Grantham Research Institute on Climate Change and the Environment

Centre for Climate Change Economics and Policy

Why the energy efficiency gap is smaller than we think: quantifying heterogeneity and persistence in the returns to energy efficiency measures

Daire McCoy and Raphaela Kotsch

November 2018

Centre for Climate Change Economics and Policy Working Paper No. 340
ISSN 2515-5709 (Online)

Grantham Research Institute on Climate Change and the Environment
Working Paper No. 306
ISSN 2515-5717 (Online)

ESRC Economic & Social Research Council

The Grantham Foundation for the Protection of the Environment

University of Leeds

LSE The London School of Economics and Political Science

The Centre for Climate Change Economics and Policy (CCCEP) was established by the University of Leeds and the London School of Economics and Political Science in 2008 to advance public and private action on climate change through innovative, rigorous research. The Centre is funded by the UK Economic and Social Research Council. Its third phase started in October 2018 with seven projects:

1. Low-carbon, climate-resilient cities
2. Sustainable infrastructure finance
3. Low-carbon industrial strategies in challenging contexts
4. Integrating climate and development policies for 'climate compatible development'
5. Competitiveness in the low-carbon economy
6. Incentives for behaviour change
7. Climate information for adaptation

More information about CCCEP is available at www.cccep.ac.uk

The Grantham Research Institute on Climate Change and the Environment was established by the London School of Economics and Political Science in 2008 to bring together international expertise on economics, finance, geography, the environment, international development and political economy to create a world-leading centre for policy-relevant research and training. The Institute is funded by the Grantham Foundation for the Protection of the Environment and the Global Green Growth Institute. It has six research themes:

1. Sustainable development
2. Finance, investment and insurance
3. Changing behaviours
4. Growth and innovation
5. Policy design and evaluation
6. Governance and legislation

More information about the Grantham Research Institute is available at www.lse.ac.uk/GranthamInstitute

Suggested citation:
McCoy D and Kotsch R (2018) *Why the energy efficiency gap is smaller than we think: quantifying heterogeneity and persistence in the returns to energy efficiency measures*. Centre for Climate Change Economics and Policy Working Paper 340/Grantham Research Institute on Climate Change and the Environment Working Paper 306. London: London School of Economics and Political Science

Why the energy efficiency gap is smaller than we think: quantifying heterogeneity and persistence in the returns to energy efficiency measures

Daire McCoy, Raphaela Kotsch*

November 2018

Abstract

The "energy-efficiency gap" is a topic that has received much attention in the academic literature. While the role of market and behavioural failures have been discussed at length, much less focus has been on quantifying the magnitude of heterogeneity and persistence that exists in the realised savings from installing measures. This paper fills a gap in the literature by providing new evidence on the persistence of savings over time for various measures and household types. Not only do households in more deprived areas experience lower energy savings, for certain measures the savings erode more quickly over time for these households. This result is important for improving our understanding of the incentives faced by households, for better evaluating the cost-effectiveness of public policies and raises new concerns over the how the costs and benefits of policies are distributed. It also suggests that the energy-efficiency gap requires less explanation than some would suggest.

Keywords: Energy efficiency gap; policy evaluation; panel data econometrics

JEL codes: C23; D12; Q40; Q48

*McCoy (corresponding author): Grantham Research Institute, London School of Economics and Political Science, Houghton Street, London, U.K. WC2A 2AE, Economic and Social Research Institute, Sir John Rogerson's Quay, Dublin, Ireland, d.m.mccoy@lse.ac.uk; Kotsch: Department of Geography and Environment, London School of Economics and Political Science, Houghton Street, London, U.K. WC2A 2AE. This work makes use of data from the National Energy Efficiency framework Database (NEED) UK. We are grateful for financial support for this research from the European Investment Bank, EIBURS grant scheme. Project title: "Policies to finance energy efficiency: an applied welfare assessment". We thank Antoine Dechezleprtre, Francois Cohen, Louis-Gaetan Giraudet and Karsten Neuhoff for securing this funding. This research has also been supported by the Grantham Research Institute and the ESRC Centre for Climate Change Economics and Policy under Grant No. ES/K006576/1. We thank seminar participants at AIEE (2017), FSR Annual Climate Conference (2017), IEB International Academic Symposium (2018), UKNEE enevecon (2018), LSE, Grantham Research Institute PDE Workshop (2018), Montpelier Energy Workshop (2018), EAERE Work Congress (2018) and BIEE (2018) for comments. The authors are extremely grateful to Arlan Brucal and Roger Fouquet for useful discussions and for providing comments on an earlier draft. We would also like to thank Kenneth Gillingham, Meredith Fowlie and Matt Kotchen for useful discussions. All remaining errors are our own.

