DISCUSSION PAPER

July 2008; revised May 2010 ■ RFF DP 08-13-REV

The Performance of Voluntary Climate Programs

Climate Wise and 1605(b)

William A. Pizer, Richard Morgenstern,
and Jhih-Shyang Shih

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

RESOURCES
FOR THE FUTURE

The Performance of Voluntary Climate Programs: Climate Wise and 1605(b)

William A. Pizer, Richard Morgenstern, and Jhih-Shyang Shih

Abstract

Despite serving as the principal basis of U.S. climate policy over the past two decades, corporate voluntary environmental programs have been subject to quite limited evaluation. The self-selection of participants—an essential element of such initiatives—poses particular challenges to researchers because the decision to participate may not be random and, in fact, may be correlated with the outcomes. The present study is designed to overcome these problems by gauging the environmental effectiveness of two early voluntary climate change programs with established track records, the U.S. Environmental Protection Agency's Climate Wise program and the U.S. Department of Energy's Voluntary Reporting of Greenhouse Gases Program, or 1605(b). Both programs provide quite flexible criteria for firms to participate. Particular attention is paid to the participation decision and how various assumptions affect estimates of program outcomes using propensity score matching methods applied to plant-level Census data.

Overall, we find quite modest effects: the reductions in fuel and electricity expenditures from Climate Wise and 1605(b) are no more than 10 percent and probably less than 5 percent. Virtually no evidence suggests a statistically significant effect of either Climate Wise or 1605(b) on fuel costs. Some evidence indicates that participation in Climate Wise led to a slight (3–5 percent) increase in electricity costs that vanished after two years. Stronger evidence suggests that participation in 1605(b) led to a slight (4–8 percent) decrease in electricity costs that persisted for at least three years.

Key Words: voluntary, regulation, energy, climate change

JEL Classification Numbers: Q2, Q4

Contents

The Performance of Voluntary Climate Programs: Climate Wise and 1605(b)

William A. Pizer, Richard Morgenstern, and Jhih-Shyang Shih[1]

I. Introduction

The energy- and greenhouse gas (GHG)–related corporate voluntary environmental programs that have served as the principal tool of U.S. climate policy for almost two decades have been subject to limited *ex post* evaluation. The self-selection of participants – an essential element of such initiatives – poses particular challenges to researchers because the decision to participate may not be random and, in fact, may be correlated with the outcomes. That is, simple comparisons of outcomes between participants and nonparticipants may have less to do with program effects and more to do with other differences between the two groups. The present study is designed to overcome these problems by using propensity score matching applied to plant-level Census data to gauge the environmental effectiveness of two early voluntary climate change programs with established track records, the U.S. Environmental Protection Agency's (EPA) Climate Wise program and the U.S. Department of Energy's (DOE) Voluntary Reporting of Greenhouse Gases Program, or 1605(b). Particular attention is paid to the participation decision and how various assumptions affect estimates of program outcomes.

The single voluntary program most subject to *ex post* evaluation – EPA's 33/50 program – is not related to energy or GHGs at all: studies of this program have benefited from the existence of the nonvoluntary Toxics Release Inventory, which served as both a source of outcome measures and a universe of observations from which to construct control and treatment groups. Interestingly, evaluations of 33/50 have yielded conflicting results. An early study found that participants reduced their emissions by 28 percent compared with nonparticipants (Khanna and Damon 1999), but a more recent paper finds no significant difference between the two groups after accounting for the Montreal Protocol requirements (Gamper-Rabindran 2006).

There are at least three reasons why the incentives for emissions reduction are different in the case of toxics versus energy and GHGs. First, since toxics are typically of local or regional

[1] Pizer: U.S. Treasury Department; his work on this paper was completed while he was a full-time senior fellow at Resources for the Future. The views expressed herein do not necessarily reflect those of the Treasury Department or the U.S. government. Morgenstern and Shih: Senior Fellow and Fellow, Resources for the Future.

concern, the prospects for local recognition for emissions reduction are probably greater than for GHGs. Second, toxics can have direct, acute impacts as well as long-term, chronic effects, such as cancer and heart disease. Meanwhile, GHGs accumulate for many years, affect the climate, and thereby affect ecosystems and overall human well-being over the longer term, often in less tangible ways. Third, with no practical opportunity for end-of-pipe abatement, reductions in energy-related GHG emissions often amount to reductions in energy use itself, which in turn, given the underlying positive price on energy, results in fuel cost savings. At the same time, toxic emissions are often an unpriced industrial byproduct whose existence was widely ignored until the Bhopal accident in the late 1980s introduced a significant threat of regulation. For all these reasons, firms may have greater incentives as well as more low-cost opportunities to reduce toxic releases than to cut energy usage or GHG emissions.

A key feature of this paper is its use of propensity score matching (Rosenbaum and Rubin 1983), a technique not commonly employed in the literature on the evaluation of voluntary environmental programs. This matching approach is contrasted with the more widely used two-stage method of Heckman and Hotz (1985). As described below, the former method puts greater emphasis on matching participants and nonparticipants, and the latter emphasizes model specification. As a source of control information, we are fortunate to have access to the confidential files of the Census Bureau's Longitudinal Research Database (LRD). Despite the relatively long track record of the two programs examined, the nature and breadth of information available on the control group, and the use of alternative modeling approaches, we find quite modest effects of the Climate Wise and 1605(b) programs.

The plan of the paper is straightforward. Section II briefly reviews the relevant literature on energy- and GHG-related voluntary programs. Section III provides background on the two programs examined, Section IV outlines the data and models used in the analysis, and Section V presents the results. Section VI discusses the conclusions.

