Does energy labelling on residential housing cause energy savings?

Vibeke H. Kjaerbye
AKF
Denmark
vha@akf.dk

Keywords

energy labelling, energy efficiency, propensity score matching

Abstract

Danish households use more than 30% of the total amount of energy being used in Denmark. More than 80% of this energy is dedicated to space heating. The same relation is seen in many OECD countries. The corresponding energy savings potential was recently estimated at 30% of the energy used in buildings. Energy labelling is seen as an important instrument to target these potential energy savings. This paper evaluates the effects of the Danish Energy Labelling Scheme on energy consumption in existing single-family houses with propensity score matching using real metered natural gas consumption and a very wide range of register data describing the houses and households. The study did not find significant energy savings due to the Danish Energy Labelling Scheme, but more research would be needed to complement this conclusion.

Introduction

Since the 1970s focus in the OECD countries has been on diminishing energy consumption, both in relation to reduce CO_2 emissions but also in relation to reduce the need for import of energy from unstable countries. Danish households use more than 30% of the total amount of energy being used in Denmark. More than 80% of this energy is dedicated to space heating (see Danish Energy Statistics 2006). The same relation is seen in many OECD countries. A large potential for energy savings exists in the Danish residential building stock due to the fact that 75% of the buildings were constructed before 1979 when the first important requirements for energy performance of building were introduced in Denmark. Recently an energy saving potential in buildings has been estimated to 30-35% of the current amount of energy used for heating in buildings (Witchen 2009). Since the early 1990s energy labelling has been regarded as an essential method for improving energy efficiency and minimizing energy consumption (see Peréz-Lombard et al 2009).

In January 2003 the EPBD (Energy Performance of Buildings Directive) was introduced to the EU countries. January 2006 was the official deadline by which the 25 Member States had to tranpose the Directive into national law (see European Commision 2002). The purpose of the EPBD is to enhance energy efficiency in buildings. One of the demands to the Member States is that they introduce a scheme for energy labelling (also called energy audits, energy performance assessments) of both new and existing buildings upon sale or rent. The EPBD is to a certain extent inspired by the Danish Energy Labelling Scheme for small buildings, since Denmark is the pioneer on energy labelling in the EU and one of few countries with actual experience on energy labelling[1]. Denmark have been using energy labelling schemes as a feasible mechanism to achieving energy savings in existing buildings since 1997 and data from the scheme has been collected in a central database.

In the US, energy labelling is also used to achieve energy efficiency in buildings. A voluntary labelling scheme (RESNET[2]

1. Germany has had a labelling scheme since 1995, but has not collected data from the scheme centrally. Belgium has voluntary energy audits (see Thomsen et al 2006).

2. Residential Energy Service Network is a national standards making body for building energy efficiency rating systems. Buildings are rated and a set of rated recommendations for cost-effective improvements that can be achieved by the

) rated more than 165,000 new homes in 2006, representing approximately 10% of the homes built in US that year (see RESNET homepage 2008). The American RESNET standards have also been adopted in Canada and in the City of Shanghai, China.[3]

In 2000, an evaluation of the Danish Energy Labelling scheme was carried out by interviewing 300 owners of energy labelled houses and 300 owners of houses without an energy label. All house owners have recently bought their house. The evaluation found that only 43% of the interviewed houseowners knew or had heard about the energy labelling scheme. Further only half of the owners of labelled houses were familiar with the existence of a label for their house/flat. By comparing investment levels and achieved enery savings for labelled and non labelled houses, the evaluation concluded that differences were ambigious. The only significant differences was found in the nature of energy renovations carried out. The labelled houses tended to have made more technical demanding improvements whereas the non labelled houses had made more aesthetic improvements like changing windows. The evaluation concludes by recommending trying another evaluation approach using a significant larger dataset to estimate the effect of the energy labelling scheme (see Madsen, Ramlau and Pedersen 2001).

This paper will follow the recommendations from the evaluation just mentioned by trying another evaluation approach and using a larger dataset in order to determine whether *the Danish Energy Labelling Scheme for Small Buildings has caused significant energy savings*. The approach in this paper is strong compared to interview based approaches because we use actual consumption of energy (metered natural gas used for heating) and a very wide range of house and household characteristics obtained from administrative register data. We estimate the effect of energy labelling on what can actually be measured and not on what a selected group of house owners answers to a list of questions. This approach allows us to assume Strong Ignorability, i.e. we control for all confounding variables determining energy labelling and energy consumption. In contrast to other estimation techniques matching is able to control for selection bias without imposing a particular parametric model for energy consumption.

Building related energy conservation has been a priority in Danish energy policy during three decades and is becoming increasingly central in both the EU and the US. The results from this paper on the effect of an energy efficiency programme are expected to be useful in designing future policies in that respect.

rated building is also produced. The ratings are used for both new and existing homes. RESNET's standards are officially recognized by the U.S. mortgage industry for capitalizing a building's energy performance in the mortgage loan, certification of "White Tags" for private financial investors, and by the federal government for verification of building energy performance for such programs as federal tax incentives, the Environmental Protection Agency's ENERGY STAR program and the U.S. Department of Energy's Building America Program (see RESNET homepage 2008).

3. The World Bank estimates that by 2015, half of the world's new building constructions will take place in China, and more than half of China's urban residential and commercial stock will have been constructed after 2000 (see RESNET homepage 2008).

The paper is based on author's PhD work which is funded by the Danish Energy Research Programme (Danish Energy Agency)[4], RUC[5] and AKF[6].

Even though there is a growing interest in buildings' energy improvements, green building ratings and eco-labelling (see e.g. Building Energy Labelling Forum Agenda 2006) there are not many evaluations of labelling schemes for buildings as tools to develop more energy efficiency in buildings. This paper contributes to the scarce empirical literature by evaluating the effect of the Danish Energy Labelling Scheme on single-family houses with propensity score matching.

