Effect of energy audits on the adoption of energy-efficiency measures by small companies

Joachim Schleich
Fraunhofer Institute for Systems and Innovation Research ISI/Grenoble Ecole de Management (GEM), Grenoble, France/Virginia Polytechnic Institute and State University, Blacksburg, Virginia, USA
Breslauer Straße 48
76139 Karlsruhe, Germany
joachim.schleich@isi.fraunhofer.de

Tobias Fleiter
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Straße 48
76139 Karlsruhe, Germany
tobias.fleiter@isi.fraunhofer.de

Simon Hirzel
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Straße 48
76139 Karlsruhe, Germany
simon.hirzel@isi.fraunhofer.de

Barbara Schlomann
Fraunhofer Institute for Systems and Innovation Research ISI
Breslauer Straße 48
76139 Karlsruhe, Germany
barbara.schlomann@isi.fraunhofer.de

Michael Mai
IREES GmbH
Schönfeldstraße 8
76131 Karlsruhe, Germany
michael.mai@irees.de

Edelgard Gruber
IREES GmbH
Schönfeldstraße 8
76131 Karlsruhe, Germany
edelgard.gruber@irees.de

Keywords

audit programme, evaluation, SME, econometric evaluation

Abstract

Energy audits for companies have been promoted since more than four decades, but no evaluation has yet been carried out which is based on the counterfactual behaviour of a large comparable control group. This paper empirically analyses the effect of the German energy audit program on small participating companies' decision to adopt energy efficiency measures. Non-parametric propensity score matching estimators are employed to estimate the average effect of the audit on more than 500 small companies (< 50 employees) participating in the program thereby relying on also more than 500 observations from a control group survey. Our findings so far are fairly robust across alternative matching algorithms and suggest that audit participation resulted in higher adoption of lower cost measures in particular, i.e. of energy efficient lighting (+ 20 percentage points) and of measures optimizing the heating system (+ 28 percentage points). The audit increased the adoption of thermal insulation measures by about 11 percentage points and of heating replacements by about 6–15 percentage points. Thus, the energy audits roughly doubled the adoption rates of energy efficient lighting, thermal insulation and heating replacements, and almost quadrupled the rates for heating optimization. For measures which companies had considered for implementation independent of the audit but had not (yet) planned to implement them, the audit effect remains the same for the lower cost measures, and increases significantly for the higher cost measures. Thus for the four measures considered, audits may contribute little to overcoming the lack of information about technology availability. Instead, energy audits appear to rather help overcome organizational barriers like intra-company priority setting, or lack of strategic importance.

Introduction

Energy audits for industry have been implemented since the mid-1970[ies] in several industrialized countries in response to the oil crisis. Today, more than 100 audit programs are estimated to be in place worldwide (Price and Lu 2013), typically involving government-funded subsidies. The EU Energy Efficiency Directive 2012/27 EU requires enterprises that are not medium sized companies (SMEs) to carry out an energy audit at least once every four years from 2015 on. EU Member States are also required to encourage SMEs to carry out energy audits. Above all, energy audits are expected to help overcome information-related barriers to energy efficiency. Better information about technology options and related energy costs savings, is expected to accelerate the adoption of energy efficiency measures.[1] Information-related barriers have been found to be prevalent in smaller organizations, in particular (e.g. Schleich and Gruber 2008, Schleich 2009), providing a rational for the focus of many government programs on SMEs.

So far however, only few studies have attempted to evaluate the effectiveness of energy audit programs in industry, typically relying on respondents' subjective responses in surveys (e.g. Fleiter et al. 2012a and 2012b). Thus, the estimated effects may suffer from social desirability and other biases. Existing analy-

1. See the overviews of industry audits by ECEEE (2014), Thollander and Palm (2013) and Price and Lu (2013).

ses for the residential sector generally provide mixed evidence on the effectiveness of audit programs (e.g. or Hirst and Goelz 1983, Frondel et al. 2013, Murphy 2014). In an early review of evaluations of utility home energy programs in the US, Hirst et al. (1981, p. 624) complain that "the lack of control groups in all but two of the evaluation efforts, […], seriously impairs the validity of the conclusions". To the best of our knowledge, no evaluation of energy audits in industry has so far relied on a comparison with a control group. Apart from costs to carry out the evaluation, the reasons likely involve heterogeneity of measures and companies, which render large sample evaluations with control groups empirically challenging.

In this paper we empirically analyse the impact of an energy audit program in Germany on the adoption of efficiency measures in small enterprises. Our evaluation relies on data from two surveys, which were carried out in Germany at about the same time. The treatment group data originally includes responses from more than 1,500 companies which previously participated in an energy audit. These energy audits were subsidized under the German energy audit program for small and medium sized companies from the industry and services sectors ("Energieberatung Mittelstand"). Our control group originally consists of about 2,000 companies participating in a representative energy use survey in the German tertiary sector. Since this sector also includes small companies from the commerce and industry sectors, the adoption decisions for alike energy efficiency measures may be compared.

Relying on the Roy-Rubin potential outcome evaluation framework (Roy, 1951, Rubin 1974), we employ non parametric propensity score matching algorithms to estimate the average treatment effects on the adoption of energy efficiency measures for companies participating in the German energy audit program. These analyses are carried out for four fairly generic energy efficiency technologies: lighting, thermal insulation of buildings, exchange of heating system and optimization of the heating system. Thus, our empirical analysis allows the effectiveness of audits to differ between measures. To prevent overestimation of the audit effect, our analysis does not include measures which companies reported to have decided to implement even before the energy audit had been carried out. In addition, we analyse whether audit effectiveness differs between the measures a company had already considered and the measures which it had not considered before. This allows assessing whether audits primarily address lack of information about energy efficiency technologies[2], or whether they rather help overcome organizational barriers like intra-company priority setting, or lack of strategic importance.[3]

The remainder of our paper is organised as follows. Section 2 outlines the evaluation framework and introduces the estimation methods. Section 3 describes the surveys and provides descriptive statistics of the data. Results are presented in Section 4. The final section discusses the main findings and concludes.