II. Literature on Energy- and GHG-Related Voluntary Programs

The early energy- and GHG-related corporate voluntary programs date to the President George H.W. Bush era, but the largest number of such programs were initiated during the Clinton administration, many in the Climate Change Action Plan in 1993. Most of these programs seek to spur adoption of particular technologies; others focus on emissions reductions per se, without regard to specific technologies. The former include such well-known programs as Green Lights, Motor Challenge, and Energy Star Buildings. Over the years, many changes in program organization and design have occurred. Energy Star, for example, has expanded considerably and now covers 45 kinds of products. Climate Wise and 1605b count among the nontechnology-oriented programs developed in the early 1990s. A major challenge in evaluating both types of programs is the establishment of a credible baseline. Most assessments, especially among the technology-oriented programs, have relied on *constructed* baselines.[2] This contrast with the approach of the present paper, which involves the *observation* of a baseline.

Climate Challenge and 1605(b), the two such programs that have been most subject to rigorous analysis, including careful attention to the self-selection issue, are both operated by DOE. Climate Challenge, begun in 1994 as a departmental initiative, had as a principal focus commitments by electric utilities to reduce, avoid, or sequester greenhouse gases by the year 2000. Utilities were invited to establish their own targets for emissions reductions, develop their own approaches for achieving the reductions, and self-report on their progress through 1605(b) and/or other means. As of this writing, Climate Challenge is no longer in operation. The 1605(b) program, strictly a reporting initiative, was open to a broader collection of industries, although electric utilities were disproportionately represented. As further described in the next section, 1605(b) began in 1994, as mandated by the Energy Policy Act of 1992. A revamped version remains in operation today. Evaluations of Climate Challenge have been conducted by Welch et al. (2000) and by Montes-Sancho et al. (2007). The sole prior evaluation of 1605(b) was conducted by Lyon and Kim (2006), who focused on the performance of electric utilities, as opposed to manufacturing firms, which are examined in the present paper.

Welch et al. (2000) report that on average, the 50 largest utilities in their sample reduced their CO_2 emissions by 6.3 million tons per firm over the sample period (1995–97), more than twice the amount initially pledged. At the same time, nonparticipating firms reduced their emissions by a larger amount. Thus, Welch et al. conclude that participation in the Climate Challenge program most likely had no impact on GHG emissions. In fact, some of their results suggest it may have had a detrimental impact on environmental performance. Nevertheless, DOE credits Climate Challenge with "… shift[ing] … the thinking of electric utility management and strategic planners to include the mitigation of greenhouse gas emissions into their corporate culture and philosophy."[3]

[2] For example, Sanchez et al. (2008) and Horowitz (2007, 2004, 2001) all use constructed baselines. For related analyses, see DeCanio (1998), DeCanio and Watkins (1998), Dowd et al. (2001), and Howarth et al. (2000), which also use constructed baselines.

[3] Climate Challenge Executive Summary, http://www.climatevision.gov/climate_challenge/execsumm/execsumm.htm.

The subsequent analysis by Montes-Sancho et al. (2007) distinguishes between early and late joiners. Like Welch et al. (2000), they find no overall difference in emissions reductions between program participants and nonparticipants. At the same time, they find that early joiners do reduce emissions significantly more than nonparticipants. Unfortunately, this favorable performance of the early joiners is cancelled out by the behavior of late joiners. The authors hypothesize that late joiners, who are smaller and subject to less political pressure than the initial participants, free-ride on the substantive efforts of the early joiners. As Lyon and Maxwell (2007) note, this suggests the participation of the late joiners "… can be viewed as a form of 'greenwash' designed to deflect attention from their actual environmental performance."[4]

Lyon and Kim (2006) examine the performance of electric utilities participating in DOE's 1605(b) program, using a Heckman-Hotz (1985) two-stage model to analyze both the decision to participate and the subsequent performance of the firm. The authors find that participants tend to be larger, with higher and more rapidly increasing emissions than nonparticipants. They also find that participation has no measurable effect on a firm's carbon intensity. Lyon and Kim conclude that participation in 1605(b) may be an attempt by firms to appear more environmentally friendly than is really the case—that is, to engage in greenwashing.

III. Background on Climate Wise and 1605b

The Climate Wise Program

Officially established by EPA in 1993, Climate Wise was a performance-based voluntary program focusing on the nonutility industrial sector to encourage the reduction of CO_2 and other GHGs without regard to specific technologies[5] via adoption of energy efficiency, renewable energy, and pollution prevention technologies. Climate Wise remained in operation until 2000, when it was renamed and placed under the agency's Energy Star umbrella. Subsequently, it was terminated altogether. At its peak, Climate Wise had enrolled more than 600 industrial firms covering several thousand facilities nationwide. EPA conducted internal studies, but there has been no outside evaluation of the program.

[4] Lyon and Maxwell (2007, 732).

[5] The stated goals of Climate Wise were to (1) encourage the immediate reduction of greenhouse gas emissions in the industrial sector through a comprehensive set of cost-effective actions; (2) change the way companies view and manage environmental performance by demonstrating the economic and productivity gains associated with 'lean and clean' manufacturing; (3) foster innovation by allowing participants to identify the actions that make the most sense for their organizations; and (4) develop productive and flexible partnerships within government and between government and industry. See Climate Wise 1998 Progress Report (U.S. EPA 1998, 2).

The program consisted of three interrelated components: (i) a pledge component asked firms to commit to taking cost-effective, voluntary actions to reduce GHG emissions; (ii) tailored assistance efforts were designed to facilitate companies' emissions-reducing efforts via a clearinghouse, workshops, and seminars; and (iii) communications activities provided public recognition for actual progress in reducing emissions.