1 Introduction

How much savings do energy efficiency measures actually deliver? Answering this seemingly straightforward question is crucial to understanding why households appear to under-invest in energy efficiency measures relative to what is socially or even privately optimal, the so-called "energy-efficiency gap".[1] However, there has been much debate on the size of this gap and the relative contribution of market failures, behavioural anomalies or model and measurement error. A recent paper by Gerarden et al. (2015) argues that the energy efficiency gap may not be as large as expected and that unobserved costs, overstated savings from adoption, consumer heterogeneity, inappropriate discount rates and uncertainty may all contribute to the low adoption rate not being as "paradoxical as it first appears."[2] This is a problem that affects both individual incentives to invest and evaluations of policies aimed at encouraging adoption of energy efficiency measures, as overstated returns make private investments seem more attractive, and makes policies appear more cost-effective than they actually are. This issue is exacerbated by the fact that many policy evaluations rely on ex-ante engineering estimates of savings, and rarely consider behavioural responses.

A November 2017 Wall Street Journal op-ed by Sam Ori explored "Why Government energy-efficiency programs sound great-but often don't work" (Ori, 2017). This piece focused on research which demonstrated that the upfront investment costs can be up to twice the actual savings, and the engineering estimates of energy savings can be more than three times what is actually realised (Fowlie et al., 2018). Other research by Allcott and Greenstone (2017) adds further evidence of overstated savings from ex-ante engineering estimates and also highlights large unobserved benefits and costs which evaluations tend to miss.

Evaluations of energy efficiency improvements tend to take a short time-scale, usually a window of 1-2 years on either side of the intervention in order to assess the magnitude of savings. This is despite the fact that time-scale has proven an important factor when examining the impact of building energy codes on energy consumption (Kotchen, 2017), and on the effect of behavioural interventions to reduce energy consumption (Allcott and Rogers, 2014). Treatment effects can vary over time due to a range of factors. This could also be the case for energy efficiency measures as specific factors related to usage patterns in any particular period may bias results both before and after, while poor installation quality or degradation in the installed equipment may affect the results post-installation. Variation over time could affect the accuracy of measurement, the attractiveness of

[1] The reluctance of some consumers to make energy saving investments that offer them seemingly positive net-present value (NPV) returns has been widely studied. For example see Hausman (1979); Blumstein et al. (1980); Jaffe and Stavins (1994); Golove and Eto (1996); Allcott and Greenstone (2012).

[2] A related literature focuses on supply side market failures, imperfect competition and suboptimal rate of innovation. See for example Goldberg (1998); Cohen et al. (2017); Brucal et al. (2017)

the investment, or the cost-effectiveness of a government scheme. Further, variations in energy prices both before and after the installation may affect both expectations and realisations of the investment's net-present value.

This research contributes by providing information on the persistence and heterogeneity of savings associated with installing energy efficiency measures. Uniquely, we demonstrate how the savings from measures change over time for different household types. Not only do households in more deprived areas experience lower energy savings, the savings erode more quickly over time - in some cases reducing by 50 percent within six years (for measures expected to last twice this amount of time). This result has important implications for improving our understanding of the investment incentives households face and also for improving our evaluations of energy efficiency policies.

In order to conduct this analysis we exploit an extremely large database of home energy efficiency upgrades and metered energy consumption[3], covering over four million households and a period of eight years. By combining statistical matching and a range of panel econometric estimators we control for unobserved heterogeneity and selection into various government schemes which funded the upgrades. Another novel feature of this analysis is that our database covers the universe of households entering into energy efficiency schemes administered by energy suppliers in the UK, thus reducing the potential for "site-selection bias" as identified by Allcott (2015).

The data allows us to examine the variation in performance depending on when measures were installed, how they perform over time; how this varies by dwelling and socioeconomic characteristics, and ultimately how this affects the cost-effectiveness of measures for different household types. Results indicate significant cross-sectional and temporal variation in energy savings, that the persistence of savings varies by the type of measure installed and the socioeconomic characteristics of the household. The measures are generally still NPV positive, and compare favourably with the cost-effectiveness of other initiatives, but the returns are much lower than expected. This research also raises concerns over distributional factors given how the costs of policies are subsequently levied on households.