2. As pointed out by Metcalf and Hasset (1999), households receiving better information on energy efficiency options through audits may decide to forego some investments in response. Frondel et al. (2013) provide empirical support that participation in energy audits reduced the propensity to adopt buildings retrofit measures for a small share of households (4 %).

3. Lack of internal priority has been found to inhibit energy efficiency technology adoption by de Groot et al. (2001) and Schleich (2009). Cooremans (2011) highlights the role of strategic importance.

Methodology

GENERAL APPROACH

We analyse the effects of audits on the adoption of energy efficiency measure within the familiar evaluation framework developed by Roy (1951) and Rubin (1974)[4]. A binary treatment indicator D_i equals 1 if company *i* participated in the audit program and zero otherwise. $Y_i(D_i)$ denotes the potential outcome of the adoption decision for *i*. The treatment effect for company *i* may then be written as

$$\tau_i = Y_i(1) - Y_i(0) \tag{1}$$

Since for each *i* only one of the potential outcomes can be observed (fundamental problem of causal inference), we employ average treatment effects. In particular, we are interested in ATT (the average effect of treatment on the treated), i.e. the effect of the audit on those companies participating in the audit program:

$$\tau_{ATT} = E(\tau/D{=}1) = E[Y(1)/D{=}1] - E[Y(0)/D{=}1] \tag{2}$$

Naturally, the counterfactual mean for those participating in the audit, i.e. $E[Y(0)/D{=}1]$ cannot be observed. If participation in the audit program was random, $E[Y(0)/D{=}1] = E[Y(0)/D{=}0]$ and τ_{ATT} would be identified. In this case, the mean outcome of companies not participating in the audit program would serve as the counterfactual outcome of participating companies.[5] Estimates of τ_{ATT} based on non-randomized experiments, however, may suffer from a selection bias. That is, observable or unobservable company characteristics, which affect the decision to participate in an audit program, also affect the decision to adopt energy efficiency measures.

KEY ASSUMPTIONS

To account for non-random participation, our identification of τ_{ATT} relies on two standard assumptions: the conditional independence assumption (CIA)[6] and the common support assumption. The CIA means that – conditional on the set of relevant covariates – treatment assignment is independent of the outcomes. Thus, the covariates are nonresponsive to the audit participation. The CIA implies that audit participation only depends on observable company characteristics and that all covariates which affect both the audit participation and the adoption decision must be observed. The common support (or overlap) assumption means that companies with the same covariates have a positive probability of participating in the audit program and also of not participating in the audit program. In other words, each company has a positive probability of being in the control group and being in the audit group. To calculate ATT, it is sufficient that potential matches exist in the control group.

4. For a comprehensive overview see Angrist and Pischke (2009) or Wooldridge (2010).

5. This assumes the stable unit-treatment value assumption to hold. That is, the treatment effect for each company *i* in (1) is not affected by the participation of other companies in the audit program (no interference). Likewise, the audits are assumed to be comparable across companies (no variation in treatment).

6. The CIA has also been referred to as "selection on observables" (Imbens, 2004) or "unconfoundedness" (Rosenbaum and Rubin 1983). Note that regression-based analyses also rely on the CIA.

MATCHING METHOD: USING PROPENSITY SCORES

To estimate the ATT we employ matching estimators. Thus, we may rely on data of companies which did not participate in the audit program, but which exhibit similar relevant characteristics as the audit group. The difference in the adoption of energy efficiency measures between the companies in the audit group and in this control group may then be attributed to the audit program. In a sense, matching mimics "randomization" by balancing the distributions of the relevant characteristics (covariates) in the audit and in the control group, so as to attain independence between a company's decision to adopt a technology and its decision to participate in an energy audit. Unlike parametric estimators, the non-parametric estimators do not rely on correctly specified functional forms or on distributional assumptions.

Specifically, we employ *propensity score matching* estimators (Rosenbaum and Rubin 1983). This involves first running a logit (or a probit) model which regresses audit participation on a set of relevant covariates. Based on the coefficients audit participation is predicted for each company *i*. These propensity scores are then used to identify companies in the control group which best match the companies in the audit group. In this sense, the propensity score aggregates the information in the relevant covariates into a single index.[7] Hence, under propensity score matching audit group companies may be paired with control group companies exhibiting quite different covariate values, but close propensity scores. The ATT is estimated by calculating the difference between the share of adoptions in the audit group companies and their matches in the control group.

We first employ the most common propensity score matching where one company from the audit group is paired with one company from the control group. This standard *nearest neighbour* estimator selects the company in the control group that has the closest propensity score (e.g. Heckman et al. 1997).[8] This is done for all observations with common support, i.e. if the propensity scores of the companies in the control group overlap with those of the companies in the audit group. In addition, we allow audit group companies to be matched with multiple control group companies. Specifically, we report results for four nearest neighbours. Likewise, we apply the *Kernel* estimator, which relies on all companies in the control group (with common support) but attaches lower weights to – in terms of propensity score – more distant control group observations (e.g. Heckman et al. 1998). Asymptotically, the discussed estimators all yield identical results, but in practical applications there is typically a trade off between a lower variance and a larger bias due to the matching. Thus, in small samples the lower standard errors associated with multiple neighbours matching or with the Kernel estimator may come at the costs of a larger bias of the ATT estimator.