To join Climate Wise, a firm had to develop a baseline estimate of its direct emissions of CO_2 (and other GHGs) for the year it joined the program or any year of its choice since 1990.[6] A more detailed emissions inventory was not required. The firm was also required to identify specific actions it proposed to undertake to reduce its emissions and, for each action, to indicate whether it was a "new," "expanded," or "accelerated" initiative. To encourage consideration of substantial reductions, EPA provided a checklist of major actions to improve equipment and processes, including those involving boiler efficiency, air compressor systems, and others. EPA also suggested fuel switching, best management practices, and the further integration of energy efficiency in new product design and manufacturing. Firms were strongly encouraged, albeit not required, to select at least some of their proposed actions from this list. EPA provided several types of technical assistance to participating firms, including a guide to industrial energy efficiency, background publications on energy efficiency and related issues, and most importantly, free phone consultation with energy experts retained by the agency. Information about financial assistance to support emissions-reducing actions was made available, including via Small Business Administration guaranteed loans and low-interest buy-downs from state providers, utility programs, and others. EPA also held an annual event open to the public to recognize the performance of outstanding Climate Wise participants.

The 1605(b) Program

Following the mandate of the Energy Policy Act of 1992, DOE's Energy Information Administration (EIA) issued formal guidelines for measuring and reporting energy and GHG reductions in a publicly available database, including provisions to ensure confidentiality of sensitive information.

Although 1605(b) involves fewer participation incentives than Climate Wise, the program does provide recognition for entities that reduce GHG emissions or sequester carbon voluntarily, and it identifies innovative and effective ways to reduce emissions. Most of the

[6] Although the Climate Wise program focused on energy efficiency and the reduction of CO_2 emissions, substantial reductions of the non-CO_2 gases were also reported, especially in the chemical and beer industries.

reporting entities are affiliated with one or more other voluntary programs sponsored by EPA or other government agencies.

Consistent with its legislative mandate, the 1605(b) program is extremely flexible. Participants can choose to report reductions at the firm or project level and can then define the reporting boundary relevant for either the firm or the project.[7] Since its inception in 1994, the number of reporting entities has doubled, from about 100 to more than 200 per year; the number of projects has more than tripled, from about 600 to more than 2,000 per year; and reported reductions in direct emissions have more than quadrupled, from 63 million metric tons in 1994 to 277 million metric tons in 2004, reflecting a 3.9 percent reduction from reported emissions in the base year. Although the electric power sector reported more entities, projects, and tons of emissions reduced than any other sector, the analysis presented in this study focuses strictly on the performance of manufacturing firms.[8]

IV. Data and Methods

For both Climate Wise and 1605(b), we combine participation data from the relevant government agencies with outcome data and control observations drawn from the LRD. As noted, we focus exclusively on the manufacturing sector.

Climate Wise Data

For Climate Wise, we obtained a list of enrolling firms in each year, from 1994 to 2000, including identifying information and whether they joined at the corporate level or as individual plants. Complete data were available for a total of 671 participants. As shown in Table 1, the

---

[8] In summarizing the benefits of 1605(b), EIA notes,

- “The program has served to teach staff at many of the largest corporations in the United States how to estimate greenhouse gas emissions and has educated them on a range of possible measures to limit emissions.
- “The program has helped to provide concrete evidence for the evaluation of activities reported to the many government voluntary programs launched since 1993.
- “Reporters have been able to learn about innovative emission reduction activities from the experiences of their peers.
- “The program has created a ‘test’ database of approaches to emission reductions that can be used to evaluate future policy instruments aimed at limiting emissions.
- “The program has helped to illuminate many of the poorly appreciated emissions accounting issues that must be addressed in designing any future approaches to emission limitations.”

See U.S. EIA 2002, 1–2.

number of corporate participants reached a peak in 1996, gradually declining to zero in 2000. The number of plant participants continued to increase until 2000.

The information on program participation was linked to detailed data from the LRD using name and, for plant participants, zip code information. We succeeded in linking 377 of the total 671 participants, including 228 corporate participants and 149 plant participants. The failure to link some participants to the Census data reflects the fact that the LRD includes only manufacturing establishments, whereas the Climate Wise program, despite its programmatic focus on manufacturing, also includes municipalities, universities, and other nonmanufacturing participants. These 377 linked participants from the original Climate Wise list translate into 2,311 facilities because many corporate participants have multiple associated facilities. The results of the linking are shown in the left panel of Table 2.

As displayed in the left panel of Table 3, the principal differences among participants and the broader universe of plants in the LRD is that Climate Wise participants are considerably larger. Our participant sample is also a very small fraction of the plants in the LRD—roughly 1 percent. This suggests that the full Census sample is unlikely to be an appropriate control group, and that a large number of plants are available from which to choose a more appropriate subgroup of controls.

Linking Climate Wise and Census data has important consequences for our ability to evaluate the effect of program participation over longer horizons. Because we are attempting to study behavior two or three years after joining, and since Census data are available only through 2001, we are forced to drop plants that enrolled in 1999 and 2000. Given the steep drop-off in new corporate participants after 1998, we do not sacrifice many observations by considering performance two to three years from the enrollment date. However, trying to discern effects four years after enrollment, with only those participants that joined between 1994 and 1997, we would have noticeably fewer observations and noisier estimates. Thus, we do not attempt to examine effects more than three years after enrollment.

1605(b) Data

For 1605(b) we obtained a list of reporting entities, sectors, years reported, and form type used, for the years between 1994 and 2001. As noted, we focus on manufacturing participants, which account for about 18 percent of reporting entities, as shown in Table 4.