The rest of this paper is organised as follows. Next section describes the Danish Energy Labelling Scheme, an earlier evaluation of the scheme and relevant literature. The data section provides a description of the data used in the analysis, highlighting the important differences between labelled and non-labelled houses and among the different categories of labelled houses. The evaluation section describes the evaluation approach chosen in this paper based on the treatment effects model of Rubin (1974), and the particular estimator employed; a kernel propensity score matching estimator. The main empirical findings are discussed in the results section. The paper ends with a number of concluding remarks.

The Danish Energy Labelling Scheme

In Denmark we have two instruments to target energy efficiency in buildings. All new buildings have to meet the requirements from the building regulations (BR) and all existing buildings have to get an energy labelled in relation to a sale. In 2008 less than 4% of the total stock of single family houses[7] were built after 2000 and therefore tightenings in the BR requirements for new buildings can only result in minor energy savings (and CO_2 reductions). In the large population of existing houses significant energy saving potential have been detected. An instrument for achieving these energy savings is energy labelling. Energy labelling has been mandatory in Denmark since 1997. In relation to a sale all existing buildings used for residential, public, trade or private services purposes must get an energy label. Different schemes targets large and small buildings. The background for this energy labelling scheme was to simplify and replace the former legislations, which had been evaluated as being confused and having questionable effect (See Madsen, Ramlau and Pedersen 2001). This paper concentrates on the effect of the Energy Labelling Scheme for Small Buildings on single-family houses. The Danish Energy Labelling Scheme was revised in 2006 to live up to the demands in the EPBD. This revision happened outside the period of data material available for this evaluation. The new labelling scheme is not significantly different from the one evaluated in this paper. The most significant difference is that the EPDB requires energy labelling for all houses both new built houses and existing ones in relation to a sale. The EPBD also requires flats to be energy labelled

4. Energi Forsknings Programmet, Energistyrelsen

5. Roskilde University Center, Denmark

6. Danish Institute of Governmental Research

7. The total stock of single family houses in 2008 was 1,077,000 and the number of single family houses constructed later than 2000 was 39,484 (see Statistics Denmark 2008).

upon sale. And the EPBD has introduced another method for calculating the label. None of these changes are expected to make a difference in relation to the effect of labelling scheme evaluated here.

It is the house sellers obligation to get his house energy labelled, when the house is put up for sale. All costs of the energy labelling are paid by the house seller, including the energy audit and the necessary calculations. The typical price of a single-family house energy label is about 300-500 Euro.[8] Energy labelling is carried out by an approved energy consultant, who must have at least five years of documented relevant experience in building technology and energy consultancy. The house seller provides the necessary information. The main objective with the scheme is to initiate energy and water savings in the housing stock by adressing consumers with factual information showing new owners or potential buyers that energy costs will account for a large part of the future costs. So far, the energy label has only provided information. Recently energy labelling schemes emerging in other EU countries seems to combine labelling with e.g. soft loans, grants or tax reductions. This has also been discussed in Denmark, but has not been implemented yet.

The energy label consists of two parts; an energy label category and energy plan. The label contains a standardised energy rating of the building containing information about the state of respectively heat, electricity and water installations and subsequently CO_2 emission impact. The rating is based on objective criterias, that is the energy consumption of the house is calculated given the characteristics of the building and its energy state, but independent of resident behaviour. Calculated energy consumption is compared to energy consumption of other similar buildings, and subsequently the calculated consumption is placed on a scale from A1 to C5[9] (A: low, B: middle and C: high). Houses in the different labelling categories are quite different, not only in terms of proposed profitable saving possibilities, but also in terms of house characteristics, and this paper will evaluate separately the average effects of energy label for the three main labelling categories, A, B and C.

The energy plan includes a proposal for profitable saving possibilities for all types of energy and water consumption facilities of the building. Furthermore, the energy plan includes estimates of necessary investments and annual savings of the proposal. The plan also reports the estimated technical lifetime of the proposal and provides the necessary details for calculating how profitable the individual proposals are under a given financing. A proposal is defined as profitable if: ((annual savings in DKK * estimated technical lifetime)/estimated necessary investment in DKK) >1.33. This part of the label gives the potential buyer information about the house that would have been hidden for him if the energy labelling scheme had not existed. Only if the new house owner carries out some or all of the proposed profitable energy saving improvements, the labelling scheme might induce future energy savings. All information used to calculate the energy label and propose profitable saving possibilities are registered in a central energy labelling database. But any improvements carried out by the new owner are not registered anywhere.

As expenses for heating normally make up the largest part of the expenses for energy consumption in a household[10], the energy labelling scheme has special focus on improvements that can reduce these expenses. Madsen, Ramlau and Pedersen (2001) examined the proposed profitable saving possibilities registered in the Energy Labelling Database and concluded that 95% of the recommended investments in energy saving improvements are related to reducing energy consumption for heating.

The energy labelling scheme is mandatory to existing houses in relation to a sale, but there are no significant consequences when the labelling is avoided. In the studied period, only 50-60% of the potential buildings were labelled when they were sold with important geographic differences in terms of the coverage of the label. Buildings in Greater Copenhagen and Funen are registered with coverage on more than 85%, while buildings in Northern Jutland are underrepresented with coverage on only 15-25% (see Madsen, Ramlau and Pedersen 2001).

According to the Danish Energy Authority 45,000-50,000 labels are issued each year. In total, more than 300,000 buildings corresponding to nearly 20% of all single-family houses in Denmark have got an energy label in the first 6½ years of the scheme. In 2002 the total annual costs of the energy labelling scheme amounted to more than 20 million Euro[11] (see Danish Energy Authority). Through the energy labelling scheme, energy savings for more than 130 million Euro were identified in 2001. The result of implementing all the possible savings would reduce the annual consumer energy cost by almost 20 million Euro. On average, the single-family houses could lower their energy costs by about 20%.

