Effectiveness of energy audits in small business organizations

Joachim Schleich [a,b,c,*], Tobias Fleiter [b]

[a] *Grenoble Ecole de Management (GEM), France*
[b] *Fraunhofer Institute for Systems and Innovation Research ISI, Germany*
[c] *Virginia Polytechnic Institute & State University, USA*

ARTICLE INFO

ABSTRACT

Energy audits for business organizations have been promoted for more than four decades, but no evaluation based on the counterfactual behaviour of a comparable large control group has yet been carried out. Seeking to close this gap, this paper analyses the effect of a German energy audit programme involving more than 1400 small manufacturing and non-manufacturing organizations. The control group observations were drawn from a parallel study involving more than 2000 organizations. Limiting the sample to business organizations with at most 50 employees, the study employs coarsened exact matching, and–as a robustness check–conventional propensity scores as well as distance-based matching to estimate the effectiveness of simple and detailed audits on the adoption of four ancillary energy efficiency measures. The findings suggest that both types of audits spur the adoption of lighting, insulation, heating systems, and operational measures to improve heating systems (operations) by between 10 and 20 percentage points. Audit effectiveness was highest for insulation measures and operations. In addition, the findings suggest a positive but diminishing interaction between audit effectiveness and organization size for lighting, insulation and operations. These results are robust across alternative matching methods.

© 2017 Elsevier B.V. All rights reserved.

1. Introduction

In response to the first oil crisis in the 1970s several industrialized countries implemented energy audit programmes for industry. Currently, it is estimated that there are more than 100 audit programmes in place worldwide (Price and Lu, 2011), typically involving government-funded subsidies. As of 2015, the European Union (EU) Energy Efficiency Directive 2012/27 requires business organizations that are not small or medium-sized enterprises (SMEs) to carry out energy audits at least once every four years. Although such audits are not obligatory for SMEs, EU member states are required to encourage SMEs to carry them out. The ultimate aim of energy audits is to overcome information-related barriers to energy efficiency (e.g., Anderson and Newell, 2004; Schleich, 2004). Better information about technology options and related energy cost savings is expected to accelerate the adoption of energy efficiency measures. Information-related barriers have been found to be prevalent in smaller organizations in particular (e.g. Schleich and Gruber, 2008; Schleich, 2009), providing a rationale for many government programmes to focus on SMEs.

Despite the popularity of energy audits, little is known about their effectiveness. Methodological challenges involved in accurately measuring the effectiveness of audits include, in particular, sample selection bias, which would result if compa-

---

nies that adopt low-cost energy efficiency measures are more likely to participate in subsidized energy audits. Moreover, some companies have implemented energy efficiency measures in the absence of audits. Unless these factors are properly accounted for in policy evaluation, sample selection would likely lead to overestimating policy effectiveness (e.g. Joskow and Marron, 1992). So far, only a few studies have attempted to evaluate the effectiveness of energy audit programmes that involve technology adoption.[1] Typically, these studies rely on subjective assessments that are collected via surveys (e.g. Thollander et al., 2007; Fleiter et al., 2012a,b; Backlund and Thollander, 2015; Paramonova and Thollander, 2016) and, therefore, are prone to social desirability bias. To the best of our knowledge, no extant evaluations of energy audits for companies have relied on comparison with a control group. Apart from the costs involved in conducting such an evaluation, assembling a control group of comparable companies and measures is challenging in practice.[2] In addition, control groups need to be sufficiently large to allow for statistical analyses. In our study we seek to close this gap.

In this paper we empirically analyse the impact of a voluntary federal German energy audit programme for SMEs ('Energieberatung Mittelstand') on the adoption of four ancillary energy efficiency measures by small businesses from manufacturing and non-manufacturing sectors. The results also provide insights into whether audit effectiveness differs across these measures. As this programme subsidizes both simple and more detailed energy audits, we explore the relative effectiveness of audit intensity for these measures. Finally, we investigate whether audit effectiveness varies with organization size. For example, audits may be less effective for larger organizations because barriers related to information or know-how are typically lower in them (e.g. Fleiter et al., 2012b). Or perhaps audit effectiveness is greater for larger organizations because they are more likely to have the financial resources needed to implement an audit's investment recommendations (e.g. Asensio and Delmas, 2017).[3] Thus, our findings provide valuable design and evaluation guidance for policies such as the EU Energy Efficiency Directive 2012/27EU, which requires non-SMEs to carry out regular energy audits.

This paper employs data from two surveys administered in Germany in 2014. The original treatment group data include responses from more than 1400 organizations that had previously participated in the German energy audit programme. The control group is recruited from a representative survey involving more than 2000 organizations in the German 'trade, commerce, services and other consumers' sector, which also includes small manufacturing enterprises.

Relying on the Roy-Rubin potential outcome evaluation framework (Roy, 1951; Rubin, 1974), we use Coarsened Exact Matching (CEM) (Iacus et al., 2011, 2012) to estimate the average treatment effect on the adoption of energy efficiency measures by organizations participating in the German energy audit programme. To allow for comparison of the CEM-based results with those of more traditional matching methods, we also apply Propensity Score Matching (PSM) and distance-based matching. The analyses were carried out for four ancillary energy efficiency technologies that are typically included in energy audits (Thollander et al., 2007; Fleiter et al., 2012b): lighting, thermal insulation of buildings, heating system exchange, and heating system optimization.

The paper is organized as follows. Section 2 provides a brief outline of the evaluation framework and introduces the estimation methods. Section 3 describes the surveys and provides descriptive statistics on the data. Section 4 presents the results. Section 5 summarizes the main findings and concludes the paper.

