P(Z_{it})^j + c_i + \tau_t + U_{it}^0 \tag{9}$$

which can be estimated using a least squares fixed effects regression. As discussed in Heckman and Vytlacil (1999, 2001, 2005), Heckman et al. (2006), Heckman et al. (2010) and Carneiro et al. (2011), evidence of nonlinearity of Y_{it} in the propensity score is indicative of selection on unobservable gain. Conversely, evidence of linearity of Y_{it} in the propensity score – formally, that $\eta_j = 0, \forall j > 1$ – is evidence that facilities do not select on unobservable gain because in that case the marginal impact of program participation is constant after controlling for X_{it}. Hence, from (9), a model specification test of joint significance of higher order polynomial coefficients can be used to assess the incidence of facility selection into RC on unobservable gain.

If Y_{it} is nonlinear in $P(Z_{it})$, estimation of the MTE follows the method of local instrumental variables (Heckman and Vytlacil 1999, 2001), given by

$$\Delta_{MTE}(x,p) = \frac{\partial E[Y_{it} | X_{it} = x, v_i, \tau_t, P(Z_{it}) = p]}{\partial p} \tag{10}$$

which given (9) yields

$$\Delta_{MTE}(p) = \bar{X}(\beta^1 - \beta^0) + \sum_{j=1}^{3} j\eta_j p^{j-1}. \tag{11}$$

To obtain an estimate of the propensity score, we first use a random effects probit regression. The next step is to estimate equation (9), from which we evaluate equation (11) at points in the support of the estimated propensity score. Prior to calculating equation (11), we analyze the distribution of the estimated propensity scores to ascertain the interval of relevant support over which the MTE is identified, which is the range of overlap in the propensity score between treated and untreated facilities. We use a paired bootstrap with 399 resamples to obtain standard errors of the MTE.

Estimation when Selection is Not on Unobservable Gain

In the event that firms select into the RC program solely in terms of observables and facility/year specific factors, or that selection depends on levels and not gains so that the average treatment effect is homogenous, equations (3) through (6) describe how the MTE becomes equivalent to the ATE. Equation (2) then becomes

$$Y_{it} = X_{it}\beta^0 + [X_{it}(\beta^1 - \beta^0)]D_{it} + c_i + \tau_t + U_{it}^0 \tag{12}$$

which coincides with a panel specification

$$Y_{it} = X_{it}\beta + \delta D_{it} + c_i + \tau_t + \epsilon_{it} \tag{13}$$

such that $\delta \equiv [X_{it}(\beta^1 - \beta^0)]$ is the homogeneous effect of treatment (conditional on X_{it}). If we assume that D_{it} is uncorrelated with ϵ_{it}, conditional on (X_{it}, c_i, τ_t), we can consistently estimate δ using panel fixed effects techniques. Moreover, under this restriction we can consider the interaction between the treatment indictor and a set of year indicators that allows us to recover a differential impact of third party certification on facility emissions over the years 2005-2010. Alternatively, we can deploy instrumental variables panel methods to consistently estimate δ in the event that there remains some correlation between D_{it} and ϵ_{it} after controlling for (X_{it}, c_i, τ_t).

Hence, the empirical strategy we employ is to first estimate the traditional panel specification both with and without instrumental variables to obtain an estimate of the ATE, assuming absence of essential heterogeneity. We then explore the possibility that the ATE depends on unobservable gain by estimating the MTE and testing for nonlinearity of the regression in terms of the propensity score as evidence of essential heterogeneity in the average effect of third party certification on facility emissions.

Empirical Model Specification

In our empirical specification of the model, the facility/time-varying covariates, X_{it}, that affect a facility's emissions are: the facility to parent firm TRI release ratio, parent firm TRI releases, the facility share of Hazardous Air Pollutants in total TRI air releases (HAP-TRI release ratio), number of facility inspections under the Clean Air Act (CAA), and the number of gases for which the county where a facility is located has been out of attainment with the National Ambient Air Quality Standards (NAAQS). Some of the differences in facility TRI emissions may be due to differences in relative facility and firm size and the first two covariates allow us to control for them. The last three covariates capture a facility's exposure to mandatory regulation under various aspects of the CAA. Like Vidovic and Khanna (2007, 2012), we anticipate that facilities with a greater HAP to TRI release ratio, which captures the exposure of facilities to regulation of HAPs, facilities located in counties that are out of attainment with the NAAQS and are under regulatory pressure, and facilities with a larger number of inspections under the CAA will face an additional incentive to reduce their TRI emissions in order to mitigate the cost and stringency of current and/or future mandatory regulation. The unobservable year specific factor accounts for changes in regulations and available technology over time, as well as any general trends in emissions, such as gradual reductions in emissions over time, that should not be erroneously attributed to third party certification (Vidovic and Khanna 2007). The facility-specific effect accounts for differences among facilities that are constant over time, such as the level of employment or in some cases management characteristics.

For the model to be identified, we require instrumental variables that appear only in the selection equation but are plausibly excluded from the outcome equation. In our case the selection equation includes parent firm level variables. The decision to participate in RC was made by the parent firm for all of its facilities jointly, while pollution performance is specific to each facility. Following Gamper-Rabindran and Finger (2013) our firm level instruments include the number of other facilities reporting under the parent firm and a dummy variable that is equal to 1 if the parent firm is publicly owned. These instruments reflect the fact that larger firms are more likely to participate in voluntary pollution abatement programs and that publicly traded firms are likely to be in closer contact with the final consumer giving them a greater incentive to participate (Vidovic and Khanna 2007). That is, we expect that facilities belonging to larger, publicly traded firms are more likely to be subject to the third party certification treatment because their parent firms are more likely to be RC participants.

5. Data Description and Sources

Our data consist of US chemical manufacturing facilities that report emissions of toxic chemicals to the TRI. We restrict our sample to facilities that report SIC 28 and/or NAICS 325 as their primary industry, representing the largest share of the facility's economic activity.[8] Andrew King provided us with a list of RC participants from 1988 to 2001. We obtained the list of current ACC participants and their certification status between 2005 and 2010 from the ACC website (http://reporting.responsiblecare-us.com/Reports/Members/RCMSC_Cmpny_Rpt.aspx, accessed May 14, 2012).

RC participation is reported at the firm level and, as is common in the literature, we assume that all facilities belonging to a participating parent firm participated in the program. We have information on the RC status for each firm in each year between 1988 and 2001. We also have information on whether firms were third party certified between 2005 and 2010, and we only count

[8] The TRI began using NAICS instead of SIC codes starting with the 2006 reporting year; submissions from previous years were assigned NAICS codes based on their 2006 information and on their SIC codes.

firms and their plants as RC participants if they obtained certification at the headquarters and at a sample of facilities for the periods 2005-2007 and 2008-2010. However, we do not have data on RC participation for the intervening years, i.e. 2002, 2003, 2004, and we assume that firms that were members in both 2001 and in 2005 remained members through the three years for which we have missing membership information.[9]

Since RC and non-RC facilities may differ systematically, for identification purposes we classify facilities strictly as either RC members or as non-RC members during our period of analysis, 1996-2010. That is, to avoid contamination of our treatment and control groups we only consider facilities that do not switch between these two groups. For example, if a facility belonged to an RC member firm between 1996 and 1999 and then it was traded to a non-RC firm in 2000, we exclude this facility from our dataset because the effects of having been an RC member may linger after the facility is traded to a non-RC firm thus contaminating our control group. The same is true for a facility that was originally not a member of RC. However, if a participating facility was traded in any year to another parent firm that was also a member of RC, we treat this facility as an RC participant and it remains in our data set. Our sampling strategy also avoids another potential problem with facility level data. Because facilities are traded across firms fairly frequently we cannot tell with certainty whether facilities traded between RC and non-RC firms at around the time that third party certification was introduced were traded in response to this structural change in the program or not, and we avoid making a somewhat arbitrary judgment call.