Data, variables and summary statistics

Our evaluation relies on data from two surveys, which were carried out in Germany at about the same time, and which included a set of identical questions about audit program participation and adoption of energy efficiency measures.[9]

AUDIT GROUP

The treatment group data includes responses from companies participating in an energy audit program. The energy audits are subsidized under the German energy audit program for small and medium sized companies (SMEs). This program was launched in 2008 as "Sonderfonds Energieeffizienz in KMU" and continues since March 2012 as "Energieberatung Mittelstand". It aims at overcoming information barriers via high-quality energy audits. Eligible companies are SMEs with up to 250 employees. In March 2012 an additional criterion was introduces which required annual energy costs to exceed 5,000 Euros.

Audits funded by the program are carried out on site by independent professional energy auditors. Grants are provided for both 'initial' and 'detailed' audits. Initial audits are short audits which last up to two days. These short audits focus on identifying major energy saving potentials and measures at the audited sites. They are especially suited for companies with little energy demand/simple energy systems and to obtain a rough overview of possible energy savings potentials in other companies. Detailed audits may last of up to ten days. Thus, the auditors can further elaborate the analysis, carry out in-depth monitoring activities and provide detailed action plans and recommendations.

In the case of initial audits, eligible companies may obtain a funding of up to 80 % of the daily rate of the energy auditor (maximum of 640 Euros/day) for up to two days. For detailed audits, funding is granted for 60 % of the daily rate of the audits (maximum of 480 Euros/day) for up to ten days. Thus, the maximum funding is 1,280 Euros and 4,800 Euros respectively. Since the introduction of the programme in 2008, more than 24,300 companies received funding via the audit program. According to Gruber et al. (2006), the 4,000–5,000 audits carried out per year represent about 20 % of the market potential.

The data we use for our analysis is based on a recent evaluation of the German audit program (Mai et al. 2014). As part of the evaluation an online survey was conducted among companies that had received funding within the program. The survey addressed the characteristics of the respective companies, the perception of the funding program, the audits and the auditors as well as detailed questions on specific energy efficiency measures.

A total of 9,981 SMEs that received funding via the program since March 2012 was contacted to participate in the survey via e-mail. With approximately 700 delivery failures, roughly 9,200 companies received the invitation to participate. The survey was open for five weeks in April/May 2014 and provided a total of 1,523 completed questionnaires. The size of the sample and the response rate (17 %) is considerably higher than during a previous evaluation of the programme carried out in 2010

7. The CIA now means "conditional on the propensity score". Thus, adjusting for the propensity score only is sufficient to eliminate confounding (Rosenbaum and Rubin 1983). The common support assumption means that the range of the propensity score of the control group must cover that of the audit group.

8. In our analysis we perform matching with replacement, i.e. a control group company may be matched with more than one audit group company.

9. Including this set of identical questions was feasible since some of the authors were involved in both surveys. Mai et al. (2014) documents the evaluation of the energy audit program. Schlomann et al. (2014) reports the findings on the energy use in the German tertiary sector. Both studies were funded by the Federal German Ministry of Economics. These reports however do not include any of the findings presented in this paper, though. They also contain the questionnaires (in German) used.

(Fleiter et al. 2012a). The responses reflect the structure of the population of funded companies quite well in terms of company size and sector distribution. The data was intensively checked for plausibility and internal validity prior to further analyses.

CONTROL GROUP

The control group data relies on responses of companies participating in a representative energy use survey in the German tertiary sector, which also includes small companies from the commerce and industry sectors.

The control group survey originally involves a total of 2013 companies and public institutions in the tertiary sector in Germany and was carried out between February and July 2014 by the international market research company GfK SE (Gesellschaft für Konsumforschung). In each organization either the energy manager or – if such a position did not exist – the person responsible for energy management was interviewed via Computer Assisted Personal Interviews (CAPI). Similar surveys had been carried out five times since 2001 by the same institutes (Schlomann et al. 2014). In the survey, the tertiary sector was defined in the same way as in the German national energy balances. The sector therefore covers all public and private services and trade, agriculture, construction and some small industrial enterprises with less than 20 employees. For the matching analysis the sub-sectors were aggregated to six sectors (see Table A1).

By design, most companies in this control group are rather small. Less than 2 % have more than 50 employees. To allow for comparable data sets, we restrict observations in both groups to small companies, i.e. to companies with less than 50 employees.

The survey gathered general company information on company size, organizational structure, or energy costs, and specific information on technologies and on energy consumption for different end uses. Unlike previous surveys for this sector, the recent survey included a separate part which asked identical questions on the four energy efficiency measures as in the audit group.

BASIS FOR COMPARISON

To allow for meaningful comparisons both surveys contain identical questions on the adoption of four generic energy efficiency measures, which are typically recommended in energy audits: lighting replacement (*lighting*), thermal insulation of the building (*insulation*), replacing the heating system (*heating*) and optimization of the heating system (*heating optimization*). These measures are fairly generic and typically explored in any energy audit, but they differ in terms of investment costs. In our combined sample, average reported investment cost for *lighting* are about 11,500 Euros, for *insulation* 72,000 Euros, for heating 61,500 Euros and for *heating optimization* 40,000 Euros. Based on respondents' self-reported information adopted lighting measures on average saved 24 % of a company's lighting electricity, insulation measures saved 17 % of heating energy, exchanging the heating system saved 21 % of heating energy, and optimizing the heating system also saved about 17 % of heating energy. Similarly, the average payback times are 4.2 years for lighting, 16.5 years for insulation, 6.8 years for heating and 5.3 years for *heating optimization*.[10]

The respondents in both groups were asked whether their company had adopted any of these measures since 2008. Audit group companies were also asked whether they had planned to implement a particular measure prior to participating in the audit program. Observations of companies who indicated that this was the case were deleted from the estimation of the ATT for this particular measure to avoid overestimating the audit effect (windfall effects). This resulted in a loss of more than 10 % of the observations. The survey also provided information on whether the company had previously considered implementing a particular measure. Accordingly, about 27 % of the companies in the audit group had previously considered (but not planned to implement) *lighting*. The corresponding figures for the other measures are 11 % for *insulation*, 19 % for *heating* and 14 % for *heating optimization*.

