RUHR

ECONOMIC PAPERS

Manuel Frondel
Karoline Krätschell
Lina Zwick

Environmental Management Systems – Does Certification Pay?

#519

Imprint

Ruhr Economic Papers

Published by

Ruhr-Universität Bochum (RUB), Department of Economics
Universitätsstr. 150, 44801 Bochum, Germany

Technische Universität Dortmund, Department of Economic and Social Sciences
Vogelpothsweg 87, 44227 Dortmund, Germany

Universität Duisburg-Essen, Department of Economics
Universitätsstr. 12, 45117 Essen, Germany

Rheinisch-Westfälisches Institut für Wirtschaftsforschung (RWI)
Hohenzollernstr. 1-3, 45128 Essen, Germany

Editors

Editorial Office

Ruhr Economic Papers #519

Responsible Editor: Manuel Frondel

ISSN 1864-4872 (online) – ISBN 978-3-86788-594-2

Ruhr Economic Papers #519

Manuel Frondel, Karoline Krätschell and Lina Zwick

Environmental Management Systems – Does Certification Pay?

Bibliografische Informationen der Deutschen Nationalbibliothek

Die Deutsche Bibliothek verzeichnet diese Publikation in der deutschen Nationalbibliografie; detaillierte bibliografische Daten sind im Internet über: *http://dnb.d-nb.de* abrufbar.

http://dx.doi.org/10.4419/86788594
ISSN 1864-4872 (online)
ISBN 978-3-86788-594-2

Manuel Frondel, Karoline Krätschell and Lina Zwick[1]

Environmental Management Systems – Does Certification Pay?

Abstract

The voluntary adoption of environmental management systems (EMS), frequently certified by third-party audits following international standards, has become a vital supplement to mandatory environmental policies based on regulation and legislation. Although there is empirical evidence that both EMS adoption and certification can effectively improve firms' environmental performance, the impact on their business performance is far from clear. Drawing upon an OECD survey including more than 4,000 manufacturing facilities, this paper fills this void by estimating the impact of both EMS adoption and certification on facilities' business performance using statistical matching techniques. While our results indicate that the pure adoption of EMS without any certification does not enhance facilities' business performance, the financial performance of certified facilities turns out to be significantly higher.

JEL Classification: O33, O38, Q28

Keywords: Environmental regulation; matching methods

November 2014

1 All, RWI and Ruhr-Universität Bochum. *– We are grateful to Nicole Darnall, Irene Henriques, Jens Horbach, Klaus Rennings, Perry Sadorsky, as well as other OECD workshop participants for helpful discussions. In particular, we appreciate the valuable comments and suggestions by Christoph M. Schmidt and Colin Vance. This paper originates from the research project "Environmental Policy Tools and Firm-Level Management: A Cross OECD-Survey of Firms" (Project number 07RIW7, research initiative "Policy Frameworks for Sustainable Innovations"). Funding by the German Federal Ministry of Education and Research (BMBF) is gratefully acknowledged.* *– All correspondence to: Manuel Frondel, RWI, Hohenzollernstr. 1-3, 45128 Essen, Germany, e-mail: manuel.frondel@rwi-essen.de*

1. Introduction

Responding to the increasing environmental concerns raised by both customers and policy-makers,[1] the voluntary adoption of an environmental management system (EMS) by firms has become a vital supplement to mandatory environmental policies based on regulation and legislation (Frondel et al., 2008:154). Among other things, adopting an EMS typically entails the monitoring of a wide range of production processes, as well as the implementation of pollution-, energy- and waste-management systems. A step further is the voluntary certification of an EMS, as it requires third-party audits according to international norms, such as the European Union Environmental Management and Auditing Scheme (EMAS) and the standards of the International Standards Organization (ISO 14001).

By meeting such norms, companies may signal both high compliance with environmental regulation and that they take account of consumers' environmental awareness (Johnstone, Labonne, 2009:720). As one kind of communication of their environmental efforts, certification may serve as a strong signal to external observers and improve a company's image. In contrast, implementing an EMS without any certification may be largely internally motivated, e.g. by cost saving expectations (Johnstone, Labonne, 2009:720).

While there is empirical evidence that both EMS adoption and certification can effectively improve facilities' environmental performance (Arimura et al. 2008) and can also encourage environmental innovation (Horbach 2008), the impact on their business performance is far from clear. On the one hand, the costs incurred by facilities may be substantial, as the implementation and certification requires significant effort with respect to designing, implementing and documenting appropriate processes, entailing direct and indirect costs associated with consulting and audit fees, employee training, etc. On the other hand, EMS implementation can lead to cost savings due to the introduction of more efficient production processes, while EMS certification may in addition improve sales opportunities

[1] Many customers prefer purchasing environmentally benign products and abstain from buying from firms that violate environmental laws (Prakash 2002, Portney and Stavins 2000), while regulators are more and more concerned with the quality of firms' environmental performance.

due to signaling effects with respect to environmental compliance. The overall impact on profitability remains unclear, though, and is an empirical issue.[2]

There are a few studies that analyze this issue. For Japanese manufacturing industries, Nishitani (2011) demonstrates that ISO 14001 certification has positive business effects for export-oriented companies. Similarly positive results are obtained by Lo et al. (2012) for the U.S. fashion and textile industries. In contrast, Cañón-de-Francia and Garcés-Ayerbe (2009) find a negative relationship between ISO 14001 certification and a firms' market value, specifically for cleaner and less internationalized companies in Spain. Relying upon a twelve-page survey developed by the Organization for Economic Cooperation and Development (OECD) and a database of about 2,100 manufacturing facilities originating from Canada, Germany, Hungary, and the U.S., Darnall et al. (2008:373) provide for evidence that manufacturing facilities that implement more comprehensive environmental management systems (EMSs) are more likely to exhibit a positive business performance.