Unlike the Climate Wise program, 1605(b) data do not include participants' enrollment dates. Thus, we use the first reporting year as the enrollment year and assume that the participants continued in the program after that, even though individual entities may not have continuous reporting years. The right panel of Table 1 displays the enrollment year information based on either firm or plant participation.

We also obtained a separate entity file from EIA that includes entity identification numbers and other information. The right panel of Table 4 shows the sector distribution for the 1605(b) program and LRD linking results. After excluding observations with missing data, the linking rate for the industrial sector is about 22 percent, or 83 participants from the original 1605(b) program list. These, in turn, correspond to 1,791 LRD facilities because corporate participants can have multiple facilities. The right panel of Table 2 summarizes the linking of the 1605(b) data to the LRD. The right panel of Table 3 provides summary statistics for the linked sample, as well as the entire LRD for the 1605(b) program.

Models and Econometric Methods

With the linked Census data described in the preceding section, we have access to variables indicating energy expenditures (separately for fuels and electricity), size (measured by the total value of shipments), location, and industry for a sample of manufacturing plants over a range of years from 1992 until 2000. We also have linked information on which plants participated in each of the two programs and the year in which they first participated. The challenge is to control for selection (i.e., the participation decision). Conceptually, we can imagine two outcomes, Y_i, for every observed plant, i: the value associated with participation, $Y_i(1)$, and the value associated with nonparticipation, $Y_i(0)$. Here, $Y_i(D_i)$ is the outcome associated with either treatment, $D_i = 1$, or nontreatment, $D_i = 0$, and is the cost of either fuels or electricity measured in natural logarithms. The ideal study would measure the treatment effect,

$$Y_i(1) - Y_i(0)$$

for each plant – that is, the percentage change in energy expenditures when a plant joins the program. The obvious problem is that for every plant, we observe either $Y_i(1)$ or $Y_i(0)$ but never both. The problem, viewed this way, is one of missing data and the fact that selection determines which data are observed and which are missing (i.e., who participates).

The simplest solution, and the one appropriate for randomized experiments, is to assume that the missing observations are *missing at random* (Rubin 1974). Under this assumption, the selection mechanism determining which outcomes are observed is *ignorable*. Formally, $D_i \perp Y_i(1), Y_i(0)$, we can measure the *average treatment effect* as

$$E\left[Y_i(1)-Y_i(0)\right]=\frac{\sum_{D_i=1}Y_i(1)}{\sum_{D_i=1}1}-\frac{\sum_{D_i=0}Y_i(0)}{\sum_{D_i=0}1}$$

that is, as the difference in average outcomes between participants and nonparticipants. Or we can estimate the treatment effect from a simple regression model

$$Y_i(D_i)=\beta_0+\beta_1 D_i+u_i$$

where we assume that u_i is uncorrelated with D_i and the treatment effect is the estimated value of β_1. Of course, in reality, observations are unlikely to be missing at random; in this regression model, we have to deal with the correlation of u_i with D_i.

One approach would be to build a structural model in which, even though selection, D_i, is dependent on an unobserved variable, we can still consistently estimate the treatment effect (Heckman and Hotz 1985). A major challenge in such an approach is to identify an excluded variable – something likely to influence participation but not the outcome. It is such a variable that ultimately allows consistent estimation, providing a source of variation in observed participation that is “random” from the standpoint of the outcome variable. Although such variables can sometimes be found or constructed, none were available for these programs.[9]

An alternative approach is propensity score matching, based on work by Rosenbaum and Rubin (1983) and more recently used by List et al. (2003) and Dehejia and Wahba (2002). When it is impossible to identify an excluded variable to create a “clean” source of variation in the participation variable, an alternative is to at least rule out outcome variation caused by *other* observable variables. Conditional on all those observed covariates, we could then assume that the participation decision is ignorable, or

$$D_i \perp Y_i(1),Y_i(0)\big| X_i$$

This condition could be met via a model such as

$$Y_i=\beta_0+\beta_1\cdot X_i+\beta_2\cdot D_i+u_i \quad (1)$$

except that it requires a correct specification of the X_i dependence, lest the estimated effect of the program remain mingled with covariates. Instead, Rosenbaum and Rubin (1993) and others

[9] Khanna and Damon (1999) use a list based on a letter that was sent to potential participants as an excluded variable. No such recruitment effort occurred for these programs. We did consider both distance to the nearest EPA regional office and local membership rates in a national environmental organization, based on suggestions from early reviewers; however, neither provided any variation in participation rate (the excluded variable has to be correlated with participation for this to work).

match participants to appropriate nonparticipants and consider the pairwise differences – essentially creating a situation where X_i and D_i are uncorrelated, so the regression is unbiased even without X_i. Thus, while Heckman-Hotz looks for clean variation in participation that is sufficiently random to allow one to ignore covariates but requires a strong assumption about an excluded variable, this approach tries to remove all sources of observed confounding variation but then requires an assumption that the remaining variation is random. When no excluded variable exists, the only alternative is the latter.

The general problem of creating a set of matched, nonparticipating observations is challenging because we would want to match the many observable variables – in our case, those describing location, industry, size, energy intensity, and growth. However, Rosenbaum and Rubin (1983) show that we need to match only the expected likelihood of participation. That is, we reduce the difficult problem of matching all of these different variables to a much simpler one of matching a summary variable describing the propensity to join the program. This greatly simplifies the creation of matched nonparticipant observations.

Our model of propensity score – the likelihood of joining the voluntary programs – depends on linear and quadratic terms involving value of shipments, cost of fuels, and cost of electricity (all in logarithms), as well as dummy variables for census region and two-digit industry classification. We also include the change in the logged value of shipments over the given time horizon as a control variable. Although this is arguably endogenous, we believe that controlling for growth is critical: we observed that faster-growing plants were more likely to join voluntary programs. It seems unlikely that this growth was *caused* by joining; therefore, we need to control for it. We use samples matched with different horizons to estimate program effects over similar horizons.