We obtain data on total TRI air releases, HAP air releases, names of parent firms, and facility names and locations from the TRI (www.rtknet.org/new/tri). We include facilities that report data to the TRI continuously from no later than 2003 through 2010. Information on the

[9] Historical data from 1988-2001 show that firms maintain continuous membership until they opt-out of RC.

number of inspections under the CAA is from the Integrated Data for Enforcement Analysis database (www.epa-echo.gov/echo/index.html); county non-attainment status with the CAA is from the EPA's Green Book (www.epa.gov/oar/oaqps/greenbk).

We define facility emissions of HAP and TRI chemicals as the annual air releases of the 1995 core chemicals, which have been reported to the TRI throughout our period of analysis. Firm emissions are the sum of emissions for all facilities reporting to each parent firm in each year.

County non-attainment status is the count of pollutants for which a county has been designated by the EPA to be out of attainment with the NAAQS. The EPA will designate a county to be in non-attainment whenever air pollution levels persistently exceed the NAAQS for six pollutants: ozone, lead, carbon monoxide, sulfur dioxide, nitrogen dioxide and particulate matter. Non-attainment counties are under pressure to reduce emissions and this provides an additional incentive for facilities located in these counties to lower their air emissions (Bi and Khanna 2012, Vidovic and Khanna 2012, Gamper-Rabindran and Finger 2013).

To construct our sample we first identified TRI facilities that operate primarily in the chemical manufacturing sector. This resulted in 6,563 facilities in the continental US. We successfully matched 4,245 facilities to 1,929 parent firms by parent firm name. We further restricted the sample to facilities that belong to multi-plant firms in order to be able to instrument for a facility's participation in RC with the characteristics of other plants belonging to the same parent firm. Because the decision to join RC was made by the parent firm rather than any particular facility, restricting the sample to multi-facility firms allows us to distinguish between a firm and a facility level decision from which we can use variation in other facilities belonging to the same firm to exact exogenous variation in firm RC participation. Allowing for one year of lags, we obtain an unbalanced panel of 935 facilities that belong to 352 parent firms between 1996 and

2010. Out of the 935 facilities, 409 facilities belonging to 102 parent firms were members of RC and 526 facilities belonging to 250 parent firms were not members of RC leading to 12,999 facility-year observations.

Table 1 summarizes our data. Comparing facilities that participated in RC to those that did not participate, we find that on average participants have higher total TRI air releases, parent firm TRI air releases, and number of inspections. On the other hand, the participants have a lower facility to firm TRI air release ratio and HAP to TRI emissions ratio. Furthermore, the difference between mean RC facility emissions and mean non-RC facility emissions declines by $5.8*10^4$ lbs after third party certification was introduced in 2005. This difference is statistically significant at the 1% level. However there is no statistical difference in the relative decline rate of emissions between these two groups of facilities between 1996 and 2010.

Figure 1 summarizes the trends in emissions over time for RC participants and non-participants separately. Emissions from all facilities are declining throughout the period between 1996 and 2010, but RC facility emissions are always higher than emissions from non-RC facilities, a result that is consistent with evidence for the early years of RC as well (Gamper-Rabindran and Finger 2013). More importantly, the (unconditional) emissions paths seem to be approximately parallel for the two groups of facilities, especially between 1996 and 2004, the time period prior to the introduction of third party certification.

It is important to note that our sample generally consists of the larger facilities in the chemical manufacturing sector. That is, compared to the approximately 6500 facilities under SIC 28/NAICS 325 that report emissions to the TRI, the year-wise mean and median total air emissions of the 1995 core chemicals for our sample of facilities is larger than the corresponding mean and median emissions for all 6500 chemical facilities. In this sense, our empirical analysis is pertinent to larger chemical facilities and may not generalize to the entire chemical manufacturing sector.

6. Results And Discussion

Analysis of the Traditional ATE

In Table 2 we examine the effect of third party certification on TRI emissions using equation (13) as our regression design under the assumption of (conditional) exogeneity of third party certification and provide a difference-in-differences estimate of the homogenous ATE. This specification exploits the panel structure of our data and includes both time-varying controls and fixed effects. In Models 1 and 2, the dependent variable is TRI air emissions measured in pounds. In Models 3 and 4, the dependent variable is the natural log of TRI air emissions. In Models 3 and 4 we also use the natural log of parent firm TRI emissions: we add one to the annual sums of emissions before taking the log to accommodate zero values.[10] To minimize the possibility of endogeneity, we lag all time varying variables by one year relative to the year in which a facility's TRI emissions are measured. We estimate all models using robust standard errors, bootstrapped and clustered by facility.[11] In Models 2 and 4 we interact time dummies with the treatment indicator to allow the effect of the treatment to change over time.

The coefficient on the treatment dummy (δ) is negative and statistically significant at the 10 percent level in the first model where the dependent variable is TRI emissions in pounds and we do not interact the treatment variable with the year dummies. This indicates that on average facilities that were third party certified under RC reduced their emissions of the TRI chemicals

[10] If facility emissions of a chemical drop below the TRI reporting threshold this is recorded as a zero value in TRI. Firm emissions in a particular year can be zero if all its facilities report zero emissions in that year. There are 360 firm-year observations (or 583 facility-year observations) from 53 unique firms for which firm TRI air releases are zero. There are 469 firm-year observations (or 811 facility-year observations) from 71 unique firms for which firm HAP emissions are zero. These are a small fraction of the total 12,999 facility-year observations in our dataset

[11] We also estimated the regressions clustering the standard errors at the parent firm level. Results are qualitatively the same as those clustered at the facility level, reported in Table 2.

compared to facilities that did not participate in RC and were not independently certified. Once we interact the treatment dummy with the year dummies in Model 2, thus allowing the average treatment effect to vary over time, the coefficient on the treatment dummy now represents the average treatment effect for 2005 and is no longer statistically significant (albeit still negative). The coefficients on the interaction terms between the treatment dummy and the year dummies are also insignificant, except for the final interaction term (treatment*year 2010) which is negative and significant at the 10 percent level. The coefficient on the treatment dummy is not statistically significant in the last two models where the dependent variable is the log of TRI emissions.

To determine the ATE in Models 2 and 4 where the treatment dummy is interacted with the year dummy, we test the hypotheses that the coefficient on the treatment dummy plus the coefficient on each of the interaction terms is statistically different from zero. In Model 2, we find that the ATE is negative and significant in 2008, 2009 and 2010 at the 10 percent, 10 percent and 5 percent levels, respectively. Since emissions were falling throughout the sample period regardless of RC or treatment status, a negative treatment effect in the later years implies that the treated facilities saw even lower emissions in these years compared to the non-treated non-RC facilities.

Models 3 and 4 compare RC and non-RC facilities in terms of the average decline rate of emissions before and after treatment. Model 3 shows that there is no difference between these two groups of facilities. In Model 4, overall treatment is positive and significant at the 5 percent level in 2006 and 2007, but not in any of the other years, including 2005. This suggests that in the early years, the treated facilities saw a lower decline rate in their emissions compared to the non-RC non-treated facilities, but that effect wore off in the later years so that in the later years there is no statistically significant difference in the decline rate of emissions between the two groups.