The rest of the paper is organised as follows; Section 2 provides the context in which this analysis takes place; Section 3 the data; Section 4 describes the methodological approach employed and considerations undertaken; Section 5 outlines the results; Section 6 provides the results of robustness checks and sensitivity analysis; Section 7 provides a concluding discussion.

[3]The National Energy Efficiency Framework Database (NEED). Further details available at: https://www.gov.uk/government/collections/national-energy-efficiency-data-need-framework

2 Background

The Supplier Obligation (SO), first introduced to the UK in 1994, has become the principal policy instrument for implementing energy efficiency improvements in the domestic sector in the UK (Rosenow, 2012). The Supplier Obligations are an example of a "Tradable White-Certificate" (TWC) scheme. These are regulatory mechanisms, employing a market-based approach to deliver energy savings. Theoretically, they can be considered a hybrid subsidy-tax instrument, in which suppliers provide subsidies for energy efficiency upgrades that are then recovered through increased energy prices (Giraudet et al., 2012), having parallels with traditional demand-side management (DSM) programmes in that companies are required to invest in projects that ultimately reduce demand for their product (Sorrell et al., 2009b).

As outlined in Bertoldi and Rezessy (2008) and Giraudet et al. (2012), SOs have three main features: an obligation is placed on energy companies to achieve a quantified target of energy savings; savings are based on standardised ex-ante calculations; the obligations can be traded with other obligated parties. This flexibility ideally allows suppliers to choose the most cost-effective way to reach their target. Suppliers bear the cost of installations in the first instance, costs are then passed through to their entire population of customers through increases in energy prices (Chawla et al., 2013). Clearly, this may have distributional consequences if certain segments of the population are less likely to avail of the schemes. To alleviate this concern, targets were imposed regarding the proportion of savings to be achieved from lower income groups.

The former Department of Energy and Climate Change (DECC)[4], sets the savings targets which are then enforced by the energy regulator, the Office of Gas and Electricity Markets (Ofgem). Ofgem sets and administers individual savings targets for each energy supplier. Energy suppliers have various options to achieve their targets such as contracting installers, subsidising energy efficiency products, cooperating with local authorities, delivery agents or supermarkets, or directly working with their customers (Rosenow, 2012).

[4] now Department for Business, Energy and Industrial Strategy (BEIS)

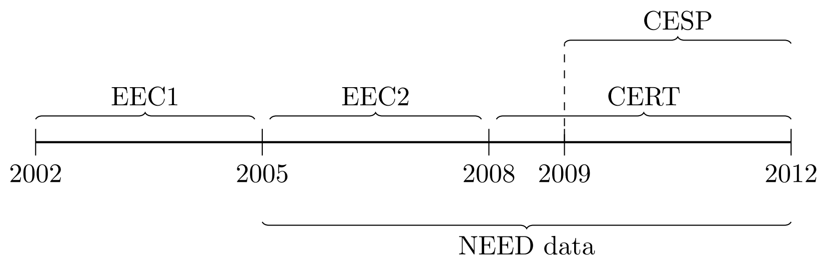

Figure 1: UK Energy Efficiency Programmes 2005-2012

Figure 1 gives an overview of SOs from 2002-2012. The first Energy Efficiency Commitment (EEC1) ran from 2002 to 2005, followed by EEC2 in 2005. In 2008, EEC2 was replaced by the Carbon Emissions Reduction Target (CERT) which ran until 2012. In 2009, the Community Energy Saving Programme (CESP) was introduced in parallel with CERT. While the main architecture of SOs did not change, the savings targets and the costs of the delivering the programmes increased over time. Rosenow (2012) provides a comprehensive overview of the main changes in each scheme from 1994 - 2012 with regards to the target, the costs, social equity implications and other changes in design. The main change concerned the target size, increasing substantially in lifetime savings from 2.7 to 494 terawatt hours (TWh) between 1994 and 2012 (Rosenow, 2012).