2. Methodology

We analyse the effect of audits on the adoption of energy efficiency measures using the common evaluation framework developed by Roy (1951) and Rubin (1974).[4] A binary treatment indicator D_i equals 1 if organization i participated in the audit programme and 0 otherwise. $Y_i(D_i)$ denotes the potential outcome of the adoption decision for i. The treatment effect for organization i can then be written as

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

In particular, we are interested in the effect of an audit on those organizations participating in the audit programme, i.e., in the Average effect of Treatment on the Treated (ATT):

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

Naturally, the counterfactual mean for those participating in the audit, i.e., $E[Y(0)/D = 1]$, is not observable (the fundamental problem of causal inferences). Since participation in the audit programme was not random, unobservable organization characteristics that affect the decision to participate in an audit programme might also affect the decision to adopt energy efficiency measures and lead to selection bias. To identify τ_{ATT} we therefore rely on the familiar Conditional Independence

---

[1] Existing analyses of energy audit programmes for private households provide mixed evidence for their effectiveness (e.g. Hirst and Goeltz, 1985; Frondel and Vance, 2013; Murphy, 2014; Alberini and Towe, 2016; Considine and Sapci, 2016).

[2] Using control groups for this kind of program evaluation is challenging since the real-world environment is less controllable than, for example, conditions in laboratory or healthcare experiments that involve clinical control groups (e.g. Vedung 1997). In particular, manufacturing companies are heterogeneous in terms of technologies, energy efficiency opportunities, and decision-making (DeCanio and Watkins, 1998; Fleiter et al., 2012c), rendering program evaluation challenging (Paramonova and Thollander, 2016).

[3] Since the costs of audits for travel, site-visits, and reporting are typically treated as overhead costs, audits are typically also more cost-efficient for organizations with higher energy use.

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

Assumption (CIA)[5]: conditional on the set of relevant (pre-treatment) covariates X, the potential outcomes [Y(1), Y(0)] are independent of assignment to the programme. We further assume that the effect of an audit on each organization is not affected by the participation of other organizations in the audit programme (no interference). Similarly, the audits are assumed to be comparable across organizations (no variation in treatment). Finally, the common support (or overlap) assumption requires covariate values where both audit and control group observations are found. We may then use matching estimators to estimate τ_{ATT} (e.g., Abadie and Imbens, 2006, 2011, 2016). Thus, we rely on data from organizations that did not participate in the audit programme but which exhibit relevant characteristics that are similar to those of members of the audit group. Any observable difference in the adoption of energy efficiency measures between organizations in the audit group and those in the control group can then be attributed to the audit programme.[6] Matching estimators for τ_{ATT} are generally constructed as the weighted average over the outcomes for nonparticipants:

$$\hat{\tau}_{ATT} = \frac{1}{N_T} \Sigma_{i \in \{D_i=1\}} \left[Y_i(1) - \Sigma_{j \in \{D_j=0\}} w_{ij} Y_j(0) \right] \tag{3}$$

where N_T is the number of participants in the treatment group and w_{ij} is a weighting function that, for each i, determines the weights attached to the outcomes of the matches j in the control group. The w_{ij} sum to 1 for each i. In essence, the sample means in (3) replace the expectations operator in (2). Ultimately, the objective of matching is to "prune" the data to attain treatment and control group samples which are balanced, i.e., which exhibit equivalent distributions of the covariates (a balancing condition). However, conditioning on X by exactly matching on the covariates fails for practical reasons if X is large or includes continuous covariates.

In this paper we employ estimators that rely on two types of matching methods: CEM, which has only recently been developed (Iacus et al., 2011, 2012)[7] and–for robustness checks–more familiar matching estimators such as PSM algorithms (Rosenbaum and Rubin, 1983) and distance-based approaches (Abadie and Imbens, 2006, 2011).

The CEM algorithm first transforms the continuous variables into strata, which enable exact matching of the treatment and control groups. Employing appropriate weights, CEM ensures that the balancing condition with respect to X holds. A weight of zero is attached to unmatched observations. A matched audit group member receives a weight of 1. The weight of a matched control group member reflects the relative frequency of audit to control group members in the control group member's stratum, compared with the relative frequency of the total number of audit to control group members in the matched sample. Employing these weights for the matched subsample and using the original (rather than the coarsened) X, the ATT can be estimated, e.g., via regression analysis. CEM has been found to outperform PSM and distance-based matching estimators with respect to bias, variance, and the balancing condition (see Iacus et al., 2011, 2012). If CEM involves many strata, however, it may be difficult to find matches, leading to a small post-matching subsample by reference to which we would estimate the ATT.

The PSM-based estimators require, first, running a binary response model that regresses audit participation on X (e.g. Caliendo and Kopeinig, 2008). Based on the resulting coefficients, audit participation is predicted for each organization i. These propensity scores then are used to identify the organizations in the control group that best match the organizations in the audit group. In this sense, the PSM aggregates the information in X into a single index. Thus, the PSM-based estimator in (3) is simply the mean difference in outcomes over the common support. Depending on the weighting function w_{ij}, various PSM methods can be applied. In this study we use the popular *nearest neighbour* method, which selects the organization in the control group with the closest propensity score, i.e., $w_{ij} = 1$ (e.g., Heckman et al., 1997). We also deploy multiple neighbours, where each neighbour is weighted equally. Using multiple neighbours typically leads to lower standard errors of $\hat{\tau}_{ATT}$ but may increase bias. In addition, we use the Mahalanobis measure of distance, which selects matches based on 'similar' values for the observable covariates (rather than for the propensity scores).[8] In practice, however, the balancing condition is often violated for the propensity score and distance-based matching procedures.

In conclusion, differences in the first stage between the CEM matching and PSM or distance-based matching may lead to very different post-matching subsamples by reference to which the ATT in the second stage is estimated. Thus, our results may be sensitive to model selection.

3. Data

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

---

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

[6] In contrast to regression analysis, matching estimators do not rely on a functional form for $E(\tau/.)$ in Eq. (2).

[7] The few applications of CEM in the energy economics literature include Alberini and Towe (2015) and Asensio and Delmas (2017).

[8] The Mahalanobis distance calculates the Euclidean distance of the covariates for treatment and control group members, using standardized covariates and adjusting for covariance.