Data

To be able to estimate an effect of the energy labelling scheme in terms of differences in energy consumption between labelled and non labelled houses, we need to define what affects energy consumption for heating in single family houses. Among the important factors is the insulation standard of the house, which can be described by the year of construction, since the construction year determines which building regulation demands the insulation standard has to meet. Also house size has a significant influence on energy consumption. The bigger house the bigger heat consumption. Building type is also important since a single floor house has a larger surface and hence uses more energy for heating, than multi floor house or a terraced house. Such information can be all be found in Danish registers. Other factors like preferences for indoor temperature also plays a significant role in energy consumption for heating. This information cannot be found directly in the registers, but is well known that households with higher income chooses a higher indoor temperature. Therefore we use information on household income to take account of differences in indoor temperature. Also significantly different energy behaviours according

8. The price of a single family house energy label was about 400 Euro in 2002 (see Lausten and Lorenzen 2003). In 2008 the price was about 700 Euro. Total costs for 2008 amounted to 32-35 million Euro (see Ditlefsen 2008).

9. The scale has been changed in 2006 when the energy labeling scheme was revised. The scale is now A-G.

10. Other expenses are on electricity and water.

11. Calculated as 50.000 labels times 400 Euro/label. See also footnote 8.

to consumers' ages has been documented in several studies (see e.g. Tonn and Eisenberg (2007) and therefore we include a variable describing age of the house owner.

Denmark has unique registers for both persons and buildings that can be combined with energy consumption data provided by the utilities. For this study a comprehensive database for the period 1999-2002 has been constructed merging data from the energy labelling database[12], data from administrative registers[13] containing socioeconomic data on house owners, (e.g. age, education, income, family composition), data from public administrative registers (BBR)[14] describing each house, both labelled houses and control group houses without an energy label (e.g. size, age, number of rooms, number of storeys) and data on exact energy consumption (metered consumption of natural gas for heating) for two geographically different areas in Denmark provided by two natural gas companies (HNG – urban area close to the capital and MidtNord – a more rural area)[15]. All this data is handled and anonymized by Statistics Denmark.The initial main dataset had information on 37,622 single family houses (27,062 houses from the HNG area and 10,560 houses from the MidtNord area). A comprehensive work has taken place in order to impose a frame and exclude missing observations in the dataset.

First, house owners living together with one or both of their parents are excluded. Compared to a normal family composition such a household will consist of more than two adults, which can both affect energy consumption and household income, which again can affect energy consumption. Also very old houses (built before 1900) are excluded. This, because old houses can have a significant different insulation standard compared to more recently built houses. Significant bigger houses (>350 m^2) are excluded under the assumption that energy use in very big houses is not comparable to energy use in average size houses. This initial trimming reduces the dataset to 24,054 houses. Finally, only houses traded in 1999-2002 are kept in the dataset. This is necessary because we only want houses in the dataset that has been purchased after the energy labelling scheme was introduced. The final total dataset includes 3,956 single-family houses. After this trimming all houses in the final dataset are between 50-350 m^2 and they are built after 1900. The houses are single-family houses used for residence only and occupied by the owner of the house. All houses use natural gas for heating and are located either in the HNG area or MidtNord area. All house owners are at least 18 years old and are not living together with their parents. The trimming of the dataset left us with about 10 % of the houses in the initial dataset. To control for possible selection bias in the final dataset, means of key characteristics of the houses in the final dataset are compared with the stock of single family houses in the initial dataset (see table 1). Household income is significantly different between the two datasets, but the other characteristics are quite similar. This must be kept in mind, when generalizing the results of this paper.

Of the 3,986 single-family houses in the final database, 2,074 houses are energy labelled and 1,912 are non-labelled houses. For each house we have between 1 and 4 observations, as we have data (e.g. metered data for energy consumption) for each house in 1 to 4 years depending on when the houseowner enters and leaves the dataset. A house is defined as labelled, if the energy label is registered within one year after the house is sold. The control group houses are non-labelled houses sold between 1999 and 2002.

The dataset is non-random in at least two respects. First, even though labelling is compulsory a household faces no significant penalty by not entering the scheme. The buildings having entered it may therefore be different from the rest of the building population. Second, the availability of energy consumption data, in this case natural gas for heating, is first limited to major city areas, where natural gas heating is provided, and secondly to areas, where the natural gas companies are willing to provide household natural gas consumption data for research purposes. The non randomness in the building stock between labelled and non labelled houses are handled in the matching process under the assumption that all differences that affect energy consumption can be described by the observed characteristics. The limitation due to availability of energy consumption data means that our results can only be generalised to areas with natural gas heating. In 2007 about 25%[16] of the total danish stock of single family houses is heated with natural gas.

Table 2[17], [18] shows that only about half of the house purchases include an energy label both in the analysed dataset (a) and in the total Danish stock of single family houses (b) (data for the total Danish stock of single family houses are from Statistics Denmark (2005). As mentioned above, earlier evaluations of the energy labelling scheme have revealed significant geographic differences in terms of the coverage of the label. In this paper the datasets are constructed to handle differences in geographical distributions of the coverage in energy labelling by including natural gas consumption data from two geographically different natural gas companies.

As described earlier, the energy labelling scheme grades houses in the range from A1-C5. Table 3[19] shows the distribution of houses in the three main categories A, B and C, and the control group in relation to number of years since the house was bought (YSB). The main percentage of the labelled houses are B-houses. The distribution of labelled houses in the three main catagories are approximately the same in both analysed dataset and total stock of labelled houses. Most houses are observed in the first and second year after the house is bought.

Our covariate set includes a range of house and household characteristics that might be correlated with labelling propensity and influence on energy consumption for heating. As house characteristics, we consider house size and house size related

12. I am very grateful for the free data provided for research purposes by the FEM Secretariate.

13. Research accces available through AKF 10 % register upon payment.

14. Data provided by Gilling Communication and Consulting (distributer of BBR-data) upon payment.

15. I am very grateful for the free data provided for research by the natural gas distribution companies, HNG and Energy MidtNord.

16. Equivalent to about 330,000 dwellings (see Naturgasfakta 2009)

17. Annual number of houses getting labelled: Data from the energy labelling database

18. Annual number of house purchases: Data for a is from the the analysed data-set and data for b is from Statistics Denmark (2005)

19. Total stock of labelled houses in Denmark: Number of labelled buildings in the energy label database by the end of 2004.