Comparing across Models 1, 2, 3, and 4 in Table 2 we conclude that there is some evidence that the introduction of third party certification had a negative average treatment effect between 2005 and 2010, and that the treatment effect seems to gather some momentum in the later years (2009-2010) compared to 2005. However, we do not find much evidence that the introduction of third party certification significantly impacted the decline rate in emissions. The fact that the negative treatment effect reported in Models 1 and 2 in Table 2 tends to disappear when we measure our dependent variable as log emissions rather than in levels, suggests that the significance result might be driven by the larger facilities in our data set.

Toffel (2006) and Russo (2009) find that early adopters of ISO 14001 experienced better environmental performance than later adopters. They argue that environmental leaders move quickly when a new opportunity arises that can differentiate them from competitors in terms of environmental performance. Based on their findings we anticipated that RC certification would lead to greater reductions in emissions in the early years of the program. On the contrary, the coefficients on the interaction terms between treatment and year dummies indicate that the benefit of the change in the program structure may have strengthened in later years. This finding may reflect the fact that firms have a three year window in which to be certified and that RC does not publicly release a firm's certification status until the end of that window.

In terms of the control variables, we find that facilities that belong to more polluting firms as measured by the total parent TRI releases, and more polluting facilities within a firm as measured by the facility to parent firm TRI ratio, had significantly higher TRI air releases, as well as the change in TRI releases. Our year indicators are negative and statistically significant in each of our models. This is interesting and important because it constitutes robust evidence that air emissions were gradually falling over the entire sample period, regardless of RC and treatment status. This trend can also be seen in Figure 1 in which we plot average (unconditional) facility emissions over time for treated and untreated facilities separately. This result is not new – Vidovic and Khanna (2007) found a similar trend in emissions reductions for the 33/50 Program that, if not controlled for, confounds the estimate of the program evaluation parameter.

On the other hand, the coefficients on the number of inspections, HAP-TRI ratio, and county non-attainment status are not statistically significant, providing no evidence that the anticipation of more stringent mandatory regulation may have had a negative effect on emissions of the TRI chemicals. These results generally suggest that these control variables do not significantly determine facility emissions. Emissions are highly dependent on productivity, which in turn is dependent on both aggregate demand and facility-specific factors that may not be observable. Such fluctuations are captured in our year indicators and facility fixed effects; in other words, given our set of fixed effects and their apparent significance, it is not surprising that we find less significance of our other control variables.

Analysis of the Traditional ATE via Instrumental Variables

As an extension of the traditional ATE, we estimate the regression described by (13) using instrumental variables. This model maintains the assumption of a homogeneous treatment effect, but allows for endogeneity in third party certification; that is, this model corresponds to the classic selection model where participation in RC is not random but where firms do not select on unobserved gains. Table 3 presents the results from these instrumental variable regressions, using the number of other facilities reporting to the parent firm and a dummy variable for whether the parent firm is publicly traded as instrumental variables. In Model 1 we measure facility emissions as aggregate TRI releases while in Model 2 the dependent variable is the natural log of TRI releases.

We first assess the relevance and validity of our instruments. Effective instruments must satisfy two conditions: they must be correlated with the included endogenous variables (relevant), and they must be orthogonal to the error term (valid). Using the Sargan-Hansen statistic (for overidentification), we fail to reject the null hypothesis that the instruments are jointly uncorrelated with the errors given p-values of 0.795 and 0.261, respectively. Using a Conditional Likelihood Ratio (CLR) test (for weak instruments), we fail to reject the null hypothesis that the excluded instruments are correlated with the included endogenous regressors given the p-values of 0.826 and 0.103, respectively. Based on our assessment, the instruments are both relevant and valid. Note that in the instrumental variables models, we do not interact treatment with the year dummies, because we do not have instrumental variables to mitigate endogeneity arising from multiple endogenous treatment variables.

Contrary to what we report in Table 2, we do not find that treatment by third party certification led to a significant change in emissions in either Models 1 or 2 in Table 3, i.e, once we allow for self-selection into RC and therefore third party certification. Nonetheless, the coefficient estimates on the control variables reported in Table 3, are very similar to the corresponding estimates from Table 2, and remain statistically significant.[12] This is initial evidence that selection into third party certification treatment is not random so that the traditional estimator assuming conditional exogeneity may be overly restrictive and it motivates our consideration of the MTE to assess the nature of firm selection into treatment.

Analysis of the MTE and Essential Heterogeneity

[12] We also estimated versions of this regression using a constructed inverse Mills' ratio to control for endogenous selection into RC. The inverse Mills' ratios were constructed from both random effects and correlated random effects probit regressions. The goodness of fit statistics were similar for both probit models, the inverse Mills ratios were highly correlated, and the results from the regressions were similar. Overall, these additional regressions did not yield qualitatively different conclusions regarding the causal impact of third party certification.

In order to identify the MTE, we require a first stage estimate of the propensity score, which we obtain via a random effects probit model. We report these regression results in Table 4. Among the right hand side variables, we include all of the factors that affect a facility's TRI emissions. Additional variables such as the number of other facilities reporting to the same parent firm as the facility in question, and a dummy variable for whether the parent firm is publicly traded serve as instruments for selection. In Model 2 of Table 4, we include year dummies. The results indicate that the two instruments are positive and statistically significant at the 1 percent level. All of the other factors are statistically insignificant.

As we described above, the support over which we estimate the propensity score determines the range of margins over which the MTE can be identified. We plot the support of the estimated propensity scores from Model 1 in Table 4 (which was chosen over Model 2 because of better fit statistics) for both treated and untreated facilities in Figure 2 – note that the estimated propensity score has full support, which ensures that we can identify the MTE over all margins of self-selection. Recall that the MTE calculated at low values of the propensity score reflects the average treatment effect for facilities that are most likely to self-select into RC (and hence be treated by third party certification), and the MTE calculated at high values of the propensity score identifies the treatment effect for facilities that are least likely to self-select into RC and be subject to third party certification starting in 2005.

Table 5 reports the estimated parameters from the cubic polynomial regressions for emissions measured in levels (Model 1) and logs (Model 2). We see that the propensity score polynomial and most interaction terms are statistically significant in each specification. The table reports results from an F-test of joint significance on the higher order polynomial terms of the propensity score, and for both models we reject the null hypothesis that the higher order terms are jointly insignificant. This confirms that firms select into RC in part on the unobservable gain from participation, a result that has not yet been established empirically in the voluntary abatement literature.

We plot the estimated MTE in Figure 3 for facility emissions measured in levels (top panel) and logs (bottom panel). The dashed line represents a 95 percent bootstrapped confidence interval around the estimated MTE. In both panels, the estimated MTE takes an inverted U-shape. For low levels of the propensity score, the MTE represents the average effect of third party certification for facilities that are most likely to be subject to third party certification. In both models, the MTE over this range is not significant, which is consistent with our standard instrumental variables results that find insignificance of treatment. It is interesting that the estimated MTE for the levels specification is significantly positive over a substantial range of the support of the propensity score, which implies that facilities that can be induced to participate over moderate manipulations in the instrumental variables (the mid-range of the function) significantly increase emissions following third party certification. Finally, facilities that are least likely to participate (high levels of propensity score) do not adjust emissions if mandated to be third party certified. The bottom panel in Figure 3 shows that while the estimated MTE becomes positive over the middle points of support of the propensity score, the MTE is never statistically significant.