[9] Including this set of identical questions was feasible since the authors were involved in both surveys through participation in projects funded by the Federal German Ministry of Economics (IREES/Fraunhofer ISI, 2014; Schlomann et al., 2015). The project reports provide copies of the questionnaires (in German).

3.1. Audit group

The treatment group data include responses from organizations participating voluntarily in an energy audit programme (IREES/Fraunhofer ISI, 2014). The energy audits are subsidized by the German energy audit programme for SMEs, which was launched in 2008 as 'Sonderfonds Energieeffizienz in KMU' and ran until March 2012 as 'Energieberatung Mittelstand'. The programme was designed to overcome information barriers via high-quality energy audits. Eligible organizations were SMEs with a maximum of 250 employees in manufacturing and non-manufacturing sectors. Thus, the German audit programme differs from the US Industrial Assessment Centers (IAC) program (Anderson and Newell, 2004), which is limited to manufacturing sector firms with fewer than 500 employees. Therefore, a relatively higher number of the recommended measures in the German energy audit programme relate to buildings rather than to industrial production processes. Like the US IAC program, however, the German energy audit programme is not part of any broader energy efficiency programme such as a voluntary agreement or energy management obligations.

Independent professional energy auditors carry out audits funded by the German programme on site. Grants are provided for both simple and detailed audits. Simple audits are carried out within a maximum of two days and focus on identifying major energy savings potentials and measurements at the audited sites. They are geared especially towards organizations with low energy use and relatively simple energy systems. Detailed audits are funded for up to ten days and include more elaborate analysis, in-depth monitoring, and detailed action plans and recommendations. Simple audits are subsidized at up to 80% of the daily rate of the energy auditor to a maximum of 640 euros/day for up to two days. For detailed audits, funding is granted for 60% of the daily rate of the audit to a maximum of 480 euros/day for up to ten days. Thus, the maximum funding for a simple audit is 1280 euros and for a detailed audit it is 4800 euros. The audit program started in 2008; through the end of 2013, more than 24,300 organizations had received funding via the programme.[10] Fleiter et al. (2012a) provide a more detailed discussion of the German audit programme.

The data used for our analysis are based on a recent evaluation of the German audit programme (IREES/Fraunhofer ISI, 2014), which included an online survey of about 9200 organizations that had received funding under the programme and for which contact details were available. This survey asked about the characteristics of the funded organizations and perceptions of the funding programme, the audits, and the auditors, and included detailed questions on specific energy efficiency measures. The survey was open for five weeks in April/May 2014 and yielded 1523 completed questionnaires, corresponding to a response rate of about 17%. The responses represent the structure of the population of funded organizations in terms of organization size and sector distribution. For each observation, the data were validated and checked for plausibility and internal consistency prior to further analysis.[11] This left a sample of 1471 observations for the audit group, approximately 66% of which had been involved in a simple audit, 14% in a detailed audit, and 20% in both types of audits. The four ancillary energy efficiency measures considered in this study are generally included in simple audits. Our empirical analysis distinguishes between organizations that had carried out only simple audits and organizations that had carried out either detailed audits or both types of audits.

3.2. Control group

The control group data are based on the 2440 responses of organizations participating in a representative energy use survey in the German combined 'trade, commerce, services and other consumers' sector.[12] This survey was administered by the market research company GfK SE between February and July 2014 (Schlomann et al., 2015). In each organization, either the energy manager or the person responsible for energy management was interviewed by trained interviewees via Computer Assisted Personal Interviews. The survey gathered general information on organization size, structure, and energy costs, and specific information on technologies and energy consumption for various end uses. Most importantly, it included a separate part comprised of identical questions pertaining to the four ancillary energy efficiency measures included in the survey of the audit group. Since the combined 'trade, commerce, services and other consumers' sector was defined in the same way for the survey and the German national energy balances, it includes non-manufacturing organizations in areas such as public and private services and trade, agriculture and construction, and also manufacturing organizations such as manufacturing enterprises with fewer than 20 employees. By design, most organizations in the control group are quite small; less than 2% have more than 50 employees. To allow for comparable data sets in the matching analysis, we restricted

---

[10] Some funding programs such as those provided by the German Bank for Reconstruction (KfW − Kreditanstalt für Wiederaufbau) offer low-interest loans for organizations that had carried out energy audits.

[11] Implausible data include "zeros" for energy savings, energy costs, or investment, energy consumption below 1000 kWh/year, or an audit duration exceeding 30 days. Data entries were considered inconsistent, for example, when respondents stated that energy saving measures had been implemented but the number of implemented measures was reported as zero.

[12] The German energy balances distinguish the four end-use sectors: industry, private households, transportation, and the combined 'trade, commerce, services and other consumers' sector.

observations in the audit and the control groups to small organizations with a maximum of 50 employees.[13] Thus, the subsequent findings apply to organizations of this size and should be extrapolated with some caution.

3.3. Energy efficiency measures

Both surveys included questions on the adoption of four categories of common, crosscutting, ancillary energy efficiency measures: (i) installation of efficient lighting (*lighting*) includes replacement of lighting systems with more efficient alternatives such as T5 fluorescent lamps or LEDs; (ii) insulation of the building envelope (*insulation*) may involve insulation of the roof and outer walls, and more effective window insulation; (iii) replacement of the heating system (*heating*) entails discarding the existing system and installing a more efficient one, such as a condensing boiler; (iv) optimization of heating system operations (*operations*) involves energetic optimization of a heating system such as through hydraulic adjustments, night-time turndown, dynamic control, or temperature reduction. The respondents in both groups were asked whether their organizations had adopted any of these measures since 2008.[14] To reduce the length of the questionnaire, respondents were asked for further details about only two, randomly chosen, measures.