Table 1. Mean characteristics for the initial and final datasets

Characteristic	Initial dataset	Final dataset
N = number of single family houses	37,622	3,986
Size of house (m2)	146	140
Energy consumption (kwh/m2)	15.76	16.85
Construction year	1957	1958
Average household income/year (Euro)	60,480	90,728

Table 2. Development in coverage of the labelling scheme in the analysed dataset (a) and in the total danish stock of single family houses (b)

	1999		2000		2001		2002		Total	
	a	b	a	b	a	b	a	b	a	b
Annual number of houses getting labelled	425	27,008	536	27,908	594	23,862	519	23,478	2074	102,256
Annual number of house purchases	1010	53,565	1049	56,140	1093	54,580	834	55,749	3986	220,034
% coverage (labelled houses/purchased)	42.08	50.42	51.11	49.71	54.35	43.72	62.23	42.11	52.03	0.46

Table 3. Distribution of houses in the control group and three main labelling categories in relation to number of years since the house was bought

	Year of house purchase (Ysb=0)	Ysb=1	Ysb=2	Ysb=3	Ysb=4	Total in the dataset	% of labelled houses	Total stock of labelled houses in Denmark	% of the total stock of labelled houses
A	53	330	275	165	75	898	21%	64,277	28%
B	101	993	752	445	208	2499	58%	106,043	46%
C	28	330	277	164	61	860	20%	58,215	25%
Controls	549	1057	1155	719	371	3851	-	-	-
Missing information						10	-	-	-
# observations	731	2710	2459	1493	715	8108	-	-	-

controls like number of toilets, number of bathrooms, number of floors, house structural characteristics like type of roof, type of outside wall, and type of heating installation, and finally year of construction.[20] As household observable characteristics we include number of children at different age intervals, age and education of the main person, and household disposable income.

A variable describing the outdoor temperature is also necessary, when we estimate energy consumption for heating. Here the variable "graddag" (degree day) describing how cold the weather has been each year, is included in the covariate set. An increasing number of degree days means a year with colder mean temperature. Other unexplained annual variations affecting energy consumption are described by dummies for each year (D1999, D2000, D2001). If there should be any structural changes in the house sales market they are caught by dummies describing the year of house purchase. To take account of geographical differences between the two areas we include a dummy variable (Mn=1 if the house is situated in the MidtNord area, and Mn=0 if the house is situated in the HNG area).

Table 4 reports descriptive statistics for a selection the covariates. Due to the fact that we do not observe all houses four years after the house is bought, this table includes the covariate distribution for the biggest of the four sub-samples used in the application, which is the sample of houses observed one year after the house is bought. From the descriptive statistics we have, a mean house in the dataset is a brick house with slate roof. The house is linked to a central heating system based on natural gas. It has two toilets and one bathroom, it is a one

20. Petersen and Gram-Hanssen (2005) point out house size and year of construction looked among the most relevant factors determining energy and water consumption in the Danish households.

Table 4. Selected descriptive statistics for the first year after house purchase (Ysb=1). Presented numbers are means for the covariates

	Description of variable	A-label	B-Label	C-label	All labelled houses	Controls
House characteristics						
Size of house	m^2	150.91	139.02	118.71	137.40	142.56
Energy consumption	M3/m^2/year	14.45	18.65	22.52	18.54	17.63
Energy consumption	Kwh/m^2/year	158.95	205.15	247.72	203.94	193.93
Construction year		1970	1959	1945	1958	1956
# Toilets	1 toilet	0.23	0.39	0.61	0.40	0.39
	2 toilets	0.69	0.56	0.34	0.54	0.54
# Bathrooms	1 bathroom	0.53	0.70	0.82	0.69	0.69
	2 bathrooms	0.46	0.29	0.14	0.29	0.29
# Floors	1 floor	0.98	0.98	0.98	0.98	0.97
	2-3 floors	0.02	0.02	0.02	0.02	0.03
Vintage class	Built 1900 – 1949	0.11	0.22	0.48	0.25	0.30
	1950- 1959	0.05	0.14	0.29	0.15	0.14
	1960-1969	0.19	0.37	0.20	0.30	0.28
	1970-1976	0.30	0.24	0.02	0.21	0.18
	1977-1981	0.15	0.03	0	0.05	0.05
	1982-1984	0.05	0.003	0	0.01	0.004
	1985-1997	0.15	0.01	0	0.04	0.05
Household characteristics						
Total household income	Euro/year	101,486	97,207	88,191	96,449	79,306
Age of house owner at year of house purchase		37.81	36.62	35.31	36.59	37.00
Number of members in household		3.35	3.34	3.19	3.31	3.29
Number of children	Age 0-6	0.74	0.73	0.72	0.73	0.47
	Age 7-14	0.33	0.35	0.25	0.33	0.31
	Age 15-17	0.08	0.09	0.07	0.08	0.05

storey house and there are four household members The house was built 1960-1969.

Table 4 highlights the differences and similarities between both the three labelling categories of labelled houses, and between labelled and control group houses. For example, we can see that A-labelled houses are the most recent built, and consistent with the trend in house size, these houses are also bigger than both B- and C- labelled houses and control group houses. C-labelled houses are smaller. As A- labelled houses are most recent, and therefore contructed under the strictest building regulations, we can observe they also use less energy/m^2 for heating than control group houses, while B- and C-labelled houses use more energy/m^2.

The socioeconomic characteristics of households in the different types of labelled houses also differ, e.g. household income. As mentioned earlier it is documented in other studies (see e.g. Petersen and Gram-Hanssen 2005) that income matters in a household's energy consumption. Lower income households tend to heat less their homes which explains the lower energy consumption level in that group.

The description of the dataset in this section points at relevant differences between labelled houses and non-labelled houses, but also among the three categories of labelled houses. These differences makes it unreliable to estimate the effect of energy labelling by simply comparing energy consumption between labelled houses and non labelled houses (control group). However the matching approach is an excellent evaluation method, because it can handle such differences in explanatory variables by balancing control houses to the characteristics of the labelled houses. The differences in terms of covariate distribution justify using different propensity scores for each type of labelled houses; A, B and C.