The ATE can be obtained by computing the unweighted average of the MTE across all observations. Though the MTE captures heterogeneity in the average effect of treatment, calculating the ATE is one way of facilitating a direct comparison of the MTE model to the difference-in-differences models that assume conditional exogeneity or restrict selection to being only on levels. We find that the ATE from the MTE model with emissions in levels is 30913.31 with a standard error of 21034.03, and the ATE for the log model is -0.09 with a standard error of

-0.08. That is, the ATEs obtained from the MTE models are both insignificant, a result consistent with the homogenous ATE obtained under the assumption of selection on levels.

To provide further insight into characteristics of facilities that appear along different margins of the propensity score, we compare distributions of each of our covariates for samples of facilities that have a propensity score less than 0.4, between 0.4 and 0.7, and greater than 0.7 for the model with emissions measured in levels. We find that the number of inspections, the number of gasses for which the county is out of attainment, parent firm TRI emissions, the facility to firm TRI release ratio, and the HAP to TRI ratio are not significantly different across the three groups. Hence, these variables do not explain why different firms end up along different margins of self-selection. We find that there are substantial differences in TRI air emissions across these three groups: firms that are most likely to self-select into the RC program are those with the lowest facility emissions. For this group, the average emissions is 99,740 lbs, with a median of 1,505 lbs and an interquartile range of [10, 15,160]. The distributions of emissions for facilities with medium and high values of the propensity score are not statistically significantly different from each other; the interesting fact is that these groups have emissions that are substantially higher than the facilities with the lowest propensity scores. The average levels of emissions for the medium and high groups, respectively, are 184,800 lbs and 125,500 lbs, with medians 16,900 lbs and 9,598 lbs and interquartile ranges [1,506, 87,240] and [951, 61,620]. Our results imply that, at best, these facilities are unlikely to change their emissions following third party certification and, at worst. may further increase their already high emissions. One possibility is that the production technology for these large facilities is more pollution intensive, and there may be fewer substitute technologies or inputs available through which to adopt and satisfy program commitments.

As an additional check, we investigate whether there are any significant changes in emissions across these three groups over time. Recall that RC firms are required to be third party certified every three years; the first three year period ended by the end of 2007, and the second period by 2010. While one might expect that a firm is more likely to meet RC standards at the end of each three year interval, we do not find any evidence that emissions change significantly within any post-third party certification year.

These insights are contrary to the literature. For example, Vidovic and Khanna (2012) find that more polluting facilities are more likely to join the 33/50 program whereas King and Lennox (2000) similarly find that dirtier firms were more likely to join RC. However no study in the voluntary abatement literature prior to ours considers selection on participation gain and our contrarian results suggest that more work is needed to assess the nature of self-selection into specific voluntary pollution abatement programs. From a policy perspective, our results are especially troubling because they call into question the effectiveness of third party certification as a tool to elicit meaningful reductions in emissions,

7. Conclusion

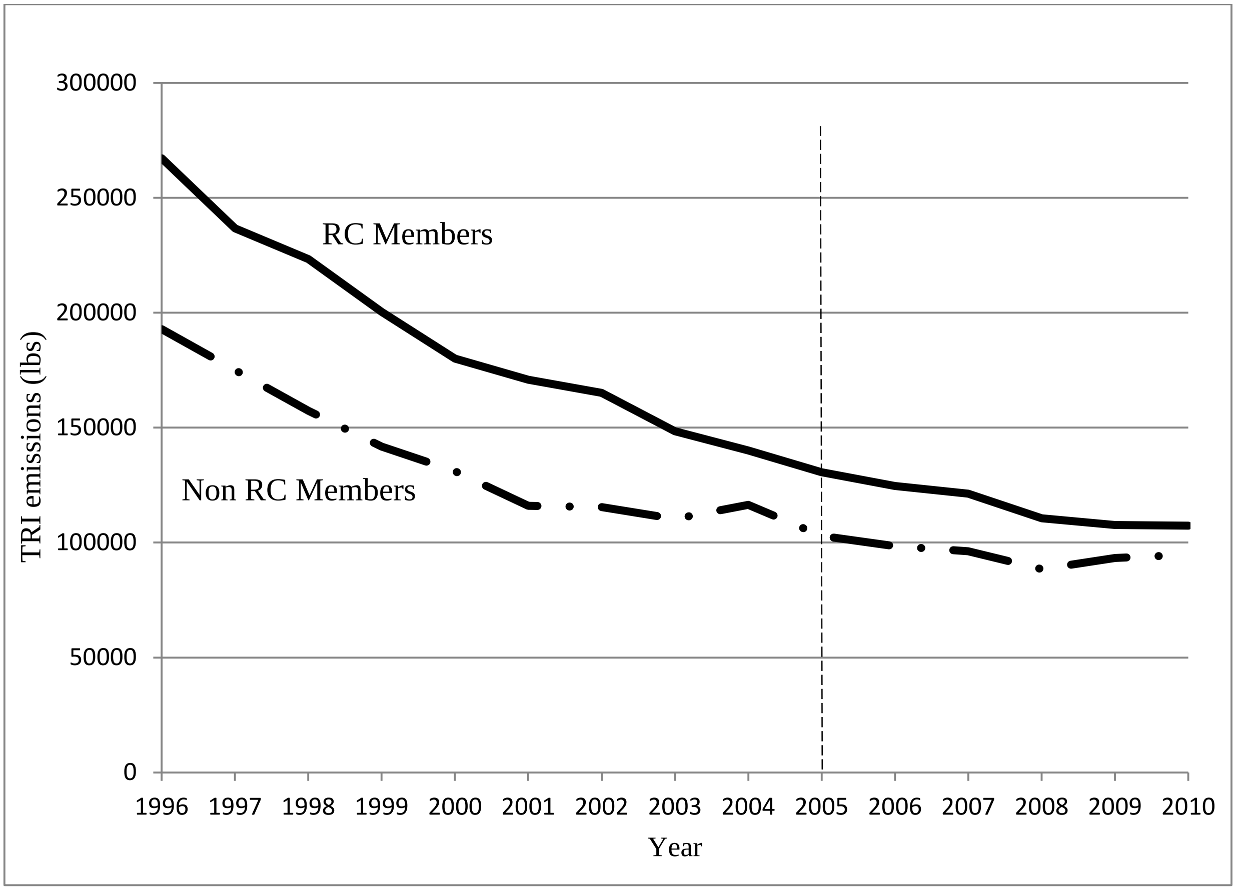
Since the tragic explosion at the Union Carbide facility in Bhopal, India, the ACC has made a very public effort to improve its environmental performance under the aegis of its flagship self-regulation program, Responsible Care. While the ACC claims that this program has been very successful in lowering hazardous emissions – by as much as 76% between 1988 and 2011 (http://responsiblecare.americanchemistry.com/FactSheet, accessed June 19, 2014), the academic literature has been more skeptical. A distinguishing feature of RC that set it apart from other self-regulation programs and especially from its primary competitor, ISO 14001, was the lack of third party oversight. The credibility of RC was suspect because this voluntary program lacked a transparent monitoring and enforcement mechanism. Perhaps in response to this and as an attempt to enhance the credibility of the RC brand, RC introduced mandatory third party certification for all participants from 2005 onwards. This marks the most significant structural change in RC since its inception.

Our goal in this paper is to assess whether the introduction of third party certification causes a facility to improve its environmental performance relative to an average facility in the chemical industry that is not in RC and not subject to third party certification, thereby enhancing the credibility of RC. To do this, we estimate the causal effect of third party certification on facility emissions using three separate estimators that allow us to assess the extent to which firms self-selected endogenously into the RC program, and if so, the nature of this selection. Our results indicate that RC firms do significantly select into the program, in part because of unobservable gains. Hence, our preferred model is the MTE model, through which we find that facilities in the RC program did not reduce emissions following third party certification relative to facilities that were not in the RC and were not third party certified. These results are consistent with our standard linear instrumental variables estimates. We find some discouraging evidence that facilities less likely to volunteer to participate in RC may increase emissions if induced to participate.