By making the clear distinction between three types of facilities – those that have not yet introduced an EMS, those that have, but without certifying it, and those that have certified their EMSs by third parties –, this article contributes to this line of inquiry by investigating whether EMS implementation and certification have distinct impacts on the business performance of facilities. To this end, we draw upon a more abundant database than that employed by Darnall et al. (2008) by adding data from three other countries that participated in the same OECD survey and use matching techniques, which have been rarely employed in the empirical literature on EMSs. In contrast to classical parametric regression methods that require assumptions on the relationship between the outcome measure and the treatment and control variables, matching is a non-parametric method of controlling for the confounding influence of covariates, rendering specification assumptions superfluous. A common drawback of both methods, however, is that both classical regression and matching techniques are crucially based on exogeneity assumptions, here, that the determinants of EMS adoption and certification are exogenous.

[2] More generally, numerous empirical studies indicate a positive relationship between the environmental and business performance of firms, see e.g. Klassen and McLaughlin (1996), Konar and Cohen (1997), Russo and Fouts (1997), and Darnall et al. (2007).

While providing for a didactic summary of matching methods, besides propensity-score matching techniques, we employ the covariate matching estimator, both of which are standard in the program evaluation literature. To correct for imperfect matching on multiple covariates, which can lead to substantial bias in limited samples (Abadie et al. 2004), we also use the bias-corrected covariate matching estimator developed by Abadie and Imbens (2011). This method is essentially based on a combination of matching and regression techniques.

Our results indicate that the pure implementation of an EMS without any certification does not enhance facilities' financial performance, a result that holds for both the whole sample and the subsample of EMS adopters. In other words, both the average treatment effect (ATE) and the average treatment effect of the treated (ATT) of EMS adoption are vanishing. For EMS certification, in contrast, we find both a statistically positive ATE and ATT, suggesting that the financial performance of companies that undergo ISO14001 or EMAS certification turns out to be significantly higher. This effect may result from better sales opportunities due to positive signaling effects associated with ISO14001 and EMAS certification.

The subsequent section describes the data and variables employed for our empirical estimations. Section 3 provides for a concise introduction to matching methods, followed by the presentation of our estimation results in Section 4. The last section summarizes and concludes.

2. Data and Variables

Our analysis is based on a survey on environmental policy tools and firm-level management practices that was initiated by the OECD and conducted in 2003 among seven OECD countries: Canada, France, Germany, Hungary, Japan, Norway, and the U.S. The database contains information on almost 4,200 facilities across all manufacturing sectors, including energy- and pollution-intensive sectors. A major task of the survey was to analyze the EMS adoption and certification decisions of facilities. Standardized questionnaires were used, encompassing questions pertaining to facility- and firm-specific characteristics, environmental behavior, the perception of the stringency of environmental regulation, etc.[3] The survey's focus was on facilities, rather than firms, as EMSs are established and certified at the facility level.

With respect to the focus of our analysis, we use the categorical information that respondents provided on a facility's overall business performance over the three years prior to 2003. On a five-point scale, the respondents indicated whether revenues (1) were so low as to produce large losses, (2) were insufficient to cover costs, (3) allowed breaking even, (4) were sufficient to make a small profit, or (5) were in excess of costs. On the basis of this information, we create a binary outcome variable that equals unity if a facility generated positive profits over the three years 2000 to 2002 and zero otherwise. (Using a categorical, rather than a binary variable is precluded when matching methods shall be employed.) As can be seen from Table 1, 59 % of the facilities report at least small, if not substantial profits for the period 2000-2002.

Among our two key explanatory variables is, first, a binary variable that indicates pure EMS adoption, yet without certification, until the end of 1999. While this holds true for 74 facilities, this definition presumes that the adoption of an EMS can only have a measurable impact on the business performance of the years 2000-2002 if it was adopted prior to this period. In a similar vein, our second key variable equals unity if a facility acquired either EMAS or ISO 14001 certifications until the end of 1999 and zero otherwise.[4] Until then, 367 facilities were certified, with most of these facilities acquiring certification within two years after EMS implementation.

Using the terminology of program evaluation, we denote EMS implementation and certification as treatment variables. In addition to these treatment variables, we control for a number of other facility-specific characteristics. To capture facility size, we employ three dummy variables reflecting the following size categories: less than 50 employees, between 50 and 249 employees, and more than 249 employees. Moreover, we include a series of dummy variables indicating the number of competitors a facility faces, another series of

[33] The questionnaire and descriptive statistics are available at: www.oecd.org/env/.

[4] As a consequence, 701 facilities that implemented an EMS or certified an EMS after 1999 are not considered in our matching exercise.

dummy variables measuring market scope, and, not least, whether the firm to which a facility belongs is listed on a stock exchange.

Table 1: Descriptive Statistics

Variable	**Mean**
PROFIT	0.590
PURE EMS	0.029
CERTIFICATION	0.130
STOCKEXCHANGE	0.167
STRINGENCY	0.168
Size Dummies:	
Less than 50 employees	0.064
Between 50 and 249 employees	0.611
More than 249 employees	0.330
#Competitors Dummies:	
Less than 5	0.271
Between 5 and 10	0.354
More than 10	0.375
Scope Dummies:	
LOCAL	0.085
REGIONAL	0.109
NATIONAL	0.416
GLOBAL	0.389

Furthermore, assuming that the environmental regulatory framework may affect facilities' business performance, we employ a measure for the perception of the degree of environmental policy stringency, which is defined to take on the value 1 if respondents consider the environmental policy regime as very stringent and equals zero otherwise. Finally, to account for country- and industry-specific differences, we include country- and sector dummies in the regressions (see Table A1 in the appendix for summary statistics).