From 2002, all programmes included a target for disadvantaged households and fuel poverty increasingly came to the fore. Eventually, CESP only allowed projects to be carried out in specific low income areas of Britain, the lowest 10-15% of areas ranked in Income Domain of the Indices of Multiple Deprivation (Hough and Page, 2015). Thus, CESP was only available in certain geographical regions. Furthermore, CESP introduced a new bonus structure that incentivised the installation of multiple measures in a single dwelling and the treatment of as many dwellings as possible in the same area (Duffy, 2013). Table 1 summarises the key features of the schemes under consideration.

Table 1: Overview of Supplier Obligations

	ECC1	ECC2	CERT	CESP
Target	62 TWh	130 TWh	494 TWh	19.25 Mt CO2
Annual costs (millions)	167	400	1,158	unknown
% savings in priority group	50%	50%	40%	10-15% most deprived areas
Measures installed				
Cavity wall insulation	791,524	1,760,828	2,568,870	3,000
Loft insulations	754,741	1,780,302	3,897,324	23,503
Replacement heating systems	366,488	2,018,812	31,986	42,898

Source: Based on information from Lees (2006, 2008); Rosenow (2012); Duffy (2013)

A key feature of all previous evaluations of the above policies is that the energy savings achieved were based on model ex-ante estimates and not actual ex-post data. Engineering model estimates tend to overstate actual savings significantly, as they are derived from lab-based estimates and factors such as occupancy and behaviour are typically not considered. This would lead to concern over the accuracy of measurement regarding both the energy savings achieved and the cost-effectiveness of various policies in delivering savings.

This work leverages an extremely large dataset of energy efficiency measures to shed new light on the savings delivered by the principal policy initiatives in the UK during the period 2002-2012.

3 Data

The National Energy Efficiency Database (NEED) contains dwelling-level data on four million households, over an eight-year period. Information comes from a range of sources including meter point electricity and gas consumption data, Valuation Office Agency (VOA) property attribute data, the Homes Energy Efficiency Database (HEED) containing data on energy efficiency measures installed, and modelled data provided by Experian on household characteristics. An overview of data types and sources is provided in Table 2.

Table 2: Data sources combined in NEED

Type of variable	Source
Energy efficiency measures	HEED/Ofgem/DECC
Energy consumption	Energy Supplier
Property attributes	VOA
Household characteristics	Experian

Source: NEED database

The remainder of this section will discuss the measures installed, energy consumption and the socioeconomic characteristics of households. Further detail on all variables contained within the dataset is provided in Table A1 and descriptive statistics in Table A2. This includes detailed dwelling information.

3.1 Measures installed

The NEED database includes measures installed through EEC2, CERT and CESP schemes. These schemes were by far the most prevalent mechanism for delivering energy savings in residential dwellings in the UK over this period. The database does not include an exhaustive list of measures installed as part of the various schemes, appliances and lighting also featured but are not included. However, as Table 3 demonstrates, insulation and heating comprised the vast majority of estimated energy savings across various schemes over this period. In total over two million measures were installed over the period within our sample, this is graphically represented in Figure 2.

Table 3: Energy savings by scheme and measure

	EEC1	EEC2	CERT
	2002-2005	2005-2008	2008-2012
Insulation	56%	75%	66.20%
Heating	9%	8%	8.20%
Lighting	24%	12%	17.30%
Appliances	11%	5%	5.90%
Other	-	-	2.40%

Source: Lees (2006, 2008); Ofgem (2013)