MATCHING VARIABLES

Both surveys collected information on variables which the empirical literature found to affect companies' adoption of energy efficiency measures.[11] We are not aware of empirical analyses of factors driving audit program participation, but implicitly assume that the same set of variables is relevant for both, technology adoption and program participation.

In particular, we include the share of energy costs on total costs (*energy cost share*) to capture companies' financial incentives to invest in energy efficiency and also the strategic importance of energy to the company.[12], [13] The number of employees is included to capture the effect of company size.[14] Adoption propensity is expected to be higher for larger companies since they have more resources available to acquire technological and financial know how, may more easily overcome information costs and other transaction costs related barriers. Larger firms may also better spread the risk of technology adoption, and may more easily acquire external funding. Companies with an energy manager (as an organizational unit) are more likely to adopt energy efficiency measures because energy managers' responsibilities typically include controlling and optimizing energy costs. Thus, they need to be informed about energy savings options, energy consumption and energy costs. Also, companies establishing the organizational unit of an energy manager are likely to attach a higher relevance to company energy use compared to companies without such a position.

The form of company ownership may also matter for technology adoption. On the one hand, subsidiaries may exhibit higher adoption rates because of positive spill-over effects (e.g. related to information) from their mother company. On the other hand, if the mother organization appropriates the benefits, i.e. the energy cost savings, the propensity to adopt may be lower for sub-

10. These figures are based on a smaller sample, since not all respondents provided information.

11. Analyses for the industry sector include Velthuijsen (1993), Thollander and Ottosson (2010), Abdelaziz et al. (2011), and Stenqvist et al. (2011) and for the tertiary sector and SMEs Schleich and Gruber (2008), Schleich (2009), Fleiter et al. (2012), Trianni and Cagno (2012), and Schlomann and Schleich (2015).

12. We also considered including energy prices. In particular, we calculated electricity prices as the ratio of electricity expenditures and electricity use. For loss of observations, however, we ended up not including this variable in the subsequent analyses.

13. Note that strictly speaking our measure of energy intensity may be affected by audit participation and hence violate the CIA. However, for the companies in our sample, energy costs account for a relatively small share of total costs (on average ca. 9 % for audit group, see Annex Tables). Also the impact of the measures considered on total company energy costs tends to be small.

14. For the propensity score matching analysis we use the log of employees.

Table 1. Descriptive statistics of the adoption of energy efficiency measures.

	Group	N	Share	Difference
lighting	audit	582	54.47%	24.04% ***
	control	585	30.43%	
insulation	audit	601	18.97%	10.31% ***
	control	589	8.66%	
heating	audit	590	21.86%	14.20% ***
	control	509	7.66%	
heating optimization	audit	622	36.01%	28.66% ***
	control	517	7.35%	

*Note: *** indicates significance at $p < 0.01$ in an individual two-tailed t-test.*

sidiaries (split-incentives problem). Similarly, renting (rather than owning) a building may decrease a company's financial incentives to adopt energy efficient technologies because renters are less likely to appropriate the benefits of the energy efficiency investment (landlord-tenant or user-investor dilemma). Finally, to capture heterogeneity across sectors, we included sector dummies. These sector dummies were constructed by aggregating sub-sectors with similar technologies and energy use patterns. For the final audit group and control group samples Table 1 provides the descriptive statistics for the outcome variables in the audit and in the control group samples and Annex Table A2 for the covariates for the lighting sample.[15] Both tables suggest that the differences in the outcome variables and the covariates are statistically significant (at $p < 0.01$). A simple comparison of adoption rates between groups in Table 1 suggests that adoption shares of all four energy efficiency measures are substantially higher in the audit group than in the control group. Also, the adoption shares of low cost measures (lighting, heating optimization) are higher than those of high cost measures (insulation, heating). Table A2 also suggests that the (average) values of most covariates differ significantly between the audit and the control groups (before matching). Compared to the control group the audit group companies are on average (slightly) less energy intensive, substantially larger, more likely to have an energy manager, less likely to be a subsidiary and more likely to rent their buildings, less likely to belong to the trade sector and more likely to belong to the metal sector.

Results

We use STATA 13 to estimate the matching estimators.[16] Table A3 in the Annex presents the findings of the logit models underlying the propensity score matching estimators. For the samples of all four measures the propensity to participate in the audit is positively and statistically significantly related to the (log of) *energy cost share* and to the number of *employees*. Having an *energy manager* is only (negatively) related to audit participation for the lighting sample. For all four samples, *subsidiary* and also *rented* are found to decrease the propensity to participate in an audit. Also, relative to the base sector (other production) most sector dummies turn out to be statistically significant for all four samples. In sum, these findings are quite intuitive.