All our variables are constructed from the answers provided by the survey respondents. This approach is far from being unproblematic, since for some variables, such as the degree of environmental policy stringency, these responses reflect both genuine variations across facilities as well as individual differences in the perception of the respondents. While in these instances the data is likely to be subject to measurement error, the majority of variables employed, such as the number of employees, provide for hard facts, rather than reflecting respondents' perceptions.

To gain a preliminary impression on the effect of both treatments, EMS implementation and certification, on facilities' business performance, we compare the shares of facilities with positive profits in the period 2000-2002 across treatment and control facilities, that is, those facilities without treatment. This share amounts to 63.0% for EMS adopters, compared to 57.4% for those facilities without EMS implementation. For those facilities with an EMAS or an ISO14001 certification the share of facilities with positive profits amounts to 62.5%.

These naïve comparisons clearly suffer from the fact that the facilities compared may have totally different characteristics. Therefore, it remains unclear whether and to what extent any difference in the above shares owes to the treatment or to differences in covariates, such as the number of competitors and market scope. Instead, reasonable comparisons are those in which treatment and control facilities are comparable in terms of their observable covariates, as well as, hopefully, with respect to their unobservable characteristics. To perform such comparisons conditional on the covariates is precisely the idea of the matching techniques presented in the subsequent section.

The principal challenge is to find appropriate matching partners, in our case treated and non-treated facilities that are as similar as possible in their pre-treatment characteristics and, ideally, differ only in their exposure to treatment, here EMS adoption and certification. To address this issue, alternate approaches are available, such as matching on the propensity-score or covariate matching, both of which are applied here.

3. Methodology

Following Abadie et al. (2004), this section provides for a didactic introduction to statistical matching. Essentially, propensity score matching is based on the estimation of the probability of taking part in a treatment – the propensity score – given the observed characteristics (Caliendo and Kopeinig, 2005). On the basis of these scores, facilities are matched to *m* potential matching partners, that is, untreated facilities with similar or even identical propensity scores. Rather than employing propensity scores, the method of covariate matching measures the distance $||\boldsymbol{z} - \boldsymbol{x}||_{\boldsymbol{V}}$ between any two vectors $\boldsymbol{x}$ and $\boldsymbol{z}$ of covariates, where $||\boldsymbol{x}||_{\boldsymbol{V}} := (\boldsymbol{x}^T \boldsymbol{V} \boldsymbol{x})^{1/2}$ is the vector norm with a positive definite matrix $\boldsymbol{V}$.[5]

Depending on the research question, one is interested in calculating the average treatment effect (ATE) or the average treatment effect on the treated (ATT):

$$ATE = \frac{1}{N}\sum_{i=1}^{N}(Y_i(1) - Y_i(0))$$

and

$$ATT = \frac{1}{N}\sum_{i=1}^{N}[Y_i(1) - Y_i(0)] \mid (W_i = 1)\,,$$

where $Y_i(1)$ and $Y_i(0)$ denote the (potential) outcomes of facility *i* when and when not exposed to the treatment, respectively, and *W* is the treatment indicator. Condition $W_i = 1$ indicates that for estimating the ATT, comparisons with matching partners are performed merely for the subsample of the treated facilities. In contrast, estimating the ATE additionally implies drawing comparisons of the outcome $Y_i(0)$ of any untreated facility with the average outcome of its matching partners from the group of treated facilities.

In practice, however, it is impossible to calculate the ATE or ATT on the basis of these formulae, as for each facility *i* merely one of either potential outcomes, $Y_i(1)$ or $Y_i(0)$, can be observed. This is called the fundamental evaluation problem (Frondel, Schmidt, 2005:518). To solve this problem, one has to provide for an estimate of the counterfactual outcome, that is, the unobserved of the two potential outcomes for each sample facility, such as the potential outcome $Y_i(0)$ of a treated facility *i*.

To this end, ideally, one would like to have one or several perfect matches for each facility *i*, that is, matching partners that are perfectly identical in the covariates. This event,

[5] Although there are no further a-priori restrictions on matrix $\boldsymbol{V}$, the Mahalanobis metric is frequently used, where $\boldsymbol{V}$ is the inverse of the sample covariance matrix of the covariates.

however, is very rare in non-experimental (observational) studies. Rather, differences in the covariates or propensity scores between facilities and their matches are hardly avoidable. In this instance, that is, when the matching is not exact, the covariate matching estimator will be biased in finite samples (Abadie et al. 2004:298).

To diminish this bias due to imperfect matching, Abadie and Imbens (2002) propose a combination of covariate matching and regression methods, which was implemented in Stata by Abadie et al. (2004). For this bias-corrected matching estimator, the potential outcomes would be predicted as follows (Abadie et al. 2004:299):

$$\hat{Y}_i(0) = \begin{cases} Y_i & if\ W_i = 0, \\ 1/|J_M(i)| \sum_{j \in J_M(i)} \left(Y_j + \hat{\mu}_0(\boldsymbol{x}_i) - \hat{\mu}_0(\boldsymbol{x}_j)\right) & if\ W_i = 1, \end{cases}$$

and

$$\hat{Y}_i(1) = \begin{cases} 1/|J_M(i)| \sum_{j \in J_M(i)} \left(Y_j + \hat{\mu}_1(\boldsymbol{x}_i) - \hat{\mu}_1(\boldsymbol{x}_j)\right) & if\ W_i = 0, \\ Y_i & if\ W_i = 1, \end{cases}$$

where $J_M(i)$ denotes the set of indices for the matches of facility *i*, $|J_M(i)|$ designates the number of matching partners of facility i and $\hat{Y}_i(1)$ and $\hat{Y}_i(0)$ are the estimates of the potential outcomes for facility *i* if it were and if it were not treated, respectively. If, for instance, facility *i* is not treated (W_i = 0), the first definition tell us that the counterfactual value $Y_i(1)$ is estimated by the average of the (bias-corrected) outcomes of the matching partners of facility *i* that enjoyed a treatment.