The policy implication of the latter finding is sobering. Our analyses indicate that third party certification *at best* did not significantly influence facility emissions, and at worst may work to increase them if certain facilities were given a strong enough incentive to participate, despite not being the most eager facilities to join. These results are not reassuring – we do not find evidence that third party certification is an effective means for improving the credibility of RC, and at worst our results indicate that policymakers may be advised not to encourage reluctant facilities to participate in RC and be subjected to third party certification.

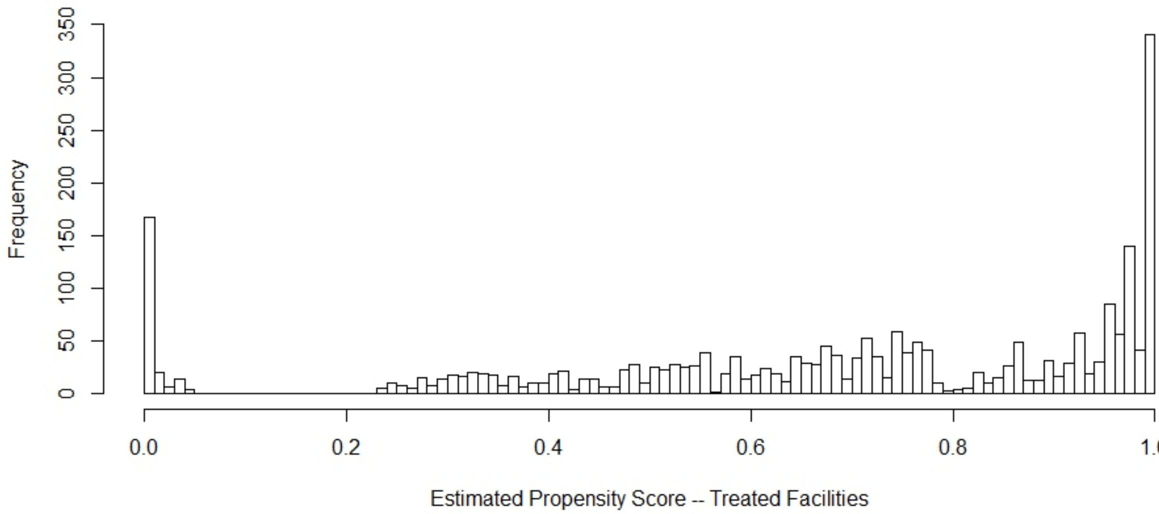
While third party certification is an important modification designed to overcome potential criticisms of voluntary abatement efforts, little econometric attention has been paid to assessing its effectiveness. Our analysis is the first in the literature to assess the nature of firm selection into voluntary abatement programs, and is the first to estimate the causal effect of third party certification on facility emissions. Much research has focused on voluntary pollution abatement programs in general, with mixed conclusions regarding their effectiveness. Our results suggest that this sensitivity may be driven by heterogeneity that is not accounted for in typical econometric models.

Figure 1: Average Trends in Emissions for Treated and Untreated Facilities



Note: Means of core 1995 TRI emissions by year and treatment status using facilities in the sample.

Figure 2: Estimated Propensity Score for Treated and Untreated Facilities



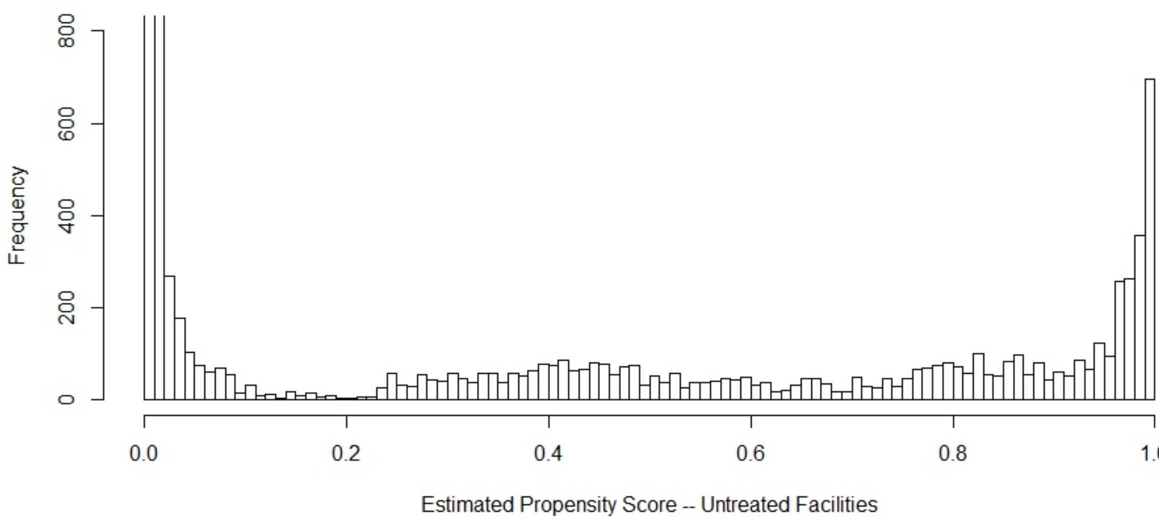
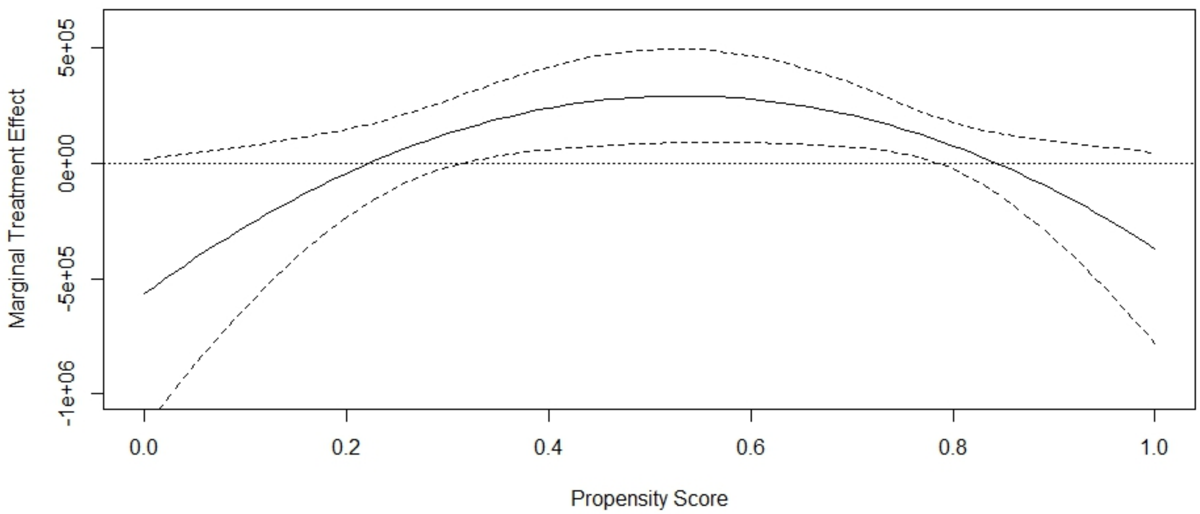
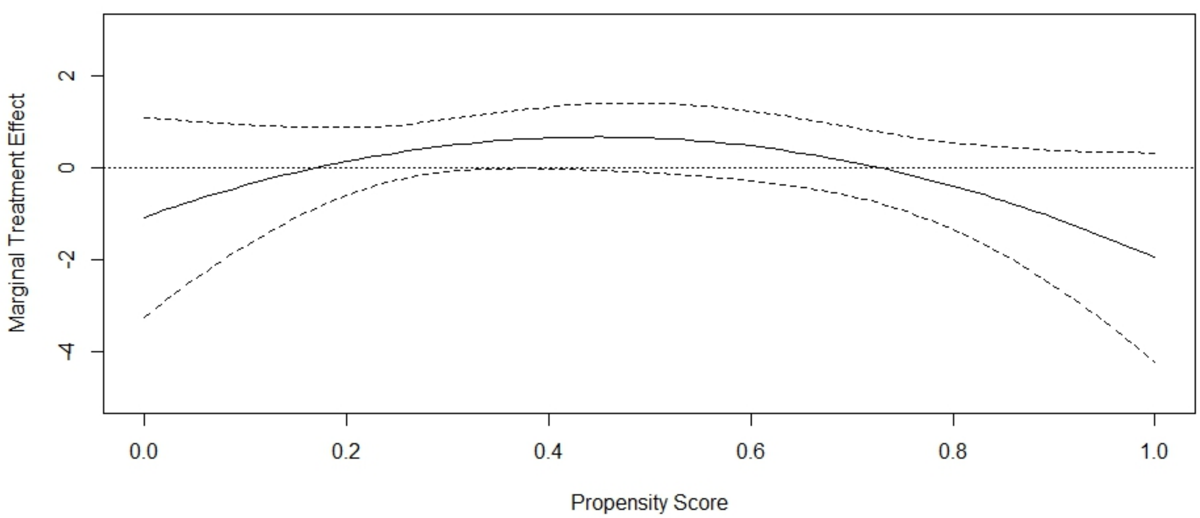