Because both Climate Wise and 1605(b) lasted a number of years, we decided to address the decision to join in a duration model framework. That is, in each period, conditional on not having joined, there is a given probability of joining based on the noted covariates and time. This allows us to combine data across years in estimating our participation model.[10] We therefore estimate a Cox proportional hazard model of the form:

---

[10] Note that, although the participants are associated with an enrollment year, the nonparticipants are not. In other words, plants may participate in various years, but this is not the case for nonparticipation. Outside of a duration model, it may not be possible to combine the data. In the Heckman-Hotz approach, we estimate effects for different cohorts of participants separately for this reason.

$$\begin{array}{c}\text{probability of joining in year } t\\ \text{(assuming plant } i \text{ has not yet joined)}\end{array} = h(t)\exp\begin{pmatrix}\beta_{size}\ln TVS_{i,t-1} + \beta_{elec}\ln EE_{i,t-1} + \beta_{fuels}\ln CF_{i,t-1}\\ +\left[\text{all quadratic combinations of size, elec, fuels}\right]\\ +\beta_{growth}\left(\ln TVS_{i,t+h} - \ln TVS_{i,t-1}\right)\\ +\sum_{\text{industries } j}\beta_j 1\left(M_i = j\right) + \sum_{\text{region } k}\beta_k 1\left(G_i = k\right)\end{pmatrix}$$

In this model, each period *t* has a baseline hazard rate *h*(*t*) defining the likelihood that a generic facility will join the program, conditional on not having joined previously. That baseline rate is then shifted by the various covariates. In the results section, below, we experiment with excluding various sets of covariates (industry dummies, region dummies, and quadratic terms).

Once estimated, we predict hazard rates – propensity scores – for each participating firm in the year it enrolls and match it to the nearest-valued nonparticipant (i.e., nonparticipant with the closest matching propensity score) in that year. We then examine the difference across each pair in the changes in fuel and electricity expenditures over one- to three-year horizons after joining, measured in natural logarithms; this difference-in-differences forms the estimate of program effectiveness.

V. Results

Before presenting our results using the propensity score approach, we give the results of using a simple regression model applied to the entire data sample for each of the two programs, and each of the two outcome measures (expenditure on fuels and electricity) (Table 5). The dependent variable measures the change between the year prior to enrollment by a group of participants (a cohort) and two years later; the results are broadly similar for one- and three-year horizons. The right-hand-side variables include all the variables used in the above propensity score matching model (value of shipments, cost of fuels, cost of electricity, growth in the value of shipments, and region and industry dummy variables) and the dummy variable indicating whether a firm joins the program in a given year. The coefficient on the participation variable, along with information on sample size and number of participants, is shown in the table.

Among the results for this simple model, we generally estimate small, statistically insignificant effects, with changes in energy expenditures of less than 10 percent. The three exceptions are a statistically significant 9 percent decline in electricity costs among 1605(b) participants in the 1994 cohort, a 6 percent increase in electricity costs among Climate Wise participants in the same cohort, and a *55 percent increase* in fuel costs among Climate Wise participants in the 1999 cohort. The first two effects are not inconsistent with our observations below, that electricity expenditures might increase among Climate Wise participants if efforts to reduce direct emissions lead to more electricity use and higher indirect emissions. In addition, a positive electricity effect could reflect a combination of specification error and the fact that larger or faster-growing firms are more likely to participate in voluntary programs. The third effect, a 55 percent increase in fuel expenditures among one cohort of Climate Wise participants, reflects an outlier in the rather small sample (96 participants) for that year that cannot be accommodated by our simple specification (the much larger estimation error indicates a much greater spread in outcomes, not a shift).

The preceding results ignore the potential for selection bias, which is the main focus of this exercise. Therefore, we now turn to the results from the propensity score matching approach shown in Tables 6 through 9. As noted in the methods section, we estimate a duration model for whether facilities enroll, using a variety of specifications. These specifications differ based on whether dummy variables are included for industry and region and whether quadratic terms are included, as indicated in the top three rows of each table. For each specification, we consider effects over one, two, and three years; we pool across all cohorts of matched pairs and report both the mean and the median differences in energy expenditures across pairs.

As with most of the estimates using the simple model in Table 5, all but one of the estimates suggest effects of less than 10 percent (the exception is 11 percent). We focus our discussion on the median estimates in the bottom half of each table because they are more robust to outlying observations of paired differences. Only 4 of these 72 median estimates are larger than 5 percent in magnitude, suggesting that any effect is probably even smaller than 10 percent. In general, the estimated changes in electricity expenditures are more likely to be statistically significant (6 of 36 estimates) than are the estimated changes in fuel expenditures (1 of 36). Interestingly, participation in 1605(b) seems to have a negative effect on electricity expenditures of perhaps several percentage points (Table 7, where 17 of 18 median estimates are negative), whereas Climate Wise appears, if anything, to have a slight positive effect on electricity expenditures (Table 9, where 14 of 18 median estimates are positive). The positive effect of Climate Wise is not present in our most general matching model (Table 9, where median estimates in column one are not significant); however, the negative effect of 1605(b) is present in this model (Table 7, where median estimates in column one are significant). Further, we found no evidence of persistence in the Climate Wise results: the effects at the three-year horizon are all lower than at two years (see bottom two rows of Table 9). Meanwhile, estimates of 1605(b) effects in four out of six models are greatest for the longest horizon (bottom row of Table 7).

Overall, we summarize these results as follows.