This paper assumes that for energy labelling to show some efficiency, the average energy savings would be positive and at least not decreasing over time. This reflecting that recommended investments can be adopted immediately or gradually. Because the houses in the three categories are quite different in relation to covariates affecting energy efficiency we further expect different effects in the average energy savings across the three labelling categories.

The Evaluation Approach

The objective of this paper is the evaluation of the Danish Energy Labelling Scheme on household energy consumption. Ideally we would estimate the causal effect of the labelling scheme by comparing the energy consumption of a group of labelled houses with the energy consumption for the same group of labelled houses, in the same periode – had they not been labelled. This is not possible since, at any point in time, a house is either labelled or not labelled, not both. The challenge in this kind of evaluation is to estimate the missing data/counterfactual (the energy consumption for a labelled house had it not been labelled). There is a vast literature on evaluation methods and the method used in this paper is developed in the albour market area (see for example Heckman and Robb 1985; Heckman, Ichimura and Todd 1997).

Second best, the evaluation of energy labelling should be based on experimental data, where a randomized treatment group of labelled houses is compared to a control group. Such experimential data does not exist and would potentially be costly to compile. Another approach would be a before-and-after comparison of the same house, also called difference-in-difference. This approach is not possible with this dataset, since we do not have data for the labelled houses in the before-label state. An analysis of the level of investments made is also not possible, since such data is not registred. Since such analysis cannot be performed, the closets we can get is to use a quasi-experimental method, where we construct a controlgroup on the basis of propensity score matching (see Rosenbaum and Rubin 1983 or Dehejia and Wahba 1999). When we observe a very wide range of variables that both influence energy labelling and energy consumption, we can apply matching to produces estimates close to what would be produced by an ideal experiment (see Heckman et al 1998).

The main purpose of matching is to re-establish the conditions of an experiment when no randomised control group is available. The matching method aims to construct the correct sample counterpart for the missing information on the treated outcomes (energy consumption for labelled houses) had they not been treated (labelled) by pairing each labelled house with a house from the non-labelled group of houses.

There are two central assumptions when using matching to estimate a treatment effect like energy labelling:

- Conditional Independence Assumption – all the energy consumption relevant differences between labelled and non labelled houses are captured in their observed variables X and energy labelling status of the house is random conditional on the observed variables X
- Common Support – we observe labelled and non labelled houses with the same observed variables X

Given these two assumptions we can use the observed mean energy consumption of the non labelled houses to estimate the mean counterfactual energy consumption the labelled houses would have had, had they not been labelled. The challenge is to ensure that the correct set of observed variables X is being used. If the right amount of information is used, matching deals well with potential bias. In our application, this set of observed variables includes a very wide range of house and household specific characteristics[21]. The data part of this paper showed quite significant differences between labelled and non labelled houses in the observed characteristics indicating a possible selection problem between the two groups of houses[22]. This is no problem for the matching approach. In fact one of the advantages with matching is that the method controls for selection on observed variables. Another advantage in the matching approach is that it does not make any assumptions on the functional form of the output (energy consumption) in contrast to traditional regression analysis.

A very often used method to perform matching is the Propensity Score Matching (PMS) (see Rosenbaum and Rubin 1983) that estimates the probability of a house being labelled given a set of observed variables. Adjustment for the propensity score suffices for removing all biases associated with differences in the observed variables between labeleld and non labelled houses. When using PSM of Rosenbaum and Rubin (1983), the matching procedure is broken down into two stages. The first stage estimates the propensity score $P(D=1|X)$ using a binary discrete choice model. The second stage matches houses on the basis of their predicted probabilities of being labelled.

Many different PSM methods can be obtained by using different methods at the first and second stage. This paper estimates the propensity score with a probit model and uses kernel matching method at the second stage. The idea of this estimator is to match the energy consumption of each labelled house with a weighted average over the set of all houses in the control group. The weights have an Epanechnikov form and their magnitude depends on the distance between predicted probability of a non labelled house and a labelled house (see Heckman, Ichimura and Todd 1997). The propensity score matching can be tested by a balancing test to make sure that the specification is sufficiently flexible. Based on this matching process we calculate the ATT – average treatment effect on the treated. Which in this case is the average effect of energy labelling on natural gas consumption for the labelled houses.

Results

This section discusses the results from estimating the average effect on natural gas consumption of the Danish Energy Labelling Scheme with PSM. The PSM analysis are performed separately for three sub-samples. The data section revealed big differences in terms of covariate distribution not only between labelled and control houses, but also between different labelling categories. In order not to obscure potential different effects, the analysis is performed separately for three samples, where the treatment group is composed by A-, B- or C-labelled houses and the control group is composed by non-labelled houses.

Furthermore, there might be differences in the effect of energy labelling according to how many years the house owner has had to carry out the recommended improvements from the energy label, since some of the recommended investments might be postponed to future years.

Three general hypothetical scenarios can be thought of in relation to the house owner's implementation of these recommendations. First, the house owner can totally ignore the recommendations or at least deliberately decide not to implement any of them. This will result in no significant difference in energy use compared to the control group. Second, the house owner decides to implement all recommendations the first year of owning the house and therefore uses less energy from year 1. This will result in a constant energy saving in all years. Third, the house owner decides to implement the recommendations gradually, resulting in an increasing energy saving across future years.

21. See the data section.

22. Another difference between the two groups of houses is that the owners of non labelled houses in fact did not respect the law, when buying a non labelled house (as energy labelling is mandatoty). It can be argued that the house seller is deliberately avoiding the energy labelling since the real estate agent would have made him aware of this legal requirement. The new owner of the house might be uninformed of the legal requirement, since he does not get the same counseling from the real estate agent, and he is therefore not deliberately disrespecting the law. If this is the case we can assume no significant differences between buyers of labelled houses and buyers of non labelled houses.

Because of the expected effect differences in A-, B- and C-labelled houses and expected differences in effect over time since the house was bought, the initial dataset is broken down into 12 sub-samples (A-, B- and C-labelled houses in 1-4 years after house purchase) and separate propensity scores, matches and ATT are estimated for each sub-sample. The propensity scores are estimated with a probit model, this leading to the predicted probability of a house being labelled given observable heterogeneity. We impose common support by excluding those labelled houses whose predicted probability is outside the range of the predicted probabilities for the control group.