These measures differ in terms of investment costs, energy cost savings, and payback times. In our combined sample, the average reported investment costs were about 5400 euros for *lighting* (N = 521), 29,000 euros for *insulation* (N = 165), 30,800 euros for *heating* (N = 197), and 18,000 euros for *operations* (N = 147). On average, the lighting measures adopted saved 25% of an organization's electricity use for lighting (N = 255), while in the case of energy use for heating, insulation measures saved 18% (N = 53), exchanging the heating system saved 23% (N = 87), and optimization of the heating system saved about 15% (N = 74) of heating costs, respectively.[15] Finally, average payback times were 5 years for *lighting* (N = 97), 9.2 years for *insulation* (N = 13), 6.9 years for *heating* (N = 43), and 6.6 years for *operations* (N = 30). Thus, the reported payback times of the energy efficiency measures considered in our sample are higher than those reported in Anderson and Newell (2004).[16]

Audit group organizations were asked also whether they had planned to implement a particular measure prior to participating in the audit programme. Observations of organizations responding positively were excluded from the estimation of the ATT for this particular measure to limit the effects of free riding on estimates of audit effectiveness. This resulted in our dropping 13% of the audit group organizations for *lighting*, 9% for *insulation*, 11% for *heating* and 8% for *operations*.

3.4. Covariates

Both surveys collected information on variables which the empirical literature suggests are correlated with the adoption of energy efficiency measures in industry (e.g., Velthuijsen 1993; de Groot et al., 2001; Thollander and Palm, 2013) and in the combined commerce and services sector (e.g. Schleich and Gruber, 2008; Schleich, 2009; Olsthoorn et al., 2017). For the PSM and distance-based matching algorithms we assume that the same set of variables is relevant for programme participation.

Our covariates include the share of energy costs of total costs (*energy cost share*) to capture both a company's financial incentive to invest in energy efficiency and the strategic importance of energy to a given organization.[17,18] *Number of employees* should capture the effect of organization size. Adoption propensity is expected to be higher for larger organizations since they have more resources available to acquire technological and financial know-how, and may more easily overcome barriers to adoption related to information costs and other transaction costs. In addition, larger firms may be better able to spread the risk of technology adoption and may find it easier to acquire external funding. Organizations with an *energy manager* (as organizational units) are more likely to adopt energy efficiency measures because an energy manager's responsibilities typically include controlling and optimizing energy costs. Also, including the formal position of energy manager may reflect the priority an organization has attached to energy use and costs. The form of organization ownership also can matter for technology adoption, e.g., know-how and information may spill over from a mother organization to a subsidiary. *Rented*

---

[13] According to the European Union definition, "small enterprises" are enterprises with fewer than 50 employees and an annual turnover of balance sheet total of at most 10 million euros. Using 49 employees rather than 50 employees as the cutoff point slightly restricts the degrees of freedom, but does not alter the findings of our analyses.

[14] We have no information indicating when the measure was implemented. Thus, some measures reported in the audit group may have been implemented before the audit. Under the CIA, this does not affect the validity of the matching analysis.

[15] The figures are calculated based on the information provided by audit group companies involved in an initial energy audit only and employing a maximum of 50 workers. Many companies, though, did not report this information.

[16] Thus, the technologies (and sectors) considered in this study are related to buildings only and differ, for example, from those considered in Anderson and Newell (2004). The latter rely on data provided by the US IAC, which includes companies from the manufacturing sector only. Compared with those in our sample, the organizations studied in Anderson and Newell (2004) are larger and the technologies include a variety of non-building-related measures such as electric motors.

[17] We considered including energy prices. In particular, we calculated electricity prices as the ratio of electricity expenditure to electricity use. However, due to loss of observations, we ultimately did not use electricity or other energy prices in the subsequent analyses.

[18] Note that, strictly speaking, our measure of energy intensity may be affected by audit participation and, hence, may violate the CIA assumption. However, for the sample at hand, we believe ex-post matching is not a problem. First, the effects of the measures considered on the energy costs of an organization are typically rather small. Based on organization responses we find that the energy savings of individual measures range from 1% (25% quantile) to 8% (75% quantile) of an organization's energy use. A typical energy cost share of 5% in our sample would be lowered by about 0.25 percentage points. Second, our findings turn out to be robust to excluding the energy costs share in the set of matching variables (see results for *CEM_no_energy* and *CEM_sectors_only* in Table 3).

Table 1
Sample means of measure adoption and covariates (simple audit).

Covariates	Group means			Difference audit vs. control group
	All	Audit	Control	
Lighting (dummy) (N = 1034)	0.409	0.537	0.318	0.219***
Insulation (dummy) (N = 1001)	0.130	0.193	0.089	0.104***
Heating (dummy) (N = 907)	0.139	0.213	0.086	0.127***
Operations (dummy) (N = 937)	0.192	0.338	0.084	0.255***
Energy cost share†	0.100	0.086	0.111	−0.025***
Employees†	12.75	20.60	7.20	13.40***
Emanager (dummy)†	0.070	0.079	0.066	0.013
Subsidiary (dummy)†	0.099	0.073	0.117	−0.045**
Rented (dummy)†	0.517	0.432	0.578	−0.145***
Hospitality (dummy)	0.098	0.108	0.091	0.017
Other services (dummy)†	0.283	0.259	0.300	−0.041
Trade (dummy)†	0.254	0.231	0.270	−0.039
Food and beverages (dummy) †	0.064	0.051	0.073	−0.021
Metal (dummy)†	0.069	0.121	0.031	0.090***
Other production (dummy)†	0.232	0.229	0.234	−0.005

***Significant at p < 0.01, **significant at p < 0.05 and *significant at p < 0.1 in a *t*-test for group means.
†Statistics reported for *lighting* subsample.

buildings are less likely to be equipped with energy-efficient technologies because, typically, the property owner makes investment decisions regarding capital-intensive measures such as space and water heating and insulation, and tenants appropriate the benefits. Finally, to capture heterogeneity across sectors, we included six sector dummies. These sector dummies were constructed by aggregating sub-sectors with similar technologies and energy-use patterns (see Table A1 in Appendix A). Thus, the survey questionnaires generated a broad set of covariates.