In both expressions, the difference $\hat{\mu}_w(\boldsymbol{x}_i) - \hat{\mu}_w(\boldsymbol{x}_j)$ represents the bias correction term that is introduced by Abadie and Imbens (2002) to cover the difference in the expected outcomes of facility *i* and its matching partners due to deviations in the covariates $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$. Only in the special case of perfect matching, that is, only if $\boldsymbol{x}_i = \boldsymbol{x}_j$, the bias correction term $\hat{\mu}_w(\boldsymbol{x}_i) - \hat{\mu}_w(\boldsymbol{x}_j)$ vanishes for both *w* = 0 and *w* = 1, since in this case there is no bias due to imperfect matching. The special case of standard matching estimation precisely employs the resulting estimates $\hat{Y}_i(1)$ and $\hat{Y}_i(0)$ with vanishing bias-correction terms to provide for estimates of the ATE and ATT:

$$\widehat{ATE} := 1/N \sum_{i=1}^{N} \{\hat{Y}_i(1) - \hat{Y}_i(0)\}$$

and

$$\widehat{ATT} := 1/N \sum_{i=1}^{N} \{Y_i - \hat{Y}_i(0) |\, W_i = 1)\} .$$

To at least partially correct for the consequences resulting from imperfect matching, that is, from differences between $\boldsymbol{x}_i$ and $\boldsymbol{x}_j$, for the conditional expectations $\mu_w(\boldsymbol{x}) := E[Y(w) |\, X = \boldsymbol{x}]$ for $w = 0$ and $w = 1$, Abadie and Imbens (2002) suggest approximating these conditional expectations by linear (regression) functions and estimating them using least squares on the matched observations:

$$\hat{\mu}_w(\boldsymbol{x}) = \hat{\alpha}_w + \hat{\beta}'_w \boldsymbol{x}.$$

That is, on the basis of the matched, rather than the full sample, two weighted least squares regressions are performed to obtain estimates of the correction terms $\hat{\mu}_w(\boldsymbol{x}_i) - \hat{\mu}_w(\boldsymbol{x}_j)$. The first regression uses the observations for only the treated of all the matched facilities, whereas the second regression employs only those of the untreated of all the matched facilities:

$$(\hat{\alpha}_w, \hat{\beta}_w) = argmin_{(\hat{\alpha}_w, \hat{\beta}_w)} \sum_{\boldsymbol{i}:\, \boldsymbol{W_i = w}} K_M(i)\,(Y_i - \hat{\alpha}_w - \hat{\beta}'_w \boldsymbol{x_i})^2,$$

where observations are weighted by K_M *(i)*, the number of times a facility is used as a match. This weighting is reasonable, because the weighted empirical distribution is closer to the distribution of covariates in which one is ultimately interested (Abadie et al. 2004:299).

Finally, it bears noting that the matching estimator consistently estimates the treatment effect of interest only if selection into treatment is purely random. This assumption, also known as unconfoundedness or “selection on the observables”, is formally stated as follows: Treatment status *W* is independent of the potential outcomes $Y_i(0)$ and $Y_i(1)$ conditional on $\boldsymbol{x}$. This is a strong assumption and, in fact, may not hold in our empirical example, as there might be unobservable factors, such as the quality of the management, that determine both the assignment to the treatment (EMS adoption and certification) and the business performance of the facility and, hence, selection into treatment might not be purely random.

A further requirement refers to a sufficient overlap in the covariate distributions between treated and control groups:

$c < Pr(W = 1 | [X = x]) < 1 - c$ for some $c > 0$.

The interpretation of this overlap assumption is straightforward: If, for instance, all the facilities with a given covariate pattern select into the treatment, there would be no matching partners for comparisons, that is, similar facilities without treatment. As a consequence, statistical matching would not be applicable.

4. Results

Throughout, our matching results are based on nearest-neighbor algorithms, with the covariate matching being based on the Mahalanobis weighting matrix and the propensity-score matching relying on the propensity scores estimated by a probit model.[6] To check the robustness of our results beyond the application of several estimation methods, we vary the number *m* of matching partners, since the final inference can critically depend on this choice (Abadie et al. 2004:298).

We find that the estimates do not change dramatically with *m*. As can also be observed from the result tables presented in this section, standard errors usually decrease monotonically with increasing *m*, an outcome that can be expected due to the increasing information provided by the inclusion of more matching partners. Continually increasing the number of matching partners is not to be recommended, though, as more and more observations are incorporated that are not sufficiently similar. In contrast, choosing a low number *m* has the disadvantage of including too little information. Abadie et al. (2004) therefore recommend using four matching partners. We thus have ignored the results for *m* = 1 entirely and have employed two to six matching partners.

Focusing first on the estimates of the average treatment effect on the treated (ATT), our interest is on the pure impact of EMS implementation on the business performance of the subsample of those EMS adopters that have not been certified yet. Our results indicate that pure EMS implementation does not have any statistically significant effect (see Table 2).

[6] The Stata commands „psmatch2“ and „teffects psmatch“ are applied for propensity-score matching and „psmatch2“ and „teffects nnmatch“ for covariate matching, respectively.

These results turn out to be quite robust across all matching techniques employed, including both the very similar, and thus not reported, covariate- and propensity-score matching estimates obtained from the Stata Command psmatch2 and the bias-corrected covariate matching estimator, whose estimates are also not reported in Table 2.