1. The effects of Climate Wise and 1605(b) on fuel and electricity expenditures are no more than 10 percent and probably less than 5 percent.
2. We found virtually no evidence of a statistically significant effect of either Climate Wise or 1605(b) on fuel costs.
3. Some statistically significant evidence suggests that participation in Climate Wise led to a slight (3–5 percent) increase in electricity costs that vanishes after two years.
4. Some statistically significant evidence suggests that participation in 1605(b) led to a slight (4–8 percent) decrease in electricity costs that persists for at least three years.

The transient, slight increase in electricity costs under Climate Wise is clearly unexpected. Two explanations seem plausible. First, participating plants may have pursued direct emissions reductions that required increased electricity use. Ignoring the indirect emissions associated with electricity use, this technically reduces emissions as defined by the program goals, but with the unintended consequence of higher indirect emissions from electricity use. Lower direct emissions might not show up in the cost-of-fuel measure because of fuel switching; for example, a shift to biomass or gas from coal might reduce emissions without changing expenditures. Alternatively, plants may have pursued nonenergy-related emissions reductions – such as the reduction of N_2O emissions at chemical plants – that are not reflected in a lower cost of fuels.

A positive effect on electricity expenditures may also reflect a failure to adequately control for growth. Although we matched participants with nonparticipants based, in part, on growth in the value of shipments, the tendency for faster-growing firms to enroll in both programs remains troubling because of its potential impact if we do not carefully control for this behavior. For example, we have no way of knowing about the underlying factors: participants might experience changes in quantities, whereas nonparticipants matched from the LRD might experience changes in prices. We cannot tease out controls that have that same pattern because details on prices and quantities are not available. If the estimated electricity expenditure growth effect really reflects an underlying and uncorrected difference in growth between participants and controls, then fixing it would presumably raise the growth rate of the control group and make the estimated program effect on electricity and fuel costs more negative.

VI. Conclusions

Thus far, the rigorous assessment of the environmental performance of energy- or GHG-related voluntary programs has been limited. The major challenge is to measure performance relative to an observed baseline. In the present study, we examine an EPA program and a DOE program, relying on confidential plant-level data for the manufacturing sector collected by the U.S. Census Bureau through 2001 to develop such a baseline based on a comparable set of nonparticipant controls. We do this via a propensity score matching approach, where we match participants to appropriate nonparticipants based on observable characteristics, and consider pairwise differences. As noted, the reductions are quite modest and, in at least one case, suggest an *increase* in electricity expenditures, although that effect vanishes after two years. At the same time, the findings of modest, albeit statistically significant, reductions in electricity expenditures for 1605(b) reporters may have implications for other government-sponsored voluntary programs. Recall that most of the entities reporting under 1605(b) are also affiliated with one or more other government-sponsored programs. Thus, the observed emissions reductions for 1605(b) reporters may reflect the influence of not only the 1605(b) program itself but also other programs. Although our separate assessment suggests that Climate Wise participation is probably not associated with significant emissions reductions, larger programs, including those that have more prescriptive participation criteria than the programs examined herein, such as EPA's Energy Star, may be more effective. Unfortunately, the EIA reporting form does not require disclosure of the names of other programs in which a firm participates.

Methodologically, our study highlights the inevitable complexity of assessing voluntary programs. Our study reinforces the work of others in emphasizing the importance of distinguishing between the participation decision and the environmental outcomes achieved. Our work also points to the value of working with micro-level data and the particular need to take special care in matching otherwise disparate samples to obtain a credible control group. This process is all the more difficult in our case, where the samples were not coded via a uniform system. In terms of estimation, we eschewed the more typical Heckman-Hotz method to selection bias because of the difficulty in finding excluded variables; instead we followed the propensity score matching approach of Rosenbaum and Rubin (1983). We believe that such an approach may have wider applicability in the future evaluation of voluntary programs.

Finally, we call the reader's attention to an intriguing observation by Lyon and Maxwell (2007) about the limits of an evaluation methodology, applied here and elsewhere, that compares the performance of participating firms with that of nonparticipants. Specifically, Lyon and Maxwell argue that if information on abatement diffuses to nonparticipants as well as to participants, then one would expect all firms to reduce their emissions at roughly the same rate, which appears to be the case in this study and the others cited earlier. Accordingly, they propose several research designs to estimate the importance of voluntary programs in diffusing information on potential efficiency gains, including potential differences in the rate of such diffusion among participants and nonparticipants, and the extent of diffusion outside traditional industry boundaries. While not definitive, these suggestions imply that further methodological development may be appropriate for the evaluation of voluntary programs, even beyond the approach adopted in the present study.

References

DeCanio, S.J. 1998. The Efficiency Paradox: Bureaucratic and Organizational Barriers to Profitable Energy-Saving Investments. *Energy Policy* 26: 441–54.

DeCanio, S.J., and W.E. Watkins. 1998. Investment in Energy Efficiency: Do the Characteristics of Firms Matter? *Review of Economics and Statistics* 80: 95–107.

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

Dowd, J., K. Friedman, and G.A. Boyd. 2001. How Well Do Voluntary Agreements and Programs Perform at Improving Industrial Energy Efficiency? Washington, DC: Office of Policy, U.S. Department of Energy.

Gamper-Rabindran, S. 2006. Did the EPA's Voluntary Industrial Toxics Program Reduce Plants' Emissions? A GSIS Analysis of Distrributional Impacts and a By-Media Analysis of Substitution. *Journal of Environmental Economics and Management* 52(1): 391–410.

Heckman, J.J., and V.J. Hotz. 1985. An Investigation of the Labour Market Earnings of Panamanian Males: Evaluating the Sources of Inequality. *Journal of Human Resources* 21(4): 507–42.