PROPENSITY SCORE MATCHING ESTIMATORS

Table 2 first presents the results for the propensity score estimator with one nearest neighbour [nn(1)]. The adjacent column presents the findings when an audit group company is matched with four nearest neighbours [nn(4)].[17] The third set of results in Table 2 show the results for the Kernel estimator.[18] The nearest neighbour estimates presented in Table 2 satisfy the common support assumption, i.e. to estimate the audit effects only audit group observations are used where the propensity scores overlap with control group observations.[19]

Our findings suggest that energy audits increase the adoption of all four energy efficiency measures for small companies participating in energy audits. The estimates of audit effectiveness for *lighting* and *insulation* hardly differ across the estimators. On average, audits increased the propensity to adopt lighting measures by about 20 percentage points for companies in the audit group. Similarly, the ATT for insulation is estimated at about 11 percentage points. For *heating* our estimates of the ATT ranges from about 6 percentage points for nn(1) to 9.5 percentage points for the Kernel estimator. The lower point estimate for the nn(1) compared to the Kernel estimator may be explained by the fact that for 60 audit companies the common support as-

15. The size of the samples differs across the four measures mainly because of differences in the number of companies which had already planned adoption differs across measures. In addition, there are small differences in the number of missing values for adopted measures across those measures. The equivalent tables to A2 for *insulation*, *heating* and *heating optimization* are available upon request.

16. We employ the *teffects psmatch* command provided in STATA 13 and the *psmatch2* command written by Leuwen and Sianesi (2003). In case of ties of propensity scores of control group matches, we allow for matches with all tied observations.

17. Results are robust to different choices for the number of neighbors.

18. Significance levels for the nearest neighbors propensity matching estimators are based on standard errors provided by the *teffects* command which recognizes that propensity scores are estimated rather than known with certainty (see Abadie and Imbens 2012). Calculating the standard errors via bootstrapping may over- or underestimate the true standard errors (in our case for the Kernel estimator).

19. In the propensity score matching with nearest neighbours 7 audit group observations were off support for *lighting*, 3 for *insulation*, 60 for *heating* and 58 for *heating optimization*.

Table 2. Audit effects (in percentage points).

		Propensity score estimators						
Measure	Group	N	nn(1)	nn(4)	N	Kernel	N	NNM(1)
lighting	audit	575	20.74 ***	18.61 ***	582	20.73 ***	1167	19.67 ***
	control	585			585			
insulation	audit	598	11.18 ***	9.86 ***	601	10.52 ***	1190	10.62 ***
	control	589			589			
heating	audit	530	5.94 *	8.00 **	572	9.54 ***	1099	14.73 ***
	control	509			509			
heating optimization	audit	564	26.95 ***	27.66 ***	607	28.81 ***	1139	26.54 ***
	control	517			517			

*Note 1: *** indicates significance at p < 0.01, ** indicates significance at p < 0.05 and * indicates significance at p < 0.1 in an individual two-tailed t-test.*

Note 2: Sample sizes for nn(4) are the same as for nn(1).

sumption was violated and they were excluded when calculating the nn(1) estimator. Thus, for these off-support audit companies, the propensity to exchange the heating system compared to their nn(1) matches must have been much higher than on average for the on-support audit companies compared to their matches. Also, the comparison of the results of nn(1) with those of nn(4) in particular suggest, that an audit group company's propensity to exchange the heating system differs less compared to its closest neighbour than compared to slightly more distant neighbours. In general, the p-values of the audit effect estimate for heating are higher than for the other measures. Finally, the findings for the algorithms in Table 2 suggest that for *heating optimization* and insulation in particular, the ATT of audits are similar to the differences in sample means in Table 1.

ROBUSTNESS CHECKS AND MISSPECIFICATION DIAGNOSTICS[20]

To check the robustness of the findings in Table 2 we first allow for alternative propensity score matching algorithms. The nearest neighbour may be quite far away in terms of propensity score. To avoid using poor matches to estimate the ATT we require the distance in propensity scores between an audit company and its control group match to be below a pre-specified threshold, the so-called *caliper distance*. Thus, only audit group companies with a match that satisfies this criterion will be included in the calculations. Since no agreed upon definition of this maximum distance exists, we follow Rosenbaum and Rubin (1985) and Austin (2011) and use 0.2 of the standard deviation of the logit of the propensity scores of the pooled audit and (matched) control group observations. The findings of the caliper estimators for nn(1) and nn(2) are very close to those presented in Table 2. The caliper-estimated ATT for *heating* is around 9 percentage points for nn(1) and nn(4), hence close to the estimate by the nn(4) and Kernel algorithms in Table 2.

To control for the impact of potentially important outliers in the audit group, we estimate the ATT by nn(1) and nn(4) but allow the common support condition to be violated. The results are very close to those of the propensity score estimators presented in Table 2. For heating, the ATT is estimated at around 9 percentage points. Thus with the possible exception of heating, imposing the common support condition does not result in important outliers to be dropped when estimating the ATT.

In addition to controlling for common support as in our nn(1) and nn(4) model runs, testing the specification of the propensity score models typically involves checking whether the balancing assumption is violated. We therefore examine whether after conditioning on the propensity score there remain systematic differences in the covariates between the audit group and the control group. The results are presented in the Annex. As can be seen for lighting in Annex Table A2, after matching the differences in the unmatched samples for the continuous variables (*energy cost share* and *employees*) are no longer statistically significant for any of the measures. Matching has also largely reduced the differences for *energy manager, subsidiary* and *rented* for all measures. For *energy manager* a statistically significant difference remains after matching for lighting only. For *subsidiary* the difference remains statistically significant for all measures but lighting. For *rented*, matching does not eliminate the difference for any of the measures. For each measure between one and three of the six sector dummies remain different even after matching (typically hotels and restaurants).