Note: savings reported are estimates

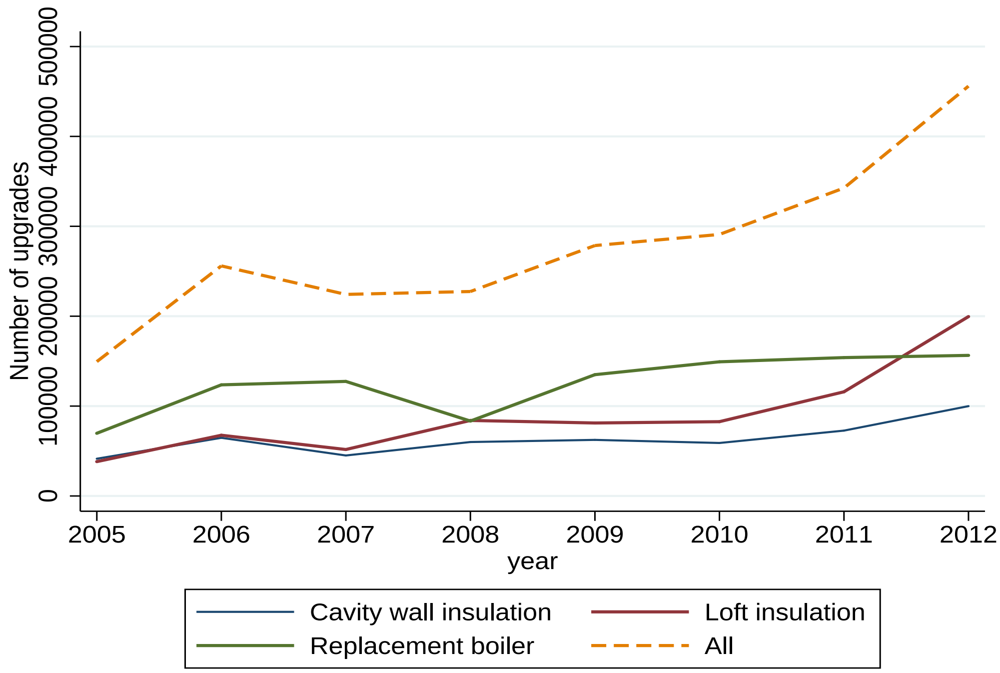

Figure 2: Energy efficiency measures installed, 2005-2012. Source: author's calculation based on NEED data

All insulation installations in our dataset were funded through government schemes. Heating upgrades were funded through both public and private means. In the early part of the sample (pre-2007) boiler installations were likely to have been funded through government schemes, however government support for replacement boilers was withdrawn during EEC2, as a combination of previous support schemes and new building regulations in 2005 had already delivered a significant penetration of new condensing boilers. Therefore the boiler data we report on is a combination of publicly and privately funded investments. As specified in the 2005 Building Regulations, all replacement boilers were required to be condensing gas or oil and have a minimum efficiency rating of 86 percent.

3.2 Energy consumption

Figure 3 illustrates that on average, gas consumption reduced by 27% between 2005 and 2012 and electricity consumption reduced by 14%. Both of these trends are encouraging signs that the various polices in place over this period were having an effect.

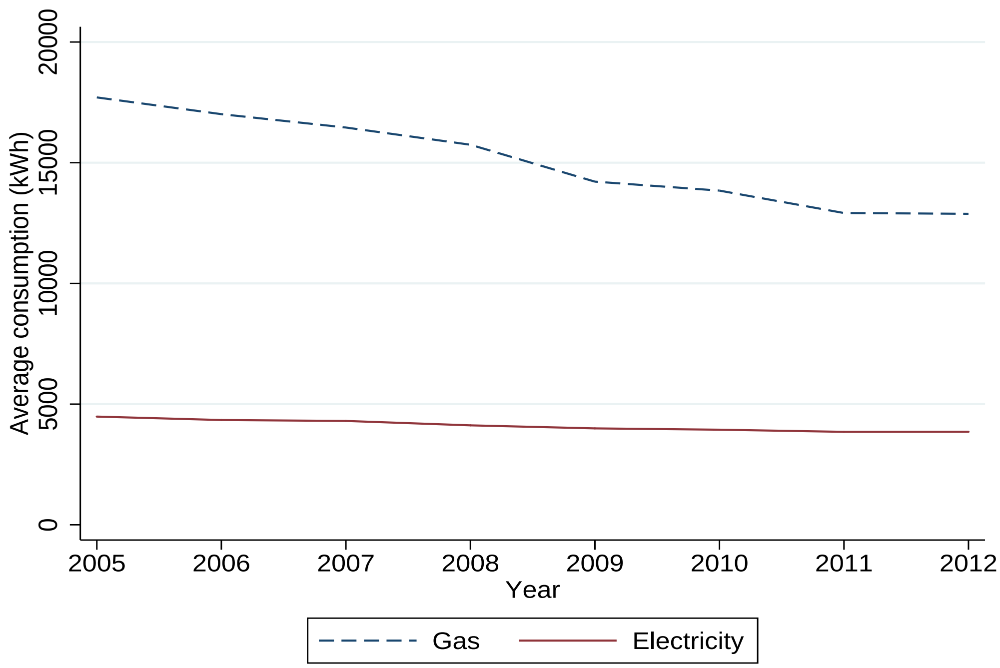

Figure 3: Average domestic energy consumption UK, 2005-2012. Source: author's calculation based on NEED data

3.3 Socioeconomic characteristics

The NEED dataset comprises information on household characteristics modelled by Experian and matched with indicators based on the geographic location of the property (DECC, 2016).[5] For reasons of data protection, the dataset was anonymised and household-level information on variables such as income and tenure-type are not available. However, the dataset does include two composite indicators of the socio-economic background of the households.