Horowitz, M. 2001. Economic indicators of Market Transformation: Energy Efficient Lighting and EPA's Green Lights. *Energy Journal* 22(4): 95–122.

———. 2004. Electricity Intensity in the Commercial Sector: Market and Public Program Effects. *Energy Journal* 25(2): 115–37.

———. 2007. Changes in Electricity Demand in the United States from the 1970s to 2003. *Energy Journal* 28(3).

Howarth, R., B. Haddad, and B. Paton. 2000. The Economics of Energy Efficiency: Insights from Voluntary Programs. *Energy Policy* 28: 477–86.

Khanna, M., and L.A. Damon. 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.

King, A.A., and M.J. Lenox. 2000. Does Membership Have Its Privileges? Analyzing Who Benefits from Industry Self-Regulation. Working Paper. Leonard N. Stern School of Business, New York University, New York. November.

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.P., and E.-H. Kim. 2006. Greenhouse Gas Reductions or Greenwash? The DOE's 1605b Program. Available at SSRN: http://ssrn.com/abstract=981730.

Lyon, T.P., and J.W. Maxwell. 2002. Voluntary Approaches to Environmental Regulation: A Survey. In *Economic Institutions and Environmental Policy*, edited by M. Franzini and A. Nicita. Hampshire, UK: Ashgate Publishing, 142–74.

———. 2007. Environmental Public Voluntary Programs Reconsidered. *Policy Studies Journal* 35(4): 723–50.

Montes-Sancho, M., M. Delmas, and M.V. Russo. 2007. Deregulation and Environmental Differentiation in the Electric Utility Industry. *Strategic Management Journal* 28(2): 189–209.

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

Rubin, D. 1974. Estimating Causal Effects of Treatments. *Journal of Educational Psychology* 66: 688–701.

Sanchez, M.C., R.E. Brown, C. Webber, and G.K. Homan. 2008. Savings Estimates for the United States Environmental Protection Agency's ENERGY STAR Voluntary Product Labeling Program. *Energy Policy* 36: 2098–108.

U.S. Energy Information Administration (EIA). 2002. *Voluntary Reporting of Greenhouse Gases 2000*. Washington, DC: U.S. EIA, Office of Integrated Analysis and Forecasting. February. www.eia.doe.gov/oiaf/1605/vrrpt/index.html (accessed May 15, 2008).

U.S. Environmental Protection Agency (EPA). 1998. *A Catalogue of the Agency's Partnership Programs*. EPA-100-B-97-003. Washington, DC: U.S. EPA, Office of the Administrator.

———. 2005. *Everyday Choices: Opportunities for Environmental Stewardship*. Technical Report by the EPA Environmental Stewardship Staff Committee for EPA Innovation Council. Washington, DC: U.S. EPA.

Welch, E., A. Mazur, and S. Bretschneider. 2000. Voluntary Behavior by Electric Utilities: Levels of Adoption and Contribution of the Climate Challenge Program to the Reduction of Carbon Dioxide. *Journal of Policy Analysis and Management* 19(3): 407–25.

Tables

Table 1. Enrollment Data for Climate Wise and DOE 1605(b) Participants

	Climate Wise			1605(b) Program		
Enrollment year	Plant	Corporate	Subtotal	Plant	Corporate	Subtotal
1994	0	8	8	0	105	105
1995	7	30	37	0	37	37
1996	38	141	179	3	23	26
1997	37	101	138	2	15	17
1998	36	70	106	8	53	61
1999	72	17	89	2	33	35
2000	144	0	144	6	53	59
Total	304	367	671	21	319	340

Table 2 . Matching of Climate Wise (CW) and 1605(b) to LRD

	CW List	LRD Plants	LRD plant–year observations (1992–2001)	1605(b)	LRD Plants	LRD plant–year observations (1992–2001)
Corporate participants with multiple plants	135	2,053	11,503	54	1,762	8,724
Corporate participants with a single plant	93	95	316	13	13	63
Plant-level participants	149	163	946	16	16	122
Total	377	2,311	12,765	83	1,791	8,909

Table 3. Sample Statistics for LRD and CW and 1605(b) Program Participants

		CW		1605(b)	
Variable	Summary statistics	Full LRD sample (1992–2001)	Program participants	Full LRD sample (1992–2001)	Program participants
ln(*TVS*) (total value of shipments)	Mean	7.61	10.87	7.80	10.99
	Standard deviation	2.30	1.81	2.34	2.17
	Plant–year observations	1,157,606	12,605	871,316	8,758
ln(*CF*) (cost of fuels)	Mean	2.54	5.31	2.69	5.36
	Standard deviation	2.12	2.23	2.18	2.32
	Plant–year observations	839,934	11,280	638,520	7,582
ln(*PE*) (purchased electricity)	Mean	3.17	6.31	3.34	6.22
	Standard deviation	2.21	1.83	2.25	2.17
	Plant–year observations	1,019,042	12,377	784,502	8,564
	Number of Plants	515,189	2,311	385,531	1,791

Table 4. The Sector Distribution for 1605(b) Reporting Entities and for Matched 1605(b) and LRD Data

Reporting counts
12
63
130
69
94
15
383

Note: Items marked with D* are included in "Other."