Figure 3: Estimated Marginal Treatment Effect of Third Party Certification



(a) Estimated MTE and 95 percent confidence bound for the model with emissions measured in levels



(b) Estimated MTE and 95 percent confidence bound for the model with emissions measured in logs

Table 1: Descriptive Statistics

Variable	Treatment Group	Control Group	Difference Between Groups
TRI releases (lbs)			
Mean	162302.30	90019	72283.30***
Standard deviation	467647.60	482150.10	
Median	15571	1349	14222
Facility to firm TRI releases			
Mean	0.12	0.22	-0.11***
Standard deviation	0.21	0.31	
Median	0.02	0.06	-0.03
Parent firm TRI releases (lbs)			
Mean	3024820	455229.10	2569591***
Standard deviation	4802283	14962510	
Median	923748	37289.33	886458.70
HAP-TRI ratio			
Mean	0.74	2.09	-1.36*
Standard deviation	3.57	5.93	
Median	0.71	0.67	0.04
Number of inspections			
Mean	0.81	0.45	0.36***
Standard deviation	3.67	1.42	
Median	0	0	0
County non-attainment status			
Mean	0.88	0.90	-0.02
Standard deviation	0.98	1.06	
Median	1	1	0
Facility-year observations	**5823**	**7176**	

Note: *** indicates means are statistically significantly different at the 1% level, ** at the 5% level, and * at the 10% level.

Table 2: Estimate of the Impact of Third Party Certification on TRI Air Releases: Exogenous Treatment

Variable	Model 1 TRI releases	Model 2 TRI releases	Model 3 Log of TRI releases	Model 4 Log of TRI releases
Treatment	-27585.10*	-18020.60	0.05	0.10
	(15253.78)	(13746.15)	(0.09)	(0.09)
Year 1997	-9882.45*	-9914.75*	-0.16***	-0.16***
	(5484.46)	(5483.93)	(0.04)	(0.04)
Year 1998	-13798.40	-13846.30	-0.22***	-0.22***
	(9015.07)	(9007.19)	(0.05)	(0.05)
Year 1999	-25359.00**	-25416.00***	-0.25***	-0.25***
	(9800.58)	(9791.77)	(0.06)	(0.06)
Year 2000	-28594.60***	-28677.00***	-0.30***	-0.30***
	(10277.14)	(10264.14)	(0.07)	(0.07)
Year 2001	-42806.30	-42896.60***	-0.38***	-0.38***
	(10072.37)	(10061.68)	(0.07)	(0.07)
Year 2002	-41045.60***	-41137.70	-0.39***	-0.39***
	(10629.22)	(10613.06)	(0.07)	(0.07)
Year 2003	-43825.00***	-43929.60***	-0.45***	-0.45***
	(12244.76)	(12236.03)	(0.07)	(0.07)
Year 2004	-36544.50**	-36660.20***	-0.47***	-0.47***
	(14335.48)	(14331.17)	(0.08)	(0.08)
Year 2005	-39744.70***	-44020.60***	-0.51***	-0.54***
	(14053.35)	(13355.97)	(0.09)	(0.09)
Year 2006	-41512.50***	-45320.70***	-0.65***	-0.72***
	(14572.96)	(14379.98)	(0.09)	(0.09)
Year 2007	-43358.40***	-45315.90***	-0.75***	-0.83***
	(14194.19)	(14103.79)	(0.09)	(0.09)
Year 2008	-49652.10***	-48542.90***	-0.68***	-0.66***
	(14742.65)	(14468.14)	(0.09)	(0.10)
Year 2009	-40787.70**	-37333.20**	-0.86***	-0.79***
	(16962.40)	(18947.05)	0.10)	(0.10)
Year 2010	-40104.60**	-35109.30*	-0.83***	-0.75***
	(16654.27)	(18543.67)	(0.10)	(0.11)
Treatment*year 2006	-	-1148.51	-	0.10
	-	(5860.60)	-	(0.08)
Treatment*year 2007	-	-5317.06	-	0.11
	-	(11156.74)	-	(0.09)
Treatment*year 2008	-	-12290.50	-	-0.08
	-	(8583.01)	-	(0.09)

Treatment*year 2009	-	-17715.60	-	-0.21**
	-	(14385.31)	-	(0.10)
Treatment*year 2010	-	-21215.50*	-	-0.23**
	-	(12857.97)	-	(0.11)
Facility to firm TRI ratio$_{(-1)}$	135847.90***	135935.80***	2.49***	2.49***
	(22243.88)	(22241.90)	(0.17)	(0.17)
Parent firm TRI releases$_{(-1)}$	0.02***	0.02***	0.20***	0.20***
	(0.00)	(0.00)	(0.03)	(0.03)
HAP-TRI ratio $_{(-1)}$	6.55	5.59	0.00	0.00
	(1326.30)	(1313.91)	(0.04)	(0.04)
Number of inspections$_{(-1)}$	-29.54	-15.72	-0.00	-0.00
	(3072.62)	(3078.01)	(0.01)	(0.01)
County non-attainment$_{(-1)}$	3939.86	3819.16	-0.05	-0.05
	(5726.39)	(5741.07)	(0.05)	(0.05)
Constant	99837.66***	100097.60***	5.34***	5.33***
	(17188.37)	(17178.11)	(0.42)	(0.43)
Number of observations	12,999	12,999	12,999	12,999
Number of groups	935	935	935	935
Tests of linear hypotheses: Treatment + interaction terms				
Treatment + Treatment*year 2006	-	-19169.06	-	0.20**
	-	(14793.34)	-	(0.09)
Treatment + Treatment*year 2007	-	-23337.62	-	0.21**
	-	(17102.56)	-	(0.10)
Treatment + Treatment*year 2008	-	-30311.02*	-	0.02
	-	(16376.30)	-	(0.11)
Treatment + Treatment*year 2009	-	-35736.10*	-	-0.11
	-	(20381.06)	-	(0.12)
Treatment + Treatment*year 2010	-	-39236.05**	-	-0.13

Note: *** indicates statistical significance at the 1% level, ** at the 5% level, and * at the 10% level. Estimation via facility-level fixed effects with bootstrapped robust standard errors clustered on facilities are in parentheses. In all models the number of inspections, HAP-TRI, facility to firm TRI, parent firm TRI and the number of gases for which a facility's county is out of attainment with NAAQS are lagged by one year relative to the year in which the dependent variable is measured. Parent firm TRI emissions are measured in natural logs in Models 3 and 4. All other variables are in levels.