Thus, while matching has significantly reduced the differences in the covariates means between the audit and the control group, we cannot preclude model misspecifications, which may impact on our estimates of the audit effects on the adoption of energy efficiency measures.

To assess the potential implications we employ the *nearest neighbour matching* (NNM algorithm). While the propensity score estimators matches on a single continuous variable, NNM uses the outcomes of similar subjects. Similarity is defined by a weighted function of the covariates for audit group companies and control group companies. Specifically, we use the STATA *teffects nnmatch* command[21]. In light of the findings of the balancing diagnostics, we request exact matching for the categorical variables *emanager*, *subsidiary* and *rented*. The findings for NNM

20. All results not shown to save space are available from the authors upon request.

21. For the weighting of the covariates *nnmatch* employs by default the Mahalanobis distance. Thus, the weights are based on the inverse of the variance–covariance matrix of the covariates. In Table 2 we report results for one nearest neighbour, but findings are robust to variations in the number of nearest neighbours. Since the NNM estimators are not consistent when matching is based on two or more continuous covariates (Abadie and Imbens 2006, 2011) we use the bias correction option implemented in STATA.

Table 3. Audit effects for previously considered measures only (in percentage points).

Measure	Group	N	nn(1)	nn(4)	N	Kernel	N	NNM(1)
lighting	audit	391	17.18 ***	20.30 ***	392	19.39 ***	977	17.33 ***
	control	585			585			
insulation	audit	157	35.35 ***	27.77 ***	157	30.85 ***	746	26.33 ***
	control	589			589			
heating	audit	235	14.68 ***	16.74 ***	247	16.28 ***	764	22.02 ***
	control	509			509			
heating optimization	audit	177	24.58 ***	23.22 ***	193	24.42 ***	716	21.08 ***
	control	517			517			

are presented in the last column in Table 2[22]. The results of this NNM estimator are presented in the last columns in Table 2. For lighting, insulation, and heating optimization, the findings of the NNM estimator are virtually identical to those of the propensity score estimators. In contrast for "heating", the estimate for the audit effect by NNM is significantly higher than by propensity score estimation. Arguably, the failure of satisfying the balancing condition by the propensity score estimators for *subsidiary* and *rented* in particular may lead the propensity score estimators to underestimate the effects of audits.

EFFECTS OF ENERGY AUDITS ON ADOPTION OF MEASURES WHICH COMPANIES HAD ALREADY CONSIDERED INDEPENDENT OF THE AUDIT

For lighting Table 3 presents the result of the matching analyses for measures which had been considered by the companies for adoption independent of the audit. Again, findings on ATT are quite similar across the various algorithms with the possible exception of heating. As in Table 2, the estimate of ATT for heating is higher when the NNM(1) estimator is employed rather than the propensity score estimators. Results from balancing diagnostics generally are better when the matching is performed for measures which had previously been considered by the companies. The findings in Table 3 suggest that for the lower-cost measures (lighting and heating optimization) the levels of the ATT are quite similar for pre-considered measures as for all measures. However, for insulation and – to a lesser extent also for heating – the effects of audits are found to be significantly higher for measures which companies had already considered independent of the energy audit.

Conclusions

Based on non-parametric matching analyses we find that the German energy audit program accelerated the adoption of four generic energy efficiency measures in small companies. In absolute percentage terms, the estimates for the ATT are highest for the lower-cost measures considered, i.e. *lighting* (20 percentage points) and *heating optimization* (28 percentage points) and lower for the higher cost measures thermal *insulation* (11 percentage points) and exchange of the *heating* system (6–15 percentage points). In relative terms, the energy audits roughly double the adoption rates of *lighting,* thermal *insulation* and *heating* replacements, and almost quadruple the rates for *heating optimization* – a measure that may more likely be overlooked by non-energy experts than the other three generic measures. These findings also suggest that the effectiveness of energy audits vary by technologies. Thus, using program effectiveness indicators like 'the number of additional measures induced by an energy audit' are likely to be misleading. The matching algorithms applied produce fairly robust results, in particular for *lighting*, *insulation* and *heating optimization.* Yet we cannot rule out that the remaining (slight) differences in some covariate means after matching lead to inconsistent propensity score matching estimates of the ATT, in particular for *heating.*[23] Our findings employing nearest neighbour matching suggest though, that our propensity score estimators likely underestimate the audit effects for heating. However, since audit participation is not random and arguably subject to self selection, unobserved heterogeneity of the propensity to participate in the energy audit may also affect the propensity to adopt the energy efficiency measures. In this case, our findings would overestimate the effectiveness of the energy audits. Likewise, if social desirability resulted in "over reporting" of adoption rates by companies in the audit group compared to the control group the ATT would be overestimated.

Our findings for measures which had been considered for adoption independent of companies' participation in energy audit provide insights into the function of energy audits. For the two lower cost measures lighting and heating optimization, the ATT does not appear to differ compared to measures which had not been considered before. For the higher cost measures (in particular for *insulation*) we even find that audit effects are significantly higher when measures had been considered by companies before. These findings suggest that for the higher cost measures audits may contribute little to overcoming the pure lack of information about technology availability. Instead, energy audits appear to rather help overcome organizational barriers like intra-company priority setting, or lack of strategic importance. In this sense energy audits – similar to the recommendations by management consulting companies – appear to facilitate the realization of measures which had been discussed, but did not gather sufficient organizational support for

22. The findings are also robust if exact matching is required for *restaurants* in addition to exact matching for *emanager*, *subsidiary* and *rented*.