Table 5. Estimated Program Effects Using a Simple Regression Model

	1605(b)				Climate Wise			
Cohort	Fuel	Electricity	Sample	Participants	Fuel	Electricity	Sample	Participants
1994	–0.05	–0.09*	15319	343	0.06	0.06*	18788	809
	(0.05)	(0.03)			(0.03)	(0.02)		
1995	–0.06	0.06	26123	193	0.06	0.04	32768	335
	(0.08)	(0.06)			(0.06)	(0.04)		
1996	–0.06	–0.17	24089	28	0.04	0.02	29111	656
	(0.20)	(0.14)			(0.05)	(0.03)		
1997	–0.14	0.04	13754	192	–0.04	–0.02	16706	835
	(0.08)	(0.05)			(0.05)	(0.03)		
1998	–0.03	0.04	22536	164	–0.04	0.01	28658	1063
	(0.09)	(0.06)			(0.04)	(0.02)		
1999	0.09	0.05	18768	162	0.55*	0.05	18702	96
	(0.11)	(0.07)			(0.14)	(0.12)		

Table 6. Effect of the 1605(b) Program on the Natural Log of Fuel Expenditures over Different Horizons: Propensity Score Matching Approach

Matching model (all models include logged value of shipments, cost of fuels, cost of electricity, and growth in shipments)							Matched sample
Industry	x	x				x	
Region	x	x			x		
Quadratic	x			x	x	x	
Mean							
1-year effect	0.02	0.03	0.04	0.07	0.04	0.04	547
	(0.03)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
2-year effect	–0.06	–0.04	0.02	–0.03	–0.11	0.01	349
	(0.06)	(0.06)	(0.06)	(0.07)	(0.06)	(0.06)	
3-year effect	–0.08	–0.01	–0.07	0.00	–0.09	–0.05	298
	(0.07)	(0.06)	(0.07)	(0.07)	(0.07)	(0.07)	
Median							
1-year effect	0.02	–0.01	0.03	0.01	–0.03	0.02	547
	(0.03)	(0.03)	(0.03)	(0.02)	(0.03)	(0.03)	
2-year effect	0.03	0.01	0.03	0.04	–0.02	0.03	349
	(0.03)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	
3-year effect	–0.05	–0.01	–0.07	–0.02	–0.07*	–0.02	298
	(0.06)	(0.05)	(0.04)	(0.05)	(0.04)	(0.05)	

Note: Data are pooled across cohorts and the difference-in-difference is based on propensity score nearest neighbor matching.

Table 7. Effect of the 1605(b) Program on the Natural Log of Fuel Expenditures over Different Horizons

Model (all models include logged value of shipments, cost of fuels, cost of electricity, and growth in shipments)							Matched sample
Industry	x	x				x	
Region	x	x			x		
Quadratic	x			x	x	x	
Mean							
1-year effect	–0.04*	0.00	–0.01	–0.02	0.00	–0.04	581
	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	(0.02)	
2-year effect	–0.03	–0.03	–0.10*	0.00	–0.08	–0.05	388
	(0.04)	(0.04)	(0.04)	(0.05)	(0.04)	(0.04)	
3-year effect	–0.07	–0.11*	–0.04	–0.03	–0.01	0.05	336
	(0.04)	(0.05)	(0.05)	(0.05)	(0.05)	(0.06)	
Median							
1-year effect	–0.04*	–0.01	–0.03	–0.03	–0.02	–0.03*	581
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.01)	
2-year effect	–0.03	–0.04	–0.05	–0.01	–0.05	–0.03	388
	(0.02)	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	
3-year effect	–0.05*	–0.08*	–0.05	–0.04	–0.03	0.01	336
	(0.03)	(0.03)	(0.04)	(0.03)	(0.04)	(0.02)	

Note: Data are pooled across cohorts and the difference-in-difference is based on propensity score nearest neighbor matching.

Table 8. Effect of the EPA Climate Wise Program on the Natural Log of Fuel Expenditures over Different Horizons

Model (all models include logged value of shipments, cost of fuels, cost of electricity, and growth in shipments)							Matched sample
Industry	x	x				x	
Region	x	x			x		
Quadratic	x			x	x	x	
Mean							
1-year effect	–0.06	–0.03	–0.04	–0.05	–0.05	–0.02	949
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
2-year effect	0.04	0.02	0.00	0.02	–0.02	–0.04	830
	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	(0.04)	
3-year effect	–0.02	–0.06	–0.09	–0.07	–0.07	–0.10	764
	(0.04)	(0.05	(0.04)	(0.04)	(0.05)	(0.05)	
Median							
1-year effect	–0.01	0.00	–0.02	–0.02	–0.01	0.00	949
	(0.03)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	
2-year effect	0.03	0.03	0.02	0.01	–0.01	–0.03	830
	(0.03)	(0.02)	(0.03)	(0.03)	(0.03)	(0.03)	
3-year effect	–0.01	0.01	–0.10	0.01	–0.04	–0.04	764
	(0.03)	(0.03)	(0.03)	(0.03)	(0.04)	(0.04)	

Note: Data are pooled across cohorts and the difference-in-difference is based on propensity score nearest neighbor matching.

Table 9. Effect of EPA Climate Wise Program on the Natural Log of Fuel Expenditures over Different Horizons

Model (all models include logged value of shipments, cost of fuels, cost of electricity, and growth in shipments)							Matched sample
Industry	x	x				x	
Region	x	x			x		
Quadratic	x			x	x	x	
Mean							
1-year effect	0.05*	0.06*	0.06*	0.04	0.04	0.08*	1004
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
2-year effect	0.04	0.05*	0.05	0.03	0.05*	0.01	888
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)	(0.02)	
3-year effect	–0.01	0.01	0.00	0.03	0.02	–0.02	837
	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	(0.03)	
Median							
1-year effect	0.00	0.02	0.03	0.01	0.01	0.02	1004
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
2-year effect	0.02	0.05*	0.03	0.03*	0.01	0.02	888
	(0.02)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	
3-year effect	–0.01	–0.01	–0.01	0.02	0.00	–0.02	837
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	

Note: Data are pooled across cohorts and the difference-in-difference is based on propensity score nearest neighbor matching.