Table 3: Estimate of the Impact of Third Party Certification on TRI Air Releases: Instrumental Variable Approach

Variable	Model 1 TRI releases	Model 2 Log of TRI releases
Treatment	-5133.71	1.16
	(117596.40)	(1.01)
Year 1997	-9630.37*	-0.16***
	(5752.99)	(0.04)
Year 1998	-13353.20	-0.22***
	(10060.84)	(0.05)
Year 1999	-24785.90**	-0.25***
	(11254.57)	(0.06)
Year 2000	-27866.59**	-0.30***
	(12160.84)	(0.07)
Year 2001	-42060.67***	-0.39***
	(11822.53)	(0.07)
Year 2002	-40389.95***	--0.40***
	(12123.26)	(0.07)
Year 2003	-43233.18***	-0.46***
	(13464.35)	(0.08)
Year 2004	-36003.87**	-0.49***
	(15272.28)	(0.08)
Year 2005	-48920.07	-1.02**
	(45774.36)	(0.47)
Year 2006	-50663.33	-1.15**
	(45650.69	(0.47)
Year 2007	-52512.63	-1.26***
	(44939.62)	(0.46)
Year 2008	-58754.13	-1.18**
	(44984.26)	(0.46)
Year 2009	-49668.43	-1.36***
	(41510.15)	(0.46)
Year 2010	-48996.59	-1.33***
	(41830.84)	(0.46)
Facility to firm TRI ratio$_{(-1)}$	136642.20***	2.51***
	(22903.00)	(0.17)
Parent firm TRI releases$_{(-1)}$	0.02***	0.20***
	(0.01)	(0.03)
HAP-TRI ratio$_{(-1)}$	9.95	0.00
	(1314.94)	(0.04)

Number of inspections$_{(-1)}$	-66.70	-0.01
	(2996.62)	(0.01)
County non-attainment$_{(-1)}$	3889.47	-0.05
	(6199.05)	(0.06)
Number of observations	12,999	12,999
Number of groups	935	935
Sargan-Hansen statistic (overidentification)	0.067	1.263
(*Chi-square p-value*)	(0.795)	(0.261)
CLR test (weak instruments)	0.05 (0.83)	2.81
(*Chi-square p-value*)	(0.826)	(0.103)

Note: *** indicates statistical significance at the 1% level, ** at the 5% level, and * at the 10% level. Estimation via instrumental variables facility-level fixed effects regression with bootstrapped standard errors clustered on facilities are in parentheses. In all models the number of inspections, HAP-TRI, facility to firm TRI, parent firm TRI, and the number of gases for which a facility's county is out of attainment with NAAQS are lagged by one year relative to the year in which the dependent variable is measured. Parent firm TRI emissions are measured in natural logs in Model 2. All other variables are in levels. We instrument for treatment with the number of other facilities reporting and a dummy variable equal to 1 if the parent firm is publicly owned.

Table 4: Random Effects Probit Model of Facility Participation in RC

Variable	Model 1	Model 2
Number of other facilities$_{(-1)}$	0.09***	0.09***
	(0.02)	(0.03)
Public Firm	1.95***	2.09***
	(0.45)	(0.51)
Year 1997	-	0.01
	-	(0.91)
Year 1998	-	-0.02
	-	(0.92)
Year 1999	-	-0.17
	-	(0.99)
Year 2000	-	-0.27
	-	(1.02)
Year 2001	-	-0.32
	-	(1.02)
Year 2002	-	-0.36
	-	(1.03)
Year 2003	-	-0.40
	-	(1.01)
Year 2004	-	-0.40
	-	(0.98)
Year 2005	-	-0.30
	-	(0.96)
Year 2006	-	-0.37
	-	(0.97)
Year 2007	-	-0.46
	-	(0.93)
Year 2008	-	-0.45
	-	(0.94)
Year 2009	-	-0.48
	-	(0.92)
Year 2010	-	-0.45
	-	(0.91)
TRI air releases$_{(-1)}$	4.44E-08	-1.18E-08
	(3.40E-07)	(3.23E-07)
HAP-TRI ratio $_{(-1)}$	-6.63E-05	6.31E-05
	(0.01)	(0.01)

Number of inspections $_{(-1)}$	-2.72E-04	2.38E-03
	(0.05)	(0.06)
County non-attainment$_{(-1)}$	-0.02	-0.02
	(0.18)	(0.18)
Constant	-2.68***	-2.47***
	(0.54)	(0.82)
Log likelihood	-676.60	-675.70
AIC	1369.19	1395.40
BIC	1428.97	1559.80

Note: *** indicates statistical significance at the 1% level, ** at the 5% level, and * at the 10% level. Standard errors are in parentheses. All variables are in levels. All time varying variables are lagged by one year relative to the year in which the dependent variable is measured. Variables that serve as instruments (excluded in the main difference-in-difference specification) are the number of facilities reporting and the public firm dummy.

Table 5: Estimate of the Impact of Third Party Certification on TRI Air Releases: Cubic Polynomial Local Instrumental Variable Approach

Variable	**Model 1** **TRI releases**	**Model 2** **Log of TRI releases**
Facility to firm TRI ratio$_{(-1)}$	6985.98	3.21***
	(14751.89)	(0.13)
Parent firm TRI releases$_{(-1)}$	0.06***	0.18***
	(0.00)	(0.01)
HAP-TRI ratio$_{(-1)}$	6.60	0.42***
	(52.27)	(0.06)
Number of inspections$_{(-1)}$	8722.93***	0.02*
	(2027.04)	(0.01)
County non-attainment$_{(-1)}$	4364.48	-0.00
	(5849.48)	(0.04)
Propensity Score	-643977.55***	-3.78***
	(95605.57)	(0.74)
Propensity Score^2	1612059.13***	3.88***
	(226278.85)	(1.45)
Propensity Score^3	-1008325.72***	-2.88***
	(147390.52)	(0.95)
P-Score * Facility to firm TRI ratio$_{(-1)}$	892184.41***	3.22***
	(49307.10)	(0.42)
P-Score * Parent firm TRI releases$_{(-1)}$	-0.04***	0.19***
	(0.00)	(0.04)
P-Score * HAP-TRI ratio$_{(-1)}$	24.94	0.50***
	(1038.47)	(0.15)
P-Score * Number of inspections$_{(-1)}$	-10508.96***	-0.03**
	(2385.49)	(0.02)
P-Score * County non-attainment$_{(-1)}$	-263.95	-0.10*
	(8395.51)	(0.05)
Number of observations	12,999	12,999
Number of groups	935	935
R Squared	0.08	0.12
F-Test for P-Score Polynomial (*p-value*)	27.71 (0.00)	6.12 (0.00)

Note: *** indicates statistical significance at the 1% level, ** at the 5% level, and * at the 10% level. The regression is estimated via facility-level fixed effects. The number of inspections, HAP-TRI, facility to firm TRI, parent firm TRI, and the number of gases for which a facility's county is out of attainment with NAAQS are lagged by one year relative to the dependent variable. Parent firm TRI emissions are measured in natural logs in Model 2. All other variables are in levels.