Table 1 provides the descriptive statistics for the outcome variables and the covariates in the final (simple) audit and control group samples.[19] The differences for all four outcome variables are statistically significant. Simple comparison of adoption rates between groups suggests that adoption shares of all four energy efficiency measures are significantly larger in the simple audit group than in the control group. In addition, the differences in adoption shares across measures reflect differences in lifetimes. Of the four technologies, lighting systems are being replaced the most frequently. Table 1 shows further that the (average) values of most non-sector covariates differ significantly between the audit and control groups. Compared with the control group, audit group organizations are, on average, somewhat less energy intensive, distinctly larger, less likely to be subsidiaries, and less likely to rent their buildings. Audit group organizations are also more likely to belong to the metals sector.

4. Results

We first present our CEM-based findings for the effectiveness of simple audits. That is, probit models on energy efficiency technology adoption were estimated using the weights calculated via CEM.[20] As robustness checks, we also report results for simple audits employing PSM and Mahalanobis distance-based matching. We then show the findings for detailed audits using CEM. Finally, we explore whether audit effectiveness depends on organization size.

4.1. CEM results for simple audits

We employ three types of CEM models, which differ in relation to the number of matching variables (see Table 2). In the first CEM model (*CEM_all*), we use all the covariates listed in Table 1.[21] This is the most restrictive model and matched approximately 43% of the organizations in the simple audit group. Annex

Table A2 in Appendix A suggests that almost all the covariates are perfectly matched. Only for *energy cost share* does a slight difference exist between the audit and the control group means.

Table 3 presents the results of estimating the probit models. The findings suggest (under CIA) that simple energy audits increase adoption of all four measures considered. Table 4 reports the average marginal effects, i.e., the ATT. It appears that a simple energy audit increases adoption of *lighting* by about 17 percentage points, of *insulation* by about 12 percentage points, of *heating* by about 11 percentage points, and of *operations* by about 13 percentage points. Based on the weighted means of

---

[19] Sample size varies slightly across the four measures mainly because the number of organizations that had already planned adoption differs across measures. In addition, there are small differences in the number of missing values across the adopted measures.

[20] This procedure implicitly assumes that the choice of audit type does not involve selection bias, i.e., unobserved factors related to audit choice also affect adoption.

[21] Actual implementation uses the natural logarithm of employees rather than the levels.

Table 2
Overview CEM procedures (simple audits).

	CEM_all	CEM_no_energy	CEM_sectors_only
Matching variables	energy cost share, employees (log), emanager, subsidiary, rented, hospitality, other services, trade, food and beverages, metal, other production	employees (log), emanager, subsidiary, rented, hospitality, other services, trade, food and beverages, metal, other production	hospitality, other services, trade, food and beverages, metal, other production
Number of matched audit group organizations (*lighting*)	212 (50%)	258 (60%)	428 (100%)
Number of observations in final probit model (*lighting*)	480 (46%)	722 (70%)	1034 (100%)

Table 3
Results of probit model for *CEM_all* for simple audits (robust standard errors in parentheses).

Variables	Lighting	Insulation	Heating	Operations
Simple audit (dummy)	0.482***	0.589***	0.552***	0.462**
	(0.158)	(0.206)	(0.184)	(0.211)
Energy cost share	−0.616	0.312	2.070	1.867
	(1.289)	(1.630)	(1.563)	(1.478)
Employees (log)	0.0191	0.0724	0.162	0.176
	(0.0917)	(0.107)	(0.122)	(0.123)
Emanager	1.005***	−0.324	0.173	−0.0319
	(0.371)	(0.539)	(0.452)	(0.401)
Subsidiary	−0.175	0.178	0.324	−0.462
	(0.315)	(0.582)	(0.597)	(0.589)
Rented	0.00579	−0.797***	−0.499**	−0.154
	(0.170)	(0.191)	(0.205)	(0.215)
Hospitality	0.562**	−0.201	−0.202	0.203
	(0.274)	(0.294)	(0.317)	(0.326)
Trade	−0.426**	−0.445*	−0.0109	−0.199
	(0.208)	(0.268)	(0.244)	(0.258)
Food and beverages	0.547	−0.624	−0.645	−0.655
	(0.444)	(0.598)	(0.529)	(0.429)
Metal	−0.576	−1.007***	−0.441	0.0907
	(0.380)	(0.384)	(0.362)	(0.367)
Other production	−0.380	−0.234	−0.315	−0.0103
	(0.235)	(0.267)	(0.283)	(0.256)
Constant	−0.252	−1.101***	−1.582***	−1.374***
	(0.272)	(0.324)	(0.352)	(0.335)
Loglikelihood	−301.7	−170.5	−158.1	−205.2
Wald Chi^2(10)	34.53***	45.31***	23.41***	19.52***
Pseudo R^2 (McFadden)	0.0833	0.120	0.0791	0.0543
N	480	478	433	397

***Significant at $p < 0.01$, **significant at $p < 0.05$, and *significant at $p < 0.1$.

adoption for the matched audit group organizations, these figures translate into adoption rates that are approximately 50% higher for *lighting* and *heating*, about 100% higher for *insulation*, and about 150% higher for *operations.*

In *CEM_no_energy*, we used all covariates except energy cost share, which turned out to be the most difficult covariate to match in *CEM_all*. This increased the number of matched audit group organizations for the *lighting* sample from 50% in *CEM_all* to 60% in *CEM_no_energy* and resulted in quite similar ATT estimates for *lighting* and *insulation*. However, the estimation of ATT in *CEM_no_energy* compared with that in *CEM_all* is lower for *heating* and higher for *operations*. Interestingly, the weighted means of all the covariates are almost perfectly balanced across the simple audit and control groups, even for *energy cost share* (see Appendix Table A2).