1. *Index of multiple deprivation* (IMD)

   NEED contains two variables describing IMDs: IMD 2010 for England and IMD 2011 for Wales. Both indicators classify Lower Layer Super Output Areas (LSOAs) according to a quintile ranking that is based on eight different domains that are incorporated using a weighting scheme. The first quintile (IMD=1) indicates the most deprived areas. Table 4 shows the composition of domains that are incorporated in the indicators and their weight in percent.

[5]Experian is a commercial organisation which produces modelled data of household characteristics at address level. These are derived from a range of sources including Census outputs and Experian's consumer survey.

Table 4: Composition of IMD in %

	England 2010	Wales 2011
Income	22.5	23.5
Employment	22.5	23.5
Health	13.5	14
Education	13.5	14
Access/barriers to services	9.3	10
Living environment/ housing	9.3	5
Physical environment	0	5
Crime [Wales: Community Safety]	9.3	5

Source: Payne and Abel (2012); ONS (2011)

2. *Fuel poverty indicator* (FP)

   Combining data from the English Housing Survey and Census data, the fuel poverty indicator indicates if households are fuel poor based on the households' income and energy requirements, as well as on fuel prices (BEIS, 2013).

4 Econometric approach

4.1 Empirical framework

Lancastrian Demand Theory suggests that households do not derive utility from consumption of a good but from the characteristics of the good and its combination with other goods (Lancaster, 1966). In this particular case, households derive benefits from their consumption of energy services and not units of energy per se. Energy services can be considered as a final good produced by the combination of energy inputs and capital equipment. The services produced include lighting, space heating, water heating. The capital equipment includes heating systems and insulation. These are intermediate goods in the production of thermal comfort or other measures of household satisfaction (Quigley and Rubinfeld, 1989).

The benefit a household derives from an improvement in the capital equipment can be realised through either a reduction in energy units required to maintain the previous level of energy service or through increased consumption of energy services, due to a reduction in their effective price. The latter effect is sometimes referred to as the "rebound effect" and has been widely studied.[6]

[6]Empirical estimates of the size of this effect vary considerably. See Sorrell et al. (2009a) for an overview. Increased energy service consumption could result in a range of other benefits to households, as has been described as the multiple benefits of energy

As outlined by Nordhaus (1996); Hunt and Ryan (2015); Fouquet et al. (2018) amongst others, not accounting for energy services in any empirical framework may result in model mis-specification and biased estimates. However, energy services can be difficult to observe. To account for this Hunt and Ryan (2015) suggest adding additional terms to the empirical framework in order to account for unobserved energy efficiency. These include adding exogenous time trends, information on past energy prices and the components of past energy prices - all of which help to account for any underlying changes in unobserved energy efficiency. Taking this on board, we propose a reduced-form framework with a rich set of fixed-effects which account for both exogenous improvements in energy efficiency and changes in energy prices without explicitly modelling these factors. In addition, by examining heterogeneity in energy savings for different household types following installation of similar capital equipment, we can implicity observe changes in energy services derived by different households.

4.2 The model

Energy consumption is determined by a range of factors such as temperature, characteristics of the dwelling and its inhabitants, and energy prices. We estimate the following baseline panel specification:

$$ln(y_{it}) = \alpha_i + \gamma_t + \rho_{rt} + \delta \sum_{j=1}^{3} D_{ijt} + \epsilon_{it} \tag{1}$$

Where y_{it} denotes consumption of either electricity or natural gas (both in kWhs) by household i in year t, α_i is a household fixed-effect, γ_t is a year fixed-effect which controls for unobserved factors which vary at an annual level such as broader macroeconomic conditions and weather patterns, ρ_{rt} is a year-by-region fixed effect to control for factors which vary at a sub-national level, such as more localised economic shocks and weather patterns, D_{it} is the treatment dummy. The key parameter of interest is δ the average treatment effect on the treated (ATT). The model is estimated as a first-differenced fixed effects panel specification controlling for unobserved time-invariant household characteristics which might affect energy consumption. Over the course of the analysis, a variety of extensions to the above are estimated, to account for interactions between upgrades and household socioeconomic characteristics, and to examine the performance of upgrades over time. All models are estimated for both gas and electricity consumption. Standard errors are clustered at the household level in all specifications. As will be described in the following sections, the data allow us to create multiple treatment and control groups. Treatment groups are created for the entire sample period and for each individual year of upgrade. This allows us to examine how treatment effects vary over time.