References

American Chemistry Council (ACC) 2013. Responsible Care. http://responsiblecare.americanchemistry.com/

Arora, S., and Cason, T., 1996. Why do firms volunteer to exceed environmental regulations? Understanding participation in EPA's 33/50 program, *Land Economics*, 72: 413–432.

Arora, S., Gangopadhyay, S., 1995. Toward a theoretical model of voluntary overcompliance, *Journal of Economic Behavior and Organization*, 28: 289–309.

Bertrand, M., Duflo, E. and Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? *Quarterly Journal of Economics*, 119 (1): 249-275.

Bi, X. and Khanna, M., 2012. Re-assessment of the impact of EPA's voluntary 33/50 program on toxic releases, *Land Economics*, 88 (2): 341–361.

Björkland, A. and Moffitt, R., 1987. The estimation of wage gains and welfare gains in self-selection models, *Review of Economics and Statistics*, 69: 42-49.

Bui, L. and Kapon, S., 2012. The impact of voluntary programs on polluting behavior: evidence from pollution prevention programs and toxic releases, *Journal of Environmental Economics and Management*, 64: 31-44.

Carneiro, P., Heckman, J. J. and Vytlacil, E. J. 2011. Estimating marginal returns to education, *American Economic Review*, 101: 2754-2781.

Carrión-Flores, C. E., Innes, R. and Sam, A. 2013. Do voluntary pollution reduction programs (VPRs) spur or deter environmental innovation? Evidence from 33/50, *Journal of Environmental Economics and Management*, 66: 444–459.

Darnall, N. and Carmin, J., 2005. Cleaner and greener? The signaling accuracy of U.S. voluntary environmental programs, *Policy Sciences*, 38: 71–90.

Gamper-Rabindran S., 2006. Did the EPA's voluntary industrial toxics program reduce emissions? A GIS analysis of distributional impacts and by-media analysis of substitution, *Journal of Environmental Economics and Management*, 52: 391-410.

Gamper-Rabindran S. and Finger, S., 2013. Does self-regulation reduce pollution? Responsible Care in the chemicals industry, *Journal of Regulatory Economics*, 43: 1-30.

Heckman, J. J., Tobias, J. L. and Vytlacil, E., 2001. Four parameters of interest in the evaluation of social programs, *Southern Economic Journal*, 68: 210-223.

Heckman, J. J. and Vytlacil, E., 1999. Local instrumental variables and latent variable models for identifying and bounding treatment effects, *Proceedings of the National Academy of*

Sciences, 96: 4730-4734.

Heckman J. J. and Vytlacil, E., 2001. Local instrumental variables, in *Nonlinear Statistical Modeling: Proceedings of The Thirteenth International Symposium in Economic Theory in Econometrics: Essays in Honor of Takeshi Amemiya*, eds. C. Hsiao, K. Morimume and J. L. Powell, New York: Cambridge University Press, 1-46.

Heckman, J. J. and Vytlacil, E., 2005. Structural equations, treatment effects, and econometric policy evaluation, *Econometrica*, 73: 669-738.

Heckman, J. J., Urzua, S. and Vytlacil, E., 2006. Understanding instrumental variables in models with essential heterogeneity, *Review of Economics and Statistics*, 88: 389-432.

Heckman, J. J., Schmierer, D. and Urzua, S. 2010. Testing the correlated random coefficient model, *Journal of Econometrics*, 158: 177-203.

Henriques, I., and Sadorsky, P., 1996. The determinants of an environmentally responsive firm: an empirical approach, *Journal of Environmental Economics and Management*, 30 (3): 381-395.

Holland, P., 1986. Statistics and causal inference, *Journal of the American Statistical Association*, 81: 945-970.

Innes, R. and Sam, A., 2008. Voluntary pollution reductions and the enforcement of environmental law: an empirical study of the 33/50 program, *Journal of Law Economics*, 51: 271 – 296.

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

Khanna, M., Quimio, W., and Bojilova, D., 1998. Toxic release information: a policy tool for environmental protection, *Journal of Environmental Economics and Management*, 36 (3): 243-266.

King, A. and Lenox, M., 2000. Industry self-regulation without sanctions: the chemical industry's Responsible Care program, *Academy of Management Journal*, 43: 698–716.

King, A., Lenox, M. and Terlaak, A., 2005. The strategic use of decentralized institutions: exploring certification with the ISO 14001 management standard, *Academy of Management*, 48: 1091–1106.

Maxwell, J. W., Lyon, T., and Hackett, S., 2000. Self-regulation and social welfare: the political economy of corporate environmentalism, *Journal of Law and Economics*, 63 (43): 583-617.

Melnyk, S. A., Sroufe, R. P., Calantone, R. L. and Montabon, F.L., 2002. Assessing the effectiveness of US voluntary environmental programmes: an empirical study, *International Journal of Production Research*, 40: 1853– 1878.

Meyer, B. D., 1995. Natural and quasi-experiments in economics, *Journal of Business and Economic Statistics*, 13 (2): 151-161.

Montabon, F., Melnyk, S. A., Sroufe, R. and Calantone, R.J., 2000. ISO 14000: assessing its perceived impact on corporate performance, *The Journal of Supply Chain Management*, Spring: 4–15.

Phillips, D., 2006. RC 14001 certification: an integrated management systems approach, Q*uality Digest*, May.

Potoski, M. and Prakash, A., 2005. Covenants with weak swords: ISO 14001 and facilities' environmental performance, *Journal of Policy Analysis and Management*, 24: 745–769.

Prakash, A., 2000. Responsible Care: an assessment, *Business and Society*, 39 (2): 183-209.

Rao, P. and Hamner, B., 1999. Impact of ISO 14000 on business performance, *working paper*.

Rivera, J., deLeon, P. and Koerber, C., 2006. Is greener whiter yet? The sustainable slopes program after five years, *Policy Studies Journal*, 34 (2): 195–224.

Rondinelli, D. A. and Vastag, G., 2000. Panacea, common sense, or just a label? The value of ISO 14001 environmental management systems, *European Management Journal*, 18: 499–510.

Sam, A., Khanna, M. and Innes, R., 2009. Voluntary pollution reduction programs, environmental management, and environmental performance: an empirical study, *Land Economics*, 85 (4): 692-711.

Segerson, K., and Miceli, T., 1998. Voluntary environmental agreements: good or bad news for environmental protection? *Journal of Environmental Economics and Management*, 36 (2): 109-130.

Toffel, M. W., 2006. Resolving information asymmetries in markets: the role of certified management programs, *Harvard Business School Working Paper*.

Vidovic, M. and Khanna, N., 2007. Can voluntary pollution prevention programs fulfill their promises? Further evidence from the EPA's 33/50 program, *Journal of Environmental Economics and Management*, 53: 180-195.

Vidovic, M. and Khanna, N., 2012. Is voluntary pollution abatement in the absence of a carrot or stick effective? Evidence from facility participation in the EPA's 33/50 program, *Environmental Resource Economics*, 52: 369-393.

Wooldridge, J., 2011. *Econometric Analysis of Cross-Section and Panel Data*. 2nd Edition, MIT Press: Cambridge.