The latter outcome is not surprising given that the balance checks presented in Table A2 and A3 of the appendix indicate no need for a bias correction.[7] In fact, Tables A2 and A3 reveal that the matching approach is successful in ascertaining covariate balancing for both the propensity-score and covariate matching.[8] That is, the matching approaches lead to subsamples of treated and control facilities that are balanced with respect to the observables, whereas there are substantial differences in the covariates for the unmatched subsamples. These results ensure that we are comparing facilities with similar characteristics in our matching analysis (Mensah et al. 2010).

Table 2: Pure EMS adoption: Average Treatment Effects on the Treated (ATT)

	Number of matching partners				
Matching Method (Stata Command)	m=2	m=3	m=4	m=5	m=6
Covariate Matching (teffects nnmatch)	-0.118 (0.072)	-0.114 (0.070)	-0.123 (0.067)	-0.099 (0.064)	-0.092 (0.062)
Propensity-Score Matching (teffects psmatch)	-0.108 (0.067)	-0.113 (0.063)	-0.109 (0.063)	-0.117 (0.063)	-0.118 (0.061)
Number of observations:	2,196				

Note: Robust standard errors are included in parentheses. ** and * denote statistical significance at the 1 and 5 percent levels, respectively.

For EMS certification, though, Tables A4 and A5 show that neither covariate-, nor propensity-score matching lead to a balancing in all covariates. In fact, in case of covariate matching, four variables – two facility-size dummies, being listed at a stock exchange, and the stringency of environmental regulation – are unbalanced. Upon propensity-score matching, some country dummies remain unbalanced. These differences in the covariate balancing indicate that propensity-score and covariate matching are based on totally

[7] For these balance checks, t tests are employed to examine whether there are substantial differences in the covariates before and after matching. While such differences are to be expected for the unmatched sample, these differences should vanish after the matching.

[8] We had to drop Hungary and Norway, as well as the furniture and recycling sectors, from the dataset, since there are no "treated" facilities, that is, EMS adopters in these countries and sectors.

different matching algorithms that may lead to very different average treatment effects (see Table 3).

Possible biases, caused by both unbalanced observables and by (small) remaining differences[9] in other covariates, can be adjusted by the correction method proposed by Abadie and Imbens (2011) when covariate matching is employed. Correcting for differences in covariates using this method may therefore yield more reliable estimates than propensity-score matching and will thus be the focus of our analysis. According to the ATT outcomes resulting from the bias-corrected matching estimator, EMS certification seems to have a positive impact on facilities' business performance for the subsample of certified companies (Table 3).

Table 3: EMS certification: Average Treatment Effects on the Treated (ATT)

	Number of matching partners				
Matching Method (Stata Command)	m=2	m=3	m=4	m=5	m=6
Covariate Matching (teffects nnmatch)	0.065 (0.043)	0.077* (0.036)	0.092** (0.033)	0.095** (0.032)	0.102** (0.032)
Covariate Matching with bias correction (teffects nnmatch)	0.074 (0.043)	0.084* (0.036)	0.093** (0.033)	0.110** (0.032)	0.112** (0.032)
Propensity Score Matching (teffects psmatch)	0.021 (0.040)	0.009 (0.040)	0.010 (0.037)	0.018 (0.036)	0.026 (0.038)
Number of observations:	2,463				

Note: Robust standard errors are included in parentheses. ** and * denote statistical significance at the 1 and 5 percent levels, respectively.

We now turn to the estimates of the sample average treatment effects (ATE) and, hence, to the question of whether the results for the subsample of treated facilities can be generalized to the whole sample. Table 4 indicates that this is the case for pure EMS implementation without any certification. Except for propensity-score matching with only two matching partners, the estimated ATE are not statistically significant, being in perfect accord with Table 2.

The ATT results for EMS certification can also be generalized to the whole sample (Table 5). Similar to the estimates reported in Table 3, we find a statistically significant

[9] Although t tests are a common method to assess matching quality, they fail to indicate by how much the differences in the covariates, and hence the bias, are reduced (Caliendo, Kopeinig 2005).

positive average treatment effect (ATE) when covariate matching is employed, suggesting the interpretation that certification seems to pay. As with the ATT results presented in Table 3, though, these effects vanish when propensity-score matching is the method of choice.

Table 4: Pure EMS adoption: Average Treatment Effects (ATE)

	Number of matching partners				
Matching Method (Stata Command)	m=2	m=3	m=4	m=5	m=6
Covariate Matching (teffects nnmatch)	0.042 (0.095)	0.045 (0.086)	0.040 (0.083)	0.045 (0.082)	0.049 (0.082)
Propensity-Score Matching (teffects psmatch)	-0.208** (0.066)	-0.182 (0.094)	-0.089 (0.086)	-0.022 (0.077)	0.004 (0.063)
Number of observations:	2,196				

Note: Robust standard errors are included in parentheses. ** and * denote statistical significance at the 1 and 5 percent levels, respectively.

In sum, while pure EMS implementation without certification does not seem to pay, certification of EMSs appears to favor the business performance of certified facilities, at least according to our covariate matching results.[10] This outcome, however, might be explained by unobservable factors, such as high management quality, that may favor the decision to implement and certify an EMS ("selection on unobservables"), but also improve a facility's business performance.

Due to such unobservable factors, selection into treatment might be far from being purely random and, hence, our estimation results might be biased. Such endogeneity problems cannot be solved by using matching approaches. In other words, as well as classical regression methods, such as discrete-choice models, matching approaches are inconsistent when the data generation process is undermined by self-selection mechanisms. In our case, indeed, self-selection may plausibly explain our empirical results: It might well be the case that the managers of the best facilities know that signaling effects matter and that an EMS only pays if it is also certified. In contrast, the managers of less successful facilities with severe management or cost problems may not be aware of this opportunity or may not be able to accomplish certification.