23. Also, our analysis did not distinguish between audit types, possibly violating the no-variation-in treatment assumption. This caveat should be considered when interpreting the results. We intend to address this potential shortcoming in the next step of our analyses.

implementation. Additional qualitative analyses may provide further insights into this issue. This finding also highlights the importance of other means (than audits) providing information about energy efficiency technologies like technology labelling, energy efficiency networks or information campaigns.

References

Abadie, A., and G. W. Imbens. 2006, Large sample properties of matching estimators for average treatment effects. Econometrica 74: 235–267.

Abadie, A., and G. W. Imbens, 2011. Bias-corrected matching estimators for average treatment effects. Journal of Business and Economic Statistics 29: 1–11.

Abadie, A. and Imbens, G., 2012, Matching on the estimated propensity score. (Download http://www.hks.harvard.edu/fs/aabadie/pscore.pdf)

Abdelaziz, E.A., Saidura, R., Mekhilef, S.(2011): A review on energy saving strategies in the industrial sector. Renewable and Sustainable Energy Reviews 15 (1), 2011, 150–168.

Anderson, S.; Newell, R., 2004, Information programs for technology adoption: the case of energy-efficiency audits. Resource and Energy Economics 26, 27–50.

Angrist, J.D. and Pischke, J.S., 2009, Mostly harmless econometrics: an empiricist's companion. Princeton University Press, Princeton, New Jersey.

Austin, P. C., 2011, Optimal caliper widths for propensity-score matching when estimating differences in means and differences in proportions in observational studies. Pharmaceutical Statistics 10, 150–161.

Cagno, E.; Trianni, A., 2014, Evaluating the barriers to specific industrial energy efficiency measures: an exploratory study in small and medium-sized enterprises- Journal of Cleaner Production 82, 70–83.

Cooremans, C. (2011). Make it strategic! Financial investment logic is not enough. Energy Efficiency, 4(4), 473–492.

de Groot, H.L.F.; Verhoef, E.T.; Nijkamp, P., 2001. Energy saving by firms: decision making, barriers and policies. Energy Economics, 23 (6), 717–740.

ECEEE (2014), Industrial audits – A comprehensive and evidence-based primer for policy-makers and decision makers in the private and public sectors, eceee Policy Guide # 1. The European Council for an Energy Efficient Economy, December.

Fleiter, T. Gruber, E.; Eichhammer, W.; Worrell, E., 2012a, The German energy audit programme for firms – a cost-effective way to improve energy efficiency? Energy Efficiency 5 (4), 447–469.

Fleiter, T.; Schleich, J.; Ravivanpong, P., 2012b, Adoption of energy-efficiency measures in SMEs – An empirical analysis based on energy audit data from Germany. Energy Policy 51, 863–875.

Gruber, E., Jochem, E., Mannsbart, W., Roser, A., Ott, V., 2006, Marktstudie Sonderfonds für Energieeffizienz in Industrie und Gewerbe, Karlsruhe: Fraunhofer Institute for Systems and Innovation Research; Institute for Resource Efficiency and Energy Strategies (IREES).

Frondel, M.; Vance, C., 2013, Heterogeneity in the Effect of Home Energy Audits: Theory and Evidence. Environmental and Resource Economics 55, 407–418.

Gillingham, K., Newell, R.G. and Palmer, K., 2009, Energy Efficiency Economics and Policy. Annual Review of Resource Economics 1: 14.1–14.23.

Heckman, J., Ichimura, H., and Todd, P., 1997, Matching as an Econometric Evaluation Estimator: Evidence from a Job Training Program, Review of Economic Studies 64, 605–654.

Heckman, J., Ichimura, H., Todd, P., 1998, Matching as an econometric evaluation estimator. Review of Economic Studies 65, 261–294.

Hirst, E.; Berry, L.; Soderstrom, J., 1981, Review of utility home energy audit programs. Energy 6, 621–630.

Hirst E.; Goeltz, R.,1985, Estimating energy savings due to conservation programmes: the BPA residential weatherization pilot programme. Energy Economics 7, 20–28.

Imbens, G.W., 2004. Non parametric estimation of treatment effects under exogeneity: a review. Review of Economics and Statistics 86, 4–29.

Leuven, E. and Sianesi, B. (2003). PSMATCH2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing. (Download: http://EconPapers.repec.org/RePEc:boc:bocode:s432001)

Mai, M. et al., 2014, Evaluation des Förderprogramms "Energieberatung im Mittelstand" Schlussbericht Im Auftrag des Bundesministeriums für Wirtschaft und Energie.

Metcalf G. E. and Hasset, K., 1999, Measuring the energy savings from home improvement investments: evidence from monthly billing data. Review of Economic Statistics 81, 516–528.

Murphy, L., 2014, The influence of energy audits on the energy efficiency investments of private owner-occupied households in the Netherlands, Energy Policy 65, 398–407.

Price, L. and Lu, H., 2011, Industrial energy auditing and assessments: A survey of programs around the world. eceee 2011 Summer Study Proceedings, eceee 2011 629–640.

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

Roy, A ., 1951, Some thoughts on the distribution of earnings. Oxford Economic Papers 3, 135–145.

Rosenbaum, P. R.; Rubin, D. B., 1985, Constructing a control group using multivariate matched sampling methods that incorporate the propensity score. The American Statistician 39, 33–38.

Rubin, D., 1974, Estimating causal effects to treatments in randomised and nonrandomised studies. Journal of Educational Psychology 66, 688–701.

Schleich, J., 2004, Do energy audits help overcome barriers to energy efficiency? An empirical analysis for Germany, International Journal of Energy Technology and Policy 2, 226–239.