First, we present evidence on the matching estimators' ability to balance the covariates for each of the sub-samples. Because of limited space we only present balancing test results for year 1. Balancing test results for year 2, 3 and 4 are available upon request. The left out results reveals the same patterns as year 1. Table 5 presents the results for the sub-samples of A-, B- and C-labelled houses and their natural gas consumption one year after the house purchase. The *SDIFF before match* estimates in columns 2, 5 and 8 are the standardised difference between the sample means of the labelled houses and the control group before the matching process. These estimates are quite large and show the mean differences in selected observed variables between the labelled and non labelled houses. They show the same kind of differences like the ones in table 3 between labelled and non labelled houses, e.g. the variable describing household income (*ln household inc*). *SDIFF before match* is > 0 in all three groups, showing that household income in the labelled houses is larger than in the non labelled houses. Comparing the *SDIFF before match* across the three groups also shows the same differences we saw in table 3 across the groups, e.g. house size (*ln house size*), where we see the same pattern like in table 4. A labelled houses are bigger than non labelled houses and B and C houses are smaller. This highlights the necessity of separate propensity score analysis for the three types of houses.

The *SDIFF after match* in columns 3, 6, 9 are the standardised difference between the sample means of the labelled houses and the control group after the matching process. These estimates show how good the matching process has performed. Compared to the *SDIFF before match* we see significant differences in the estimates for all variables in all three groups. All the standardised differences between labelled and non labelled houses are significantly smaller, and the p-values shows that the probability for the means of the labelled and non labelled houses to be equal is quite close to 1.

Rosenbaum and Rubin (1985) suggest the critical value for reasonable bias not to be bigger than 20%. As seen in tables 5 the Epanechnikov kernel matching balances treatment and control samples quite well, with no covariate presenting *SDIFF after match* bigger than 20%, and almost all covariates with biases after matching smaller than 10%. When comparing across the three sub-samples, the *SDIFF after match* for the B-labelled houses are in general the smallest.

We use the Epanechnikov kernel PSM with bandwidth 0.06.[23] The point and 95% confidence interval estimated ATT for each sub-sample are presented in table 6. Columns 3-5 show the estimated ATT for A-labelled houses in 1-4 years after the house is bought, columns 7-9 present the estimated ATT for B-labelled houses in 1-4 years after the house is bought, and columns 11-13 present the estimated ATT for C-labelled houses in 1-4 years after the house is bought.

Summing up, the matching procedure performs very well for the different sub-samples. The p-values of Hotellings t-test presented in column 4, 7 and 10 complement the picture offered by the Standardized Bias Difference.

Practically all ATT, with the exception of A-labelled houses for years 1 and 2 after labelling, are insignificant at 5%, rejecting the hypothesis of average energy savings due to labelling in Denmark several years after a house is labelled. We find only a significant negative ATT for A-houses for years 1 and 2.

Sensitivity analyses have been conducted to test for the robustness of the results. The results are robust to different specifications for the PSM algorithm. Concretely, estimated ATT with a logit propensity score, different bandwidths, or bi-weight kernel do not depart from the reported results in the paper.[24, 25]

The non significant ATT estimates can be interpreted in several ways. First, no difference in energy consumption between labelled and non labelled houses can represent a case where both labelled and non labelled houses carries out energy renovations independent of the energy label. Second, it can represent a case where neither labelled nor non labelled houses carries out energy renovations. Unfortunately the Danish registers do not contain information on whether or not energy renovations have been carried out, so based on the register data used in this paper it is not possible to say which case is the true interpretation.

Based on other studies of the Danish Energy Labelling Scheme we find support for the first case in the evaluation from 2001. Here they found that non labelled houses were energy renovated for more or less the same amount of money as were the energy labelled houses, and they gained energy savings at similar levels (see Madsen, Ramlau and Pedersen 2001).

Other studies support the second case e.g. Jensen (2004). Jensen investigates why house owners do not invest in energy efficient solutions. Among other things he finds that it is not classical barriers like money constraints, lack of interest or knowledge.The problem is that house owners find other factors more important than energy consumption and energy savings. For them the visual improvements of the house are more important, e.g. new kitchen, new bathroom. A study conducted by Danish Energy also supports the second case. It shows that only 20% of house owners are willing to spend more than 4,000 Euro on energy improvements of their house, and they are not willing to accept a return on their investment on more than 6 years (see Danish Energy 2007). Unfortunately, a large amount of the proposed investments is not paid back before 20-40 years after the investment is made (see Madsen, Ramlau and Pedersen 2001). Based on this, a large part of the proposed investments in energy renovation will not be carried out due to the long pay back periods.

23. The psmatch2 procedure is used in this application (see Leuven and Sianesi 2003).

24. Results of other bandwidths are available upon request.

25. Results from the sensitivity analysis can be presented upon request.

Table 5. Covariates balance analysis between labelled and non-labelled houses based on Epanechnikov kernel matching. Dependent variable = log (consumption of natural gas /size of house). Year 1 after house purchase