[10] We have also estimated the (additional) effect of certification by employing just those facilities as controls that have implemented an EMS, but have not acquired any certification. However, these results appear to be highly unreliable, because matching does not provide for balanced observables, most likely due to the low number of 74 controls.

Table 5: EMS certification: Average Treatment Effects (ATE)

Matching Method (Stata Command)	Number of matching partners m=2	m=3	m=4	m=5	m=6
Covariate Matching (teffects nnmatch)	0.082* (0.032)	0.078* (0.032)	0.076* (0.036)	0.057 (0.036)	0.053 (0.035)
Covariate Matching Bias correction (teffects nnmatch)	0.078* (0.032)	0.081* (0.032)	0.079* (0.036)	0.070 (0.037)	0.074* (0.035)
Propensity-Score Matching (teffects psmatch)	0.024 (0.042)	0.023 (0.044)	0.040 (0.043)	0.051 (0.041)	0.059 (0.042)
Number of observations:	2,463				

Note: Robust standard errors are included in parentheses. ** and * denote statistical significance at the 1 and 5 percent levels, respectively.

5. Summary and Conclusion

Together with ISO 9001, ISO 14001 is among the best known international management standards and the dominant environmental management system (EMS) in the world. Both systems have been implemented by more than a million organizations in 175 countries and, according to the ANSI-ASQ National Accreditation Board, far more organizations could benefit from implementing a certified management system, as implementation can provide a competitive marketing and sales edge (ANAB, 2013). Empirical evidence for this claim, however, is sparse.

Drawing upon an OECD survey on environmental policy tools and facility-level management practices and using propensity-score- as well as covariate-matching methods, this paper has investigated the value of accredited certification by addressing the question of whether either pure EMS implementation or certification favor facilities' business performance. As collapsing the two decisions on EMS implementation and certification can be misleading in any analysis (King et al. 2006), we have made a clear distinction between both. Making this distinction may be interpreted as differentiating between cost and signaling effects: While a positive impact of EMS implementation on a facility's business performance may be an indication of cost savings owing to the introduction of more efficient production processes, a positive effect of EMS certification may result from better sales opportunities due to signaling effects.

Our results indicate that pure EMS implementation without certification does not affect business performance, suggesting the absence of cost effects. In contrast, ISO 14001 and EMAS certification seem to have a positive impact on the financial performance of facilities, suggesting that companies can benefit from signaling effects associated with the certification of EMSs. This is in line with the observed behavior of facilities to certify EMSs shortly after their adoption and the view of the ANSI-ASQ National Accreditation Board. Given potential endogeneity biases that cannot be healed by matching methods, however, further research is needed to actually be able to substantiate the claim by the ANSI-ASQ National Accreditation Board that ISO 14001 and other management system standards can provide a solid foundation on which to build an organization, one that can withstand the test of time and challenges of the marketplace.

Appendix

Table A1: Additional Descriptive Statistics

Variable	Mean	Description
Sector dummies (ISIC codes)		
Food (15-16)	0.101	Food products, beverages and tobacco products
Textiles (17-19)	0.049	Textiles, leather and footwear
Wood and paper (20-22)	0.105	Wood and paper products, publishing and printing
Chemicals (23-25)	0.153	Fuel, chemicals, rubber and plastic
Minerals (26)	0.036	Non-metallic mineral products
Metals (27-28)	0.200	Basic metals and fabricated metal products
Machinery (29-33)	0.240	Machinery, electrical and optical equipment
Transport (34-35)	0.070	Transport equipment
Furniture (36)	0.026	Furniture
Recycling (37)	0.006	Recycling
Country dummies		
Canada	0.061	
France	0.064	
Germany	0.215	
Hungary	0.111	
Japan	0.358	
Norway	0.074	
USA	0.117	

Table A2: Covariate Balance Check for EMS adoption using Individual t Tests (Propensity-Score matching)

		Matched				
Variable	**Unmatched**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
Size						
Between 50 and 249	-5.02**	0.60	0.23	-0.09	0.14	0.26
More than 249	6.08**	-0.59	-0.28	-0.08	-0.34	-0.39
Stock exchange	8.08**	0.17	0.11	0.17	0.07	0.03
Stringency	5.58**	0.87	0.93	0.65	0.80	0.76
#Competitors						
Less than 5	-0.08	0.19	0.13	0.24	0.27	0.36
Between 5 and 10	0.73	-0.77	-0.74	-0-89	-0.89	-0.80
Scope						
Local	-0.89	-1.12	-0.87	-0.60	-0.49	-0.31
Regional	-0.58	-0.14	0.09	0.07	0.23	30.05
National	-1.23	1.01	0.54	0.04	-0.07	0.00
Country Dummies						
Canada	6.04**	-0.37	-0.31	-0.19	-0.19	-0.31
France	-0.88	1.01	1-23	1.18	1.14	1.12
Germany	-1.73	0.11	0.07	-0.06	-0.09	-0.15
Japan	-5.07**	-0.33	-0.22	-0.17	-0.13	-0.11
Sector dummies						
Chemicals	3.66**	0.00	-0.25	-0.28	-0.04	-0.03
Food	-1.67	0.00	0.00	-0.09	-0.034	-0.45
Machinery	-1.58	-0.89	-0.89	-0.68	-0.64	-0.57
Metals	0.95	-0.29	0.00	0.00	0.04	0.16
Textiles	0.01	1.16	1.01	1.05	1.16	0.93
Transport	-0.84	1.42	1.10	1.17	0.87	0.82
Wood and paper	0.60	0.37	0.41	0.31	0.15	0.08

Note: **and * denote significance at the 1 and 5 percent levels, respectively. Reference categories are: "Size: Less than 50 employees", "#COMPETITORs: more than 10", "Scope: global", "Country: USA", "Sector: Minerals". Covariate balance check was performed by using the Stata commands psmatch2 and pstest. We dropped Hungary, Norway, furniture and recycling from the dataset, since there are no "treated" facilities in these countries and sectors.