Finally, we employ a rather unrestrictive CEM, which matches on sector dummies only. Under *CEM_sectors_only*, all the organizations in the simple audit group are matched; the estimates of ATT for *lighting*, *insulation*, and *heating* are similar to those derived from *CEM_all*, but the estimate of the ATT for *operations* is substantially higher. Appendix Table A2 suggests that the weighted means of all but the sector dummy covariates differ considerably across the simple audit and control groups.

4.2. Robustness checks

We also estimated the ATT based on PSM and Mahalanobis distance-based matching estimators, employing the full set of covariates and allowing for alternative numbers of nearest neighbours. Appendix Table A3 summarizes the findings. The

Table 4
CEM results for ATT of simple audits (robust standard errors in parentheses).

Measures	CEM_all	CEM_no_energy	CEM_sectors_only
Lighting	0.173***	0.181***	0.218***
	(0.053)	(0.049)	(0.028)
	N = 480	N = 722	N = 1034
Insulation	0.116***	0.095***	0.095***
	(0.036)	(0.032)	(0.021)
	N = 478	N = 726	N = 1001
Heating	0.111***	0.067*	0.124***
	(0.035)	(0.035)	(0.022)
	N = 433	N = 619	N = 907
Operations	0.134**	0.189***	0.237***
	(0.055)	(0.030)	(0.022)
	N = 397	N = 597	N = 937

***Significant at $p < 0.01$, **significant at $p < 0.05$, and *significant at $p < 0.1$.

Table 5
CEM results for ATT of detailed audits (robust standard errors in parentheses).

Measures	CEM_no_energy
Lighting	0.184***
	(0.065)
	N = 317
Insulation	0.163***
	(0.038)
	N = 302
Heating	0.112**
	(0.051)
	N = 264
Operations	0.221**
	(0.050)
	N = 276

***Significant at $p < 0.01$ and **significant at $p < 0.05$.

estimated ATTs based on the PSM and the Mahalanobis distance estimators vary somewhat from the number of nearest neighbours and are mostly quite similar across both methods. For *operations,* though, the ATT of the PSM estimator appears slightly higher than the ATT of the Mahalanobis distance estimator. In general, the findings are quite similar to those found using the CEM estimator for *CEM_all* in

Table 4 even though the samples are different. The only exception might be the ATT for *operations*, which is much smaller for *CEM_all* than for the PSM or Mahalanobis distance estimators. Nevertheless, the balancing assumption was found to be violated for all the PSM models presented in Table A3 in Appendix A. After conditioning on the propensity score there is a small but statistically significant systematic difference for *energy cost share* and *employees* (but not for any of the other covariates), between the audit and control groups.

4.3. CEM results for detailed audits

Only about one-third of the organisations in the audit group had carried out detailed energy audits. Thus, estimating the effectiveness of detailed energy audits involves far fewer observations than estimating the effectiveness of simple audits, undermining the reliability of the estimates. This holds in particular for *CEM_all*, the most restrictive model, where on average across the four measures only 58 detailed audit group members (i.e., about 39% of all members of the detailed audit group) could be matched. Table 5 therefore reports the findings for *CEM_no_energy*, where the sample size is much larger. As in the case of simple audits, for *CEM_no_energy* (but not for *CEM_sectors_only*) the weighted means of all the covariates (including also *energy cost share*) are nearly perfectly balanced across the detailed audit and control groups. Comparing Tables 4 and 5 suggests that the effectiveness of detailed audits and simple audits is generally quite similar for the four measures considered. Arguably, detailed energy audits may be more effective than simple audits for *insulation.*

4.4. Effects of organization size on audit effectiveness

We also investigate whether audit effectiveness varies with organization size. For example, audits may be less effective for larger organizations because information barriers are lower. Conversely, audit effectiveness may be higher for larger organizations because they have better in-house knowledge and can more easily mobilize the financial resources required to implement an audit's investment recommendations.
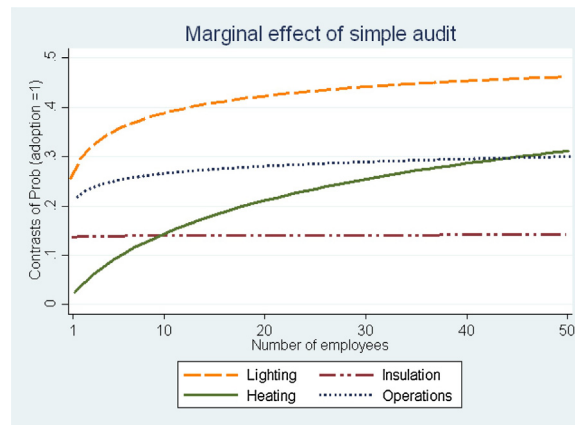
Fig. 1. Effectiveness of simple energy audits by organization size.

To explore whether the effectiveness of energy audits varies by organization size, we include an interaction term (employees (log)*simple audit) in the probit regression for *CEM_all*. As noted by Ai and Norton (2003) and further elaborated by Greene (2010), in the case of probit (and logit) models this interaction term does not capture the true interaction effect, which—because of the non-linearity—depends also on the levels of all the other variables. Following Greene (2010), we employ a graphical analysis, thereby assuming the median value of zero in the audit group sample for *emanager*, *subsidiary* and *rented*, the mean of *energy costs*, and the modes of the sector dummies (i.e., other production).[22] Fig. 1 suggests that the effectiveness of the audit increases up to a certain level with organization size for *lighting*, *heating*, and *operations* and then remains constant. In contrast, for *insulation* audit effectiveness does not appear to vary with organization size.[23]

5. Conclusions

Employing coarsened exact matching, we found that the German energy audit programme accelerated adoption of energy efficiency measures in small business organizations. While the few earlier studies evaluating the effects of energy audits on organizations' adoption of energy efficient technologies rely on the subjective responses of participating organizations and do not employ control groups, our findings provide more robust evidence pertaining to the effectiveness of energy audits.