Schleich. J.; Gruber, E., 2008, Beyond case studies: Barriers to energy efficiency in commerce and the services sectors. Energy Economics 30, 449–464.

Schlomann, B.; Geiger, B.; Gruber, E.; Kleeberger, H. (Hrsg.) (2014). Energieverbrauch des Sektors Gewerbe, Handel, Dienstleistungen (GHD) in Deutschland für die Jahre 2006 bis 2011. ISI-Schriftenreihe "Innovationspotenziale". Stuttgart: Fraunhofer Verlag. ISBN 978-3-8396-0691-9.

Schlomann B., Schleich J., 2015. Adoption of low-cost energy efficiency measures in the tertiary sector – an empirical analysis based on energy survey data. Renewable & Sustainable Energy Reviews 43, 1127–1133.
Stenqvist, C., Nilsson, L. J., Ericsson K., Modig, G. (2011). Energy management in the Swedish pulp and paper industry – the daily grind that matters. Proceedings of the eceee 2011 Summer Study, 3, 2011 767–776.
Thollander, P. and Palm, J., 2013, Improving Energy Efficiency in Industrial Energy Systems: An Interdisciplinary Perspective on Barriers, Energy Audits, Energy Management, Policies, and Programs, Springer, London 2013.
Trianni, A., Cagno, E. (2012). Dealing with barriers to energy efficiency and SMEs: Some empirical evidences. Energy, 37 (1), 494–504.
Velthuijsen, J. 1993. Incentives for investment in energy efficiency: an econometric evaluation and policy implications. Environmental and Resource Economics, 3. (2), pp. 153–169.
Wooldridge, J. M., 2010, Econometric Analysis of Cross Section and Panel Data. 2nd Edition. MIT Press, Cambridge, Massachusetts.

Annex

Table A1. Overview of sector composition.

Aggregated sectors	Detailed sub-sectors
Hotels and restaurants	Hotels, restaurants
Trade	Retail trade, wholesale trade
Other services	Car repair industry, laundries and dry cleaners, banks, insurance companies, publishing houses, rest of services
Metal	Metal processing
Foodstuffs	Bakeries, butchers, rest of food
Other production	Textiles, trucking, wood working and processing, paper processing and printing, construction, agriculture, horticulture

Table A2. Means of covariates between audit and control group before and after nn(1) matching (for lighting).

Covariates		Unmatched/ Matched	Mean		Difference between groups
			Audit group	Control group	
energy cost share	share	U	0.091	0.111	***
		M	0.092	0.100	
employees (log)	numbers	U	2.692	1.488	***
		M	2.678	2.664	
emanager	0/1 dummy	U	0.115	0.065	***
		M	0.117	0.153	*
subsidiary	0/1 dummy	U	0.076	0.118	**
		M	0.077	0.090	
rented	0/1 dummy	U	0.423	0.583	***
		M	0.428	0.369	*
hotels and restaurants	0/1 dummy	U	0.115	0.094	
		M	0.113	0.129	
trade	0/1 dummy	U	0.222	0.275	**
		M	0.224	0.230	
services	0/1 dummy	U	0.253	0.292	
		M	0.252	0.285	
metal	0/1 dummy	U	0.125	0.031	***
		M	0.122	0.085	**
foodstuffs	0/1 dummy	U	0.067	0.074	
		M	0.068	0.043	*
other production	0/1 dummy	U	0.218	0.234	
		M	0.2209	0.2278	

*Note 1: *** indicates significance at p < 0.01, ** indicates significance at p < 0.05 and * indicates significance at p < 0.1 in an individual two-tailed t-test.*

Table A3. Logit results of audit participation.

Covariates	lighting		insulation		heating		heating optimization	
energy cost share	1.07	*	1.64	***	1.55	**	1.63	**
	(0.61)		(0.60)		(0.61)		(0.62)	
employees	1.43	***	1.45	***	1.46	***	1.54	***
	(0.09)		(0.09)		(0.09)		(0.10)	
emanager	-0.26	***	-0.12		-0.27		-0.26	
	(0.26)		(0.25)		(0.26)		(0.26)	
subsidiary	-0.91	***	-0.47	*	-0.62	**	-0.84	**
	(0.25)		(0.26)		(0.26)		(0.27)	
rented	-0.53	***	-0.53	***	-0.72	***	-0.56	***
	(0.15)		(0.15)		(0.16)		(0.15)	
hotels and restaurants	0.73	***	1.26	***	0.90	***	1.33	**
	(0.27)		(0.26)		(0.27)		(0.28)	
trade	0.24		0.54	***	0.45	**	0.62	***
	(0.22)		(0.21)		(0.22)		(0.22)	
services	0.75	***	1.23	***	0.94	***	1.21	**
	(0.22)		(0.22)		(0.23)		(0.23)	
metal	0.99	***	0.80	***	0.70	**	0.72	
	(0.34)		(0.30)		(0.32)		(0.31)	
foodstuffs	-0.10		0.14		0.12		0.04	
	(0.30)		(0.31)		(0.34)		(0.31)	
constant	-3.16	***	-3.56	***	-3.22	***	-3.53	***
	(0.28)		(0.28)		(0.29)		(0.30)	
LR(Chi2)	463.41	***	461.56	***	426.85	***	471.05	***
Pseudo R2	0.2864		0.2798		0.2813		0.3002	
N	1167		1190		1099		1139	

*Note 1: *** indicates significance at $p < 0.01$, ** indicates significance at $p < 0.05$ and * indicates significance at $p < 0.1$ in an individual two-tailed t-test.*