Table A3: Covariate Balance Check for EMS adoption using Individual t Tests (Covariate Matching)

		Matched				
Variable	**Unmatched**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
Size						
Between 50 and 249	-5.02**	-1.27	-1.18	-1.27	-1.32	-1.44
More than 249	6.08**	1.18	1.07	1.14	1.18	1.27
Stock exchange	8.08**	0.43	0.51	0.51	0.55	0.63
Stringency	5.58**	1.05	1.42	1.42	1.42	1.61
#Competitors						
Less than 5	-0.08	0.39	0.66	0.75	0.55	0.39
Between 5 and 10	0.73	-0.26	-0.51	-0.43	-0.21	-0.06
Scope						
Local	-0.89	0.34	0.34	0.43	0.42	0.46
Regional	-0.58	-0.27	-0.18	-0.07	-0.11	-0.18
National	-1.23	0.00	0.24	0.13	0.14	0.12
Country Dummies						
Canada	6.04**	0.00	0.00	0.09	0.15	0.19
France	-0.88	0.00	0.00	0.11	0.17	0.22
Germany	-1.73	0.00	0.07	0.00	-0.04	0.00
Japan	-5.07**	-0.32	-0.41	-0.45	-0.58	-0.50
Sector dummies						
Chemicals	3.66**	-0.09	-0.06	-0.05	-0.08	0.00
Food	-1.67	0.00	0.00	0.00	0.00	0.00
Machinery	-1.58	0.00	-0.08	-0.18	-0.33	-0.54
Metals	0.95	0.10	0.07	0.10	0.20	0.20
Textiles	0.01	0.00	0.11	0.17	0.20	0.22
Transport	-0.84	0.00	0.00	0.00	0.00	0.00
Wood and paper	0.60	0.00	0.00	0.00	0.00	0.12

Note: **and * denote significance at the 1 and 5 percent levels, respectively. Reference categories are: "Size: Less than 50 employees", "#COMPETITORs: more than 10", "Scope: global", "Country: USA", "Sector: Minerals". Covariate balance check was derived by using the Stata commands psmatch2 and pstest. We dropped Hungary, Norway, furniture and recycling from the dataset, since there are no "treated" facilities in these countries and sectors.

Table A4: Covariate Balance Check for EMS certification using Individual t Tests (Propensity-Score Matching)

		Matched				
Variable	**Unmatched**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
Size						
Between 50 and 249	-12.17**	0.85	0.73	0.89	0.97	0.92
More than 249	15.46**	-0.88	-0.80	-0.94	-0.94	-0.76
Stock exchange	9.32**	-1.12	-1.16	-1.24	-1.29	-1.59
Stringency	1.18	-0.49	-0.03	0.51	0.91	0.81
#Competitors						
Less than 5	0.79	-0.47	-1.12	-1.23	-1.52	-1.47
Between 5 and 10	0.03	-0.52	-0.38	0.45	0.64	0.49
Scope						
Local	-4.56**	-0.45	-0.45	-0.45	-0.54	-0.56
Regional	-0.86	-0.72	-0.08	0.09	-0.05	-0.04
National	-2.83**	1.20	1.42	1.22	1.26	1.40
Country Dummies						
Canada	-1.96*	0.49	0.52	0.39	0.04	-0.45
France	-0.93	0.61	0.11	0.21	0.00	0.00
Germany	-0.20	-1.99*	-1.76	-1.49	-1.15	-1.33
Hungary	-3.71**	0.16	-0.10	-0.12	0.00	0.21
Japan	6.09**	3.38**	3.67**	3.30**	3.02**	3.20**
Norway	-0.43	-2.33*	-2.50*	-2.39*	-2.23*	-2.32*
Sector dummies						
Chemicals	4.09**	-1.95*	-1.50	-1.09	-0.80	-0.79
Food	-2.61**	-0.97	-0.84	-0.80	-0.92	-0.86
Furniture	-2.26*	-0.20	-0.13	0.00	0.00	-0.07
Machinery	6.53**	0.92	0.48	1.08	1.10	1.20
Metals	-1.06	0.89	1.02	0.57	0.48	0.29
Recycling	0.85	0.71	-1.01	-0.55	-0.38	-0.20
Textiles	-3.29**	-0.12	-0.17	0.06	0.21	0.31
Transport	1.55	-1.22	1.6110	0.11	-0.35	-0.48
Wood and paper	-4.44**	0.20	0.27	0.05	0.16	0.07

Note: **and * denote significance at the 1 and 5 percent levels, respectively. Reference categories are: “Size: Less than 50 employees“, #COMPETITORs: more than 10”, “Scope: global”, “Country: USA”, “Sector: Minerals”. Covariate balance check was performed by using the Stata commands psmatch2 and pstest.