efficiency (Ryan and Campbell, 2012)

4.3 Identification

4.3.1 The problem of unobserved heterogeneity

The fixed effects estimators described above are based on the assumption of conditional mean independence or unconfoundedness, selection on observables or ignorability (Caliendo and Kopeinig, 2005; Angrist and Pischke, 2009; Wooldridge, 2010), which requires that both of the following equations hold:

$$E[Y_{it}^{0}|A_i, t, X_{it}, D_{it}] = E[Y_{it}^{0}|A_i, t, X_{it}] \tag{2}$$

and

$$E[Y_{it}^{1}|A_i, t, X_{it}, D_{it}] = E[Y_{it}^{1}|A_i, t, X_{it}] \tag{3}$$

Thus, it assumes that D_{it} is strictly exogenous and as good as randomly assigned conditional on A_i (Angrist and Pischke, 2009). As we are primarily interested in the effect on the households who enrolled in the schemes - the average treatment effect on the treated (ATT), and not necessarily the effect on the whole population - the average treatment effect (ATE), the condition of unconfoundedness can be relaxed and equation (3) can be ignored. The parameter of interest is, the ATT is defined as:

$$ATT = E[Y_{it}^{1} - Y_{it}^{0}|D_{it} = 1] \tag{4}$$

There is strong evidence that the presence of unobserved heterogeneity leads to inaccurate estimates of the ATE and ATT in a fixed-effects OLS setting (Ferraro and Miranda, 2017; Gibbons et al., 2014). Self-selection bias occurs as households voluntarily decide to apply upgrades in their homes or take part in government funded schemes, potentially causing the treatment and control group to differ systematically in aspects that both affect their likelihood of taking part in energy efficiency programs, and their energy consumption, causing the failure of the conditional mean independence assumption (Wooldridge, 2010). Unobserved heterogeneity between households means that households respond differently to common shocks. For instance, increasing energy prices might lead to different behaviour of low and high income households. Second, the crucial assumption of a linear model with additive and homogeneous effects implies that the fixed effect estimates give a weighted average based on the frequency of groups as well as the sample variances within groups (Gibbons et al., 2014). This is problematic as the fixed effects estimator overweights groups that have larger variance of treatment conditional upon other covariates and underweights groups with smaller conditional variance if heterogenous treatment is prevalent (Ferraro and Miranda, 2017). One strategy to overcome this threat and to obtain consistent and unbiased estimators is to pre-process the data through statistical matching (Wooldridge, 2010). The following section outlines this approach.

4.4 Matching

Policy evaluations of secondary data typically employ statistical matching, along with differences-in-differences estimation, or exploit the longitudinal nature of the data with a panel fixed-effects specification. However, both of these measures may suffer from bias through either unobserved temporal effects or unobserved heterogeneity. Recent research has shown that by combining these methodologies, the accuracy of evaluations can approach that achieved by a randomised-controlled trial (RCT) (Ferraro and Miranda, 2017).

Coarsened-exact matching (CEM) is a non-parametric statistical procedure which improves the estimation of causal effects by reducing imbalance in observed variables between treatment and control groups (Iacus et al., 2008; Blackwell et al., 2009). Iacus et al. (2012) compare CEM with a range of other matching methods using Monte Carlo simulations and conclude that CEM has superior performance in terms of the bias and variance of the ATT. Alberini and Towe (2015) use a similar approach in an analysis of home energy audits in the state of Maryland.

The CEM procedure initially coarsens continuous variables into discrete bins, each with their own "Bin Signature" or strata. Households are then exactly matched to another household in the same strata. The algorithm works by selecting strata that make treated and control groups balanced with respect to X. This can be performed for multiple variables. Weights are produced in which unmatched observations receive a weight of zero. Matched units receive a weight of one if they belong to the treatment group and a weight of $\frac{m_C}{m_T}\frac{m_T^s}{m_C^s}$ if they belong to the control group. This is formally described by Iacus et al. (2012) as:

$$w_i = \begin{cases} 1 & i \in T^s \\ \frac{m_C}{m_T}\frac{m_T^s}{m_C^s} & i \in C^s \end{cases} \tag{5}$$

Where w_i is the calculated weight for $observation_i$, (T^s, C^s) the treated and control units in stratum s, (m_T, m_C) the total number of treated and control units and (m_T^s, m_C^s) the number of treated and control units in stratum s. The trade-off with CEM is choosing between the bin width or number of strata and proportion of disgarded unmatched observations. Overly narrow bin widths result a large number of strata, making matching more difficult and resulting in more unmatched observations. Overly wide bin widths result in a low number of strata, making matching easier but observations may not be sufficiently similar. This is more of a problem when working with small sample sizes.

In this case we are concerned with matching the group that received the energy efficiency upgrades with the group that did not. Covariates on which the matching is performed should be predictors of household energy consumption and simultaneously impact the uptake of energy efficiency upgrades. The IMD of the area in which the household resides, provides information on the household's socioeconomic environment, an important predictor of energy consumption and energy efficiency uptakes (Hamilton et al., 2014). Dwelling characteristics, such as the period in which the dwelling was built has an important impact on residential energy consumption (Brounen et al., 2012; Harold et al., 2015). Regional differences in weather patterns will also affect fuel consumption. In addition, Alberini and Towe (2015) provide evidence that matching solely based on dwelling or household characteristics is not sufficient and can be optimised if past energy usage is also included. Taking into consideration all these factors, matching is performed on the following variables: property age, fuel-type, energy consumption in the pre-treatment years, location of dwelling and the IMD of the area in which the household resides. By performing matching on energy consumption in prior years, we can account for unobservable household and property characteristics that might vary over time, such as the household size, composition and appliance usage. The large sample size allows us to perform 1:1 matching in which each treatment household is matched with a similar control household.

Matching is performed multiple times. In the first instance a matched control group is created for all households who installed upgrades at any point during the sample period (T_{ALL}, C_{ALL}). Following this, a matched control group is created for each individual year in which households installed measures between 2006 and 2011 $(T_{2006}, C_{2006})(T_{2007}, C_{2007})$, (T_{2008}, C_{2008}), (T_{2009}, C_{2009}), (T_{2010}, C_{2010}), (T_{2011}, C_{2011})[7]. Thus multiple treatment and control groups are created. This enables us to systematically vary the length of time in both pre-treatment and post-treatment data and to compare how treatment effects vary at different points in time.

4.4.1 Quality of matching

The quality of the matching process depends on the similarity in the distribution of covariates between treated and matched control group. This is commonly assessed by comparing the standardised difference and variance ratio of the variables in both groups, before and after matching (Caliendo and Kopeinig, 2005). The standardised difference is the difference in sample means in the treated and control group, divided by the corresponding sample variances. Formally:

[7] 2005 and 2012 are omitted to allow for pre-and post-upgrade comparisons

$$d = \frac{\overline{\mathrm{x}}_{treatment} - \overline{\mathrm{x}}_{control}}{\sqrt{\frac{s^2_{treatment}+s^2_{control}}{2}}} \tag{6}$$

It allows for a comparison of balance which is independent of the sample size and measurement unit (Austin, 2009). The smaller the difference, the better, and it is recommended that this ratio should not exceed 10 percent (Austin, 2009).

The variance ratio measures the ratio of the mean variance in the treated and control group for each covariate. Formally:

$$F = \frac{s^2_{treatment}}{s^2_{control}} \tag{7}$$

This should be close to unity (Austin, 2009; Ferraro and Miranda, 2017). A significant divergence from this indicates that the matching model is misspecified. Further methods of balance diagnostic include assessing the magnitude of the difference between treatment and matched control group covariates using tests for statistical significance. However, the use of the t-test for balance testing is criticised for several reasons under which the most problematic is the dependence on the sample size. For instance, randomly discarding control units will always increase the balance, falsely indicating a better balance (Imai et al., 2008).

As can be demonstrated by Figure 4 and Tables B1 and B2 our extremely large sample size allows a high level of precision in matching. A high degree of balance is achieved on both variables used in matching and variables not used in matching as can be seen from the standardised differences, variance ratios and the distributions of matched electricity and gas consumption.