The percentage increase in adoption in response to energy audits varied across the four types of ancillary measures analysed. The effect was strongest for *insulation*, i.e., the measure with the highest upfront costs, and for *operations*, arguably the most likely of the four measures to be overlooked by non-energy experts. Since these measures are related also to particularly high energy savings, our findings suggest that cost-efficient energy audits in small organizations should focus on—among the measures considered in this study—insulation measures and measures optimizing heating system operations. On the other hand, evaluations based on 'the number of additional measures induced by an energy audit' as an indicator of programme effectiveness (e.g., Fleiter et al., 2012b) could be misleading since, implicitly, they assume audits to be equally effective across measures.

In general, for the four measures considered we find no strong evidence that detailed audits are more effective than simple audits. This result may be explained by the fact that these measures are relatively simple ancillary measures, which are typically identified and analysed in simple audits. For more complex core-production measures, we expect detailed audits to be more effective than simple audits.

Our findings for small business organizations with up to 50 employees suggest that energy audits are more effective for lighting and insulation measures and for measures optimizing heating systems in the case of larger organizations. Arguably, the ability to follow up on energy audit recommendations is more robust for larger organizations because they have better in-house knowledge and the financial resources needed to implement an audit's investment recommendations. However, this interaction effect between organization size and audit effectiveness appears to vary by measure and to diminish with organization size. In sum, for the four cross-cutting measures considered we found no evidence that energy audits are less effective for larger organizations. In this sense, the EU Energy Efficiency Directive 2012/27EU, which obliges non-SMEs to carry out regular energy audits, is likely to be effective. Since our analysis was carried out for organizations with up to 50 employees, this conclusion should be treated cautiously.

---

[22] Qualitatively, the findings are robust to other combinations of variable values.

[23] To corroborate these findings we used the operators implemented in Stata to specify factorial interactions to formally test for interaction effects. These tests support our findings from the graphical analysis for all four measures at $p < 0.01$.

The CEM algorithm applied produced robust results under various sets of covariates and varying numbers of matched audit group organizations. These findings are consistent also with the findings obtained using conventional PSM and Mahalanobis distance-based matching, although the balancing condition could not be met for these procedures.

Finally, while our analysis sheds some critical light on the effectiveness of the German energy audit programme and the findings appear to be robust across the models employed, we cannot rule out the possibility that self-selection leads to upward bias in our estimates of audit effects. Unobserved factors may have affected both the propensity to participate in an energy audit and the propensity to adopt energy efficiency measures. While it would be difficult to implement in practice, true random assignment of organizations into treatment and control groups would avoid this potential, but untestable, source of bias. Thus, the magnitudes of the effects should be interpreted with some caution. Arguably, accounting for differences between audit and control groups is generally more relevant when evaluating programs which target core-production processes rather than—as is the case in our study—cross-cutting and ancillary measures.

Acknowledgements

Appendix A.

Table A1
Description of sectors.

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

Table A2
Weighted means of variables in the (simple) audit and control groups (for *lighting*).

Model	Variable	Audit group	Control group
CEM_all	Energy cost share[†]	0.056	0.051
	Employees (log)[†]	2.371	2.384
	Emanager[†]	0.042	0.042
	Subsidiary[†]	0.047	0.047
	Rented[†]	0.406	0.406
	Hospitality[†]	0.075	0.075
	Other services[†]	0.255	0.255
	Trade[†]	0.052	0.052
	Food and beverages[†]	0.108	0.108
	Metal[†]	0.198	0.198
	Other production[†]	0.311	0.311
CEM_no energy	Energy cost share	0.097	0.104
	Employees (log)[†]	2.281	2.279
	Emanager[†]	0.027	0.027
	Subsidiary[†]	0.201	0.074
	Rented[†]	0.415	0.415
	Hospitality[†]	0.105	0.105
	Other services[†]	0.221	0.221
	Trade[†]	0.031	0.031
	Food and beverages[†]	0.050	0.050
	Metal[†]	0.267	0.267
	Other production[†]	0.326	0.326
CEM_sectors_only	Energy cost share	0.109	0.086
	Employees (log)	1.559	2.696
	Emanager	0.059	0.079
	Subsidiary	0.100	0.072

Table A2 (*Continued*)

Model	Variable	Audit group	Control group
	Rented	0.547	0.432
	Hospitality†	0.107	0.107
	Other services†	0.259	0.259
	Trade†	0.231	0.231
	Food and beverages†	0.051	0.051
	Metal†	0.121	0.121
	Other production†	0.229	0.229

† Matching variable.
Results vary slightly across the four measures because the samples differ.

Table A3
Estimates of ATT using propensity scores and Mahalanobis distance-based matching.

	Number of nearest neighbours				
	1	4	8	12	N
Propensity score					
Lighting	0.166***	0.183***	0.168***	0.176***	1034
Insulation	0.092*	0.085**	0.079**	0.083**	1001
Heating	0.120**	0.089**	0.081**	0.075*	907
Operations	0.209***	0.240***	0.233***	0.237***	937
Mahalanobis distance					
Lighting	0.142**	0.191***	0.170***	0.165***	1034
Insulation	0.100**	0.095***	0.108***	0.106***	1001
Heating	0.105**	0.119***	0.106***	0.102***	885
Operations	0.133**	0.187***	0.167***	0.165***	887

***Significant at $p < 0.01$, **significant at $p < 0.05$, and *significant at $p < 0.1$.

References

Abadie, A., Imbens, G., 2006. Large sample properties of matching estimators for average treatment effects. Econometrica 74, 235–267.
Abadie, A., Imbens, G., 2011. Bias-corrected matching estimators for average treatment effects. J. Bus. Econ. Stat. 29, 1–11.
Abadie, A., Imbens, G., 2016. Matching on the estimated propensity score. Econometrica 84 (2), 781–807.
Ai, C., Norton, E., 2003. Interaction terms in logit and probit models. Econ. Lett. 80 (1), 123–129.
Alberini, A., Towe, C., 2015. Information v. energy efficiency incentives: evidence from residential electricity consumption in Maryland. Energy Econ. 52, S30–S40.
Anderson, S., Newell, R., 2004. Information programs for technology adoption: the case of energy-efficiency audits. Resour. Energy Econ. 26, 27–50.
Angrist, J.D., Pischke, J.S., 2009. Mostly Harmless Econometrics: an Empiricist's Companion. Princeton University Press Princeton, New Jersey.
Asensio, O., Delmas, M., 2017. The effectiveness of US energy efficiency building labels. Nat. Energy 2, 17033.
Backlund, S., Thollander, P., 2015. Impact after three years of the Swedish energy audit program. Energy 82, 54–60.
Caliendo, M., Kopeinig, S., 2008. Some practical guidance for the implementation of propensity score matching. J. Econ. Surv. 22 (1), 31–72.
Considine, T.J., Sapci, O., 2016. The effectiveness of home energy audits: a case study of Jackson, Wyoming. Resour. Energy Econ. 44, 52–70.
de Groot, H.L.F., Verhoef, E.T., Nijkamp, P., 2001. Energy saving by firms: decision making, barriers and policies. Energy Econ. 23 (6), 717–740.
DeCanio, S., Watkins, W., 1998. Investment in energy efficiency: do the Characteristics of firms matter? Rev. Econ. Stat. 80 (1), 95–107.
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 Effic. 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.
Fleiter, T., Hirzel, S., Worrell, E., 2012c. The characteristics of energy-efficiency measures – a neglected dimension. Energy Policy 51, 502–513.
Frondel, M., Vance, C., 2013. Heterogeneity in the effect of home energy audits: theory and evidence. Environ. Resour. Econ. 55, 407–418.
Greene, W., 2010. Testing hypotheses about interaction terms in nonlinear models. Econ. Lett. 107, 291–296.
Heckman, J., Ichimura, H., Todd, P., 1997. Matching as an econometric evaluation estimator: evidence from a job training program. Rev. Econ. Stud. 64, 605–654.
Hirst, E., Goeltz, R., 1985. Estimating energy savings due to conservation programmes: the BPA residential weatherization pilot programme. Energy Econ. 7, 20–28.
IREES/Fraunhofer ISI, 2014. Evaluation des Förderprogramms 'Energieberatung im Mittelstand' Schlussbericht Im Auftrag des Bundesministeriums für Wirtschaft und Energie. (Download: h https://www.bmwi.de/Redaktion/DE/Publikationen/Studien/evaluation-des-foerderprogramms-energieberatung-im-mittelstand.pdf?__blob=publicationFile&v=3).
Iacus, S.M., King, G., Porro, G., 2011. Multivariate matching methods that are monotonic imbalance bounding. J. Am. Stat. Assoc. 106 (493), 345–361.
Iacus, S.M., King, G., Porro, G., 2012. Causal inference without balance checking: coarsened exact matching. Polit. Anal. 20, 1–24.
Imbens, G.W., 2004. Non parametric estimation of treatment effects under exogeneity: a review. Rev. Econ. Stat. 86, 4–29.
Joskow, P.L., Marron, D.B., 1992. What does a negawatt really cost? Evidence from utility conservation programs. Energy J. 13, 41–73.
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.
Olsthoorn, M., Schleich, J., Hirzel, S., 2017. Adoption of energy efficiency measures for non-residential buildings: technological and organizational heterogeneity in the trade, commerce and services sector. Ecol. Econ. 136, 240–254.
Paramonova, S., Thollander, P., 2016. Ex-post impact and process evaluation of the Swedish energy audit policy programme for small and medium-sized enter-prises. J. Clean. Prod. 135, 932–949.
Price, L., Lu, H., 2011. Industrial energy auditing and assessments: a survey of programs around the world. Eceee 2011 Summer Study Proceedings, 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. Oxf. Econ. Pap. 3, 135–145.

Rubin, D., 1974. Estimating causal effects to treatments in randomised and nonrandomised studies. J. Educ. Psychol. 66, 688–701.

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

Schleich, J., 2004. Do energy audits help reduce barriers to energy efficiency? An empirical analysis for Germany. Int. J. Energy Technol. Policy 2, 226–239.

Schleich, J., 2009. Barriers to energy efficiency: a comparison across the German commercial and services sector. Ecol. Econ. 68, 2150–2159.

Schlomann, B., Wohlfarth, K., Kleeberger, H., Hardi, L., Geiger, B., Pich, A., and Roser, A., 2015. Energieverbrauch des Sektors Gewerbe, Handel, Dienstleistungen (GHD) in Deutschland für die Jahre 2011 bis 2013. Institut für Ressourceneffizienz und Energiestrategien (IREES), GfK, Fraunhofer ISI. Lehrstuhl für Energiewirtschaft und Anwendungstechnik Prof. Dr.-Ing. U. Wagner, Karlsruhe, München, Nürnberg.

Thollander, P., Palm, J., 2013. Improving Energy Efficiency in Industrial Energy Systems: An Interdisciplinary Perspective on Barriers, Energy Audits, Energy Management, Policies, and Programs. Springer, London, pp. 2013.

Thollander, P., Rohdin, P., Danestig, M., 2007. Energy policies for increased industrial energy efficiency: evaluation of a local energy programme for manufacturing SMEs. Energy Policy 35, 5774–5783.

Vedung, E., 1997. Public Policy and Program Evaluation. Transaction Publishers New Brunswick, New Jersey.

Velthuijsen, J., 1993. Incentives for investment in energy efficiency: an econometric evaluation and policy implications. Environ. Resour. Econ. 3, 153–169.

Wooldridge, J.M., 2010. Econometric Analysis of Cross Section and Panel Data, 2nd edition. MIT Press, Cambridge, Massachusetts.