	Matched A-labelled houses			Matched B-labelled houses			Matched C-labelled houses		
# labelled houses	317			973			324		
# control houses	1045			1045			1045		
Variable name	SDIFF before match	SDIFF after match	p>|t|	SDIFF before match	SDIFF after match	p>|t|	SDIFF before match	SDIFF after match	p>|t|
Mn	-39.3	-0.1	0.980	-36.3	0.7	0.783	-30.6	-1.1	0.846
Built 1900-1949	-46.6	0.0	0.994	-18.9	-1.1	0.791	37.5	-0.1	0.989
Built 1950-1959	-28.9	-0.7	0.904	-0.3	1.4	0.758	38.4	-0.7	0.939
Built 1960-1976	30.4	-2.3	0.793	15.1	-2.3	0.634	-56.0	-2.2	0.579
Built 1977-1981	34.7	6.4	0.492	-13.3	-0.9	0.822	-	-	-
Built 1982-1984	12.1	-6.2	0.572	-1.3	0.8	0.855	-	-	-
Built 1985_1997	35.1	1.9	0.840	-23.3	0.6	0.814	-	-	-
Ln household inc	49.4	1.5	0.763	43.6	-0.6	0.815	25.3	-0.5	0.938
Ln house size	31.9	-4.0	0.577	-4.5	-1.6	0.722	-69.2	-1.2	0.875
A_yob	7.9	-1.3	0.860	-2.9	-3.2	0.437	-15.1	0.8	0.907
Household members *(antper)*	3.1	2.3	0.756	3.0	-1.5	0.717	-8.4	-2.2	0.748
Antper2	-2.9	1.1	0.865	-1.8	-1.1	0.783	-12.5	-1.7	0.776
Kids age 0-6	29.6	4.3	0.615	32.2	-1.3	0.793	31.2	0.6	0.943
Kids age 7-14	5.1	-1.7	0.829	6.1	-0.9	0.843	-9.9	-4.1	0.587
Kids age 15-17	11.5	1.7	0.845	12.7	-0.5	0.928	8.7	-0.1	0.986
Outer wall concrete	-21.1	2.2	0.714	-0.1	0.3	0.954	21.8	-1.7	0.844
Outer wall other	3.5	3.1	0.700	-0.1	0.4	0.927	16.9	-0.4	0.967
Roof cement	30.7	1.3	0.882	-11.5	-0.4	0.920	-35.4	0.0	0.999
Roof tile	-17.2	1.2	0.874	-9.1	-1.7	0.701	26.9	-2.7	0.740
Roof other	-10.0	-1.6	0.834	3.1	1.3	0.775	0.8	0.7	0.928
Suppl. heat other	5.6	-0.7	0.935	-2.3	-0.0	0.994	-	-	-
Suppl. heat wood b	5.5	-0.1	0.988	8.8	-2.3	0.618	-12.6	-4.6	0.542
Suppl. heat open f	-3.7	1.8	0.805	2.0	1.5	0.741	-5.9	1.8	0.802
Suppl. heat solarp	7.2	-0.1	0.995	-2.3	-0.6	0.893	-	-	-
Toilet_1	-34.9	2.1	0.769	0.4	-1.2	0.791	45.4	-3.8	0.630
Toilet_3	3.4	0.7	0.931	-2.5	-0.9	0.841	-8.6	0.9	0.904
Bath_2	33.6	-3.7	0.654	-1.4	1.3	0.774	-37.6	-0.5	0.945
Bath_3	0.4	-0.4	0.963	-2.5	-2.4	0.590	-5.5	3.9	0.516
Floors_2	-2.5	1.8	0.808	-6.5	0.3	0.933	-2.8	-1.8	0.813
Education_1	-17.6	-2.3	0.755	-13.2	1.3	0.761	-21.3	-1.0	0.889
Education_3	1.2	-0.6	0.938	2.1	-1.4	0.762	22.0	3.0	0.714
Education_4	14.6	1.6	0.849	12.0	-1.5	0.758	-0.3	-2.4	0.761
Education_0	-16.6	-2.1	0.732	-12.6	-2.2	0.575	-12.6	-0.7	0.914
Graddag	21.5	0.9	0.908	22.1	3.0	0.490	30.0	-1.6	0.914
D2002	0.3	-1.9	0.808	9.4	0.2	0.966	0.6	-1.3	0.872
Yob99_00	-18.9	0.2	0.983	-	-	-	11.5	-0.9	0.908

Note: p-value is the probability for the means of labelled and and non labelled to be equal. P-values are estimated by Hotellings t-test.

Table 6. Propensity Score Matching Estimation of Average Treatment Effect for the Treated

	A-labelled house				B-labelled house				C-labelled house			
	N1	[	~ ATT	]	N1	[	~ ATT	]	N1	[	~ ATT	]
Years after house buy												
1	1362	-0.214	-0.141	-0.103	2018	-0.063	-0.009	0.023	1369	-0.027	0.073	0.229
2	1417	-0.129	-0.082	-0.023	1889	-0.032	-0.001	0.060	1422	-0.073	-0.010	0.072
3	869	-0.078	-0.009	0.077	1155	-0.024	0.012	0.081	877	-0.072	0.007	0.078
4	437	-0.212	-0.068	0.042	565	-0.111	-0.059	0.006	424	-0.115	0.032	0.169

Note: [= 95% lower bound.] = 95% upper bound. ATT: point estimate of Average Treatment Effect..Epanechnikov kernel PSM with bandwidth 0.06. Bootstrap Standard errors used to construct confidence interval (see Leuven and Sianesi.2003)

When comparing the results of this paper with other studies of the Danish Energy Labelling Scheme the conclusion is not clear-cut. We find no effect of the labelling scheme in terms of differences in the natural gas consumption between labelled and non labelled houses. But whether or not this means that no energy renovations are carried out is not possible to conclude. A complementary study would be needed to confirm the results from this statistical approach by verifying whether energy improvements were made or not.

A third interpretation could be that house owners starts to implement the suggested energy renovations later than the periode of time for which we have data (maximum 4 years after house purchase). E.g. because of house buyers' lack of money for some years after house purchase before being able to invest in energy renovations.

As mentioned earlier in the paper, there were significant differences between the labelled and the control group when looking at the registered characteristics. These are supposed to be corrected by the statistical approach used. It cannot be tested whether the corrections made on observed variables also adjust the possible differences in energy behaviours. It is therefore left to future research tyo investigate whether the similarities between both groups are due to the fact that both groups implemented (or not) the same level of energy improvements, or because differences in energy behaviours would compensate higher levels of energy improvements in the "labelled" group for example.