Table A5: Covariate Balance Check for EMS certification using Individual t Tests (Covariate Matching)

		Matched				
Variable	**Unmatched**	**m=2**	**m=3**	**m=4**	**m=5**	**m=6**
Size						
Between 50 and 249	-12.17**	-3.17**	-4.29**	-5.07**	-5.63**	-6.01**
More than 249	15.46**	3.12**	4.20**	5.00**	5.59**	5.99**
Stock exchange	9.32**	1.88	2.48*	2.62**	3.08**	3.53**
Stringency	1.18	1.59	1.90	1.97	2.22*	2.33*
#Competitors						
Less than 5	0.79	0.04	0.09	0.15	0.19	0.19
Between 5 and 10	0.03	-0.84	-1.01	-1.14	-1.17	-1.01
Scope						
Local	-4.56**	0.12	0.08	0.06	0.00	0.00
Regional	-0.86	0.32	0.51	0.61	0.80	1.05
National	-2.83**	-0.24	-0.24	-0.24	-0.45	-0.65
Country Dummies						
Canada	-1.96	0.00	0.00	0.00	0.00	0.03
France	-0.93	0.00	0.06	0.04	0.10	0.20
Germany	-0.20	-1.12	-1.26	-1.41	-1.42	-1.30
Hungary	-3.71**	-0.38	-0.20	-0.08	-0.06	-0.05
Japan	6.09**	1.24	1.22	1.24	1.18	0.98
Norway	-0.43	0.00	0.00	0.04	0.06	0.10
Sector dummies						
Chemicals	4.09**	0.10	0.06	0.05	0.13	0.19
Food	-2.61**	0.00	-0.05	-0.04	0.03	0.10
Furniture	-2.26*	0.00	0.00	0.00	0.00	0.00
Machinery	6.53**	0.04	0.16	0.10	0.02	-0.09
Metals	-1.06	-0.20	-0.27	-0.18	-0.20	-0.19
Recycling	0.85	0.00	0.00	0.00	0.00	0.00
Textiles	-3.29**	0.00	0.00	0.00	0.00	0.00
Transport	1.55	0.00	0.00	0.04	0.03	0.02
Wood and paper	-4.44**	0.10	0.07	0.00	0.00	0.03

Note: **and * denote significance at the 1 and 5 percent levels, respectively. Reference categories are: "Size: Less than 50 employees", "#COMPETITORs: more than 10", "Scope: global", "Country: USA", "Sector: Minerals". Covariate balance check was derived by using the Stata commands psmatch2 and pstest.

References

Abadie, A., Drucker, D., Herr, J.L. and G.W: Imbens (2004) Implementing matching estimators for average treatment effects in Stata. *The Stata Journal* 4(3), 290-311.

Abadie, A., Imbens, G.W. (2011). Bias-Corrected Matching Estimators for Average Treatment Effects. *Journal of Business & Economic Statistics* 29(1), 1-11.

ANAB (2013) The Value of Accredited Certification. ANSI-ASQ National Accreditation Board. http://www.anab.org/resources/value.aspx.

Arimura, T.H., Hibiki, A., Katayama, H. (2008) Is a Voluntary Approach an Effective Environmental Policy Instrument? A Case for Environmental Management Systems. *Journal of Environmental Economics and Management* 55, 281-295.

Caliendo, M. and S. Kopeinig (2005), Some Practical Guidance for the Implementation of Propensity Score Matching, IZA Discussion Paper No. 1588.

Cañón-de-Francia, J. And C. Garcés-Ayerbe (2009) ISO 14001 Environmental Certification: A Sign Valued by the Market? *Environmental and Resource Economics* 44(2), 245-62.

Darnall, l. N., Jolley G.J and B. Ytterhus (2007) Understanding the relationship between a facility's environmental and financial performance, in: Johnstone N. (ed.) Environmental policy and corporate behaviour. Edward Elgar, Cheltenham.

Darnall, N., Henriques , I.and P. Sadorsky (2008) Do environmental management systems improve business performance in the international setting? *Journal of International Management* 14(4), 364-376.

Frondel, M., Horbach, J. and K. Rennings (2008) What triggers environmental management and innovation? Empirical evidence for Germany. *Ecological Economics* 66(1), 153-160.

Frondel, M. and C.M. Schmidt (2005) Evaluating environmental programs: The perspective of modern evaluation research. *Ecological Economics* 55 (4), 515-526.

Horbach, Jens (2008): Determinants of Environmental Innovation – New Evidence from German Panel Data Sources, Research Policy 37, 163-173

Johnstone, N. and J. Labonne (2009) Why do manufacturing facilities introduce environmental management systems? Improving and/or signaling performance. *Ecological Economics* 68(3), 719-730.

King, A., Lenox, M., Terlaak, A. (2006) The Strategic Use of Decentralized Institutions: Exploring Certification with the ISO 14001 Management Standard. *Academy of Management Journal*, 48(6), 1091-1106.

Khanna, M. and W.R.Q: Anton (2002) Corporate environmental management: regulatory and market-based pressures. *Land Economics* 78(4), 539-558.

Klassen, R.D and C.P. McLaughlin (1996). The Impact of Environmental Management on Firm Performance. *Management Science* 42(8), 1199-214.

Konar, S. and M.A. Cohen (1997) Information as Regulation: The Effect of Community Right to know Laws on Toxic Emissions. *Journal of Environmental Economics and Management* 32(1), 109-24.

Lo, C. K. Y., Yeung, A. C. L. and T.C.E. Cheng (2012): The Impact of Environmental Management Systems on Financial Performance in Fashion and Textiles Industries. *International Journal of Production Economics* 135(2), 561-67.

Mensah, J., Oppong, J. and C.M. Schmidt (2010) Ghana's national health insurance scheme in the context of the health MDGs: an empirical evaluation using propensity score matching. *Health Economics* 19(1), 95-106.

Nishitani, K. (2011) An Empirical Analysis of the Effects on Firms' Economic Performance of Implementing Environmental Management Systems. *Environmental and Resource Economics* 48, 569–86.

Portney, P.R. and R.N. Stavins (2000), Public Policies for Environmental Protection, Resources of the Future, 2nd edition, Washington, DC.

Prakash, A. (2002) Green Marketing, Public Policy and Managerial Strategies. *Business Strategy and the Environment* 11(5), 285-297.

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

Russo, M.V. and P.A. Fouts (1997) A Resource-based Perspective on Corporate Environmental Performance and Profitability. *Academy of Management Journal* 40(3), 534-59.