Conclusion

The main purpose of this paper has been to evaluate the Danish Energy Labelling Scheme. The evaluation has been carried out by merging data on consumption of natural gas for heating in single-family houses with register data on house specific characteristics and household characteristics. We have used propensity score matching to estimate average treatment on the treated (ATT) – average labelling effect in terms of natural gas consumption on labelled houses. Propensity score matching has been carried out on 12 sub-samples; A-, B- and C-labelled houses observed 1-4 years after the house purchase. Where the time of house purchase also is the time, where the new house owners get the energy label information. The sub-samples were constructed to examine whether the effect of the energy labelling scheme on energy consumption – if such could be found significant – would depend on the energy related state of the house and/or would be related to the time passed since the house was labelled. The hypothesis being: if a significant effect on consumption of natural gas could be found, then we would expect it to be negative.

With exception of A-labelled houses in the first two years after house purchase, all the estimated ATT are insignificant, and therefore our empirical results cannot support the hypothesis of significant average energy savings due to the Danish Energy Labelling Scheme several years after a house is bought. Whether the insignificant differences in energy consumption between labelled and non labelled houses means that energy saving renovations are carried out independent of the energy label, or energy saving renovations are not carried out at all is not possible to conclude. Not even when comparing with other studies of the Danish Energy Labelling Scheme.

The empirical results of this paper support the findings from other studies on the Danish Energy Labelling Scheme. For instance, the Madsen, Ramlau and Pedersen (2001) study that found very small and close to statistical insignificant differences in investment levels and energy savings in labelled and non-labelled houses, Jensen (2004) who identified a number of barriers for realising energy savings in buildings and the Gram-Hanssen and Jensen (2006) study that found the respondents remembered the label, but they have not really used the information from the label. Finally, complementary studies using on-field inspection (to verify whether energy improvements were made) on smaller samples would be needed to confirm the conclusion from this study based on statistical approach and actual energy consumption data (but without data on energy investments made). Other existing studies tend to show that an energy labelling scheme alone is not sufficient to induce renovation works including energy improvements.

References

Building Energy Labelling Forum Agenda (2006): *Building Energy Labelling Forum* held Friday December 1, 2006, in Toronto. Office of Energy Efficiency, Commercial and Institutional Buildings, Natural Resources Canada.

Danish Energy (2007): *CATINÉT research for ELFOR* March 2007.

Danish Energy Authority: Energistyrelsen. www.ens.dk.

Dehejia, R.H. and S. Wahba (1999): Causal Effects in Non-Experimental Studies: Evaluating the Evaluation of Training Programs. *Journal of the American Statistical Association* 94: 1053-1062.

Ditlefsen, J. (2008): *Email correspondence*. Danish Energy Agency. December 2008.

EA Energy Analysis (2008): *En vej til flere og billigere energibesparelser*. Teknisk bilag M1. Rundspørge blandt parcelhusejere. Only available in danish.

European Comission (2002): *The Directive 2002/91/EC (EPBD, 2003.)* www.buildingsplatform.org/cms/index.php?id=8

Gram-Hanssen, K. and O.M. Jensen (2006): *Energimærkning af enfamiliehuse – en kvalitativ analyse.* Notat. Statens Byggeforskningsinstitut [Danish Building Research Institute]. Only available in Danish.

Heckman, J.J. and R. Robb (1985): Alternative Methods for Evaluating the Impact of Interventions. In: J.J. Heckman and B. Singer (eds.): *Longitudinal Analysis of Labor Market Data*. New York, Cambridge University Press: 156-245.

Heckman, J.J., H. Ichimura and P. Todd (1997): Matching as an Econometric Evaluation Estimator. *Review of Economic Studies* 65: 261-294.

Heckman, J.J, H. Ichimura, J. Smith and P. Todd (1998): *Characterising Selection Bias Using Experimental data*. Econometrica, Vol. 66, N°5, 1017-1098.

Jensen, O.M. (2004): *Barrierer for realisering af energibesparelser i bygninger. Statens Byggeforskningsinstitut.* Only available in Danish.

Lausten and Lorenzen (2003): *Danish Experience in Energy Labelling of Buildings.* COWI. 2003

Leuven, E. and B. Sianesi (2003): psmatch2: Stata module to perform full Mahalanobis and propensity score matching, common support graphing, and covariate imbalance testing.

Madsen, B.B.; M. Ramlau and N.B. Pedersen (2001): *Evaluering af Energimærkningsordningen. Slutrapport.* Made by COWI for Energistyrelsen. Only available in Danish.

Naturgasfakta (2009): www.naturgasfakta.dk. Website in Danish about natural gas in Denmark.

Peréz-Lombard, L., J. Ortiz, R. González and I. R. Maestre (2009): A Review of Benchmarking, Rating and Labelling Concepts Within the Framework of Building Energy Certification Schemes. *Energy and Buildings* 41 (2009): 272-278.

Petersen, K.N. and K. Gram-Hanssen (2005): *Husholdningernes energi- og vandforbrug. Afhængighed af socioøkonomiske baggrundsvariable.* SBI 2005:09. Statens Byggeforskningsinstitut. Only available in Danish.

RESNET homepage (2008): www.natresnet.org. Homepage for Residential Energy Network Service.

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

Rosenbaum, P. and D. Rubin (1985): Reducing Bias in Observational Studies Using Subclassification on the Propensity Score. *Journal of the American Statistical Association* 79: 516-524.

Rubin, D.B. (1974): Estimating Causal Effects of Treatment in Randomized and Nonrandomized Studies. *Journal of Education Psychology* 66: 688-701.

Statistics Denmark (2005): *Statistical ten-year review 2005.*

Statistics Denmark (2008): *New from Statistics Denmark.* Statement of dwellings. NR. 280, 27th of June 2008.

Thomsen, K.E.; K.B. Wittchen, O.M. Jensen and S. Aggerholm (2006): *Applying the EPBD to improve the Energy Performance Requirements for Existing Buildings – ENPER-EXIST.* WP3: Building stock knowledge.

TONN, B. and EISENBERG, J. (2007): The aging US population and residential energy demand. Energy Policy, 35(1), 743-745.)

Wittchen, K.B. (2009): Potentielle energibesparelser i det eksisterende byggeri. SBI2009:05. Statens Byggeforskningsinstitut, Ålborg Universitet. Only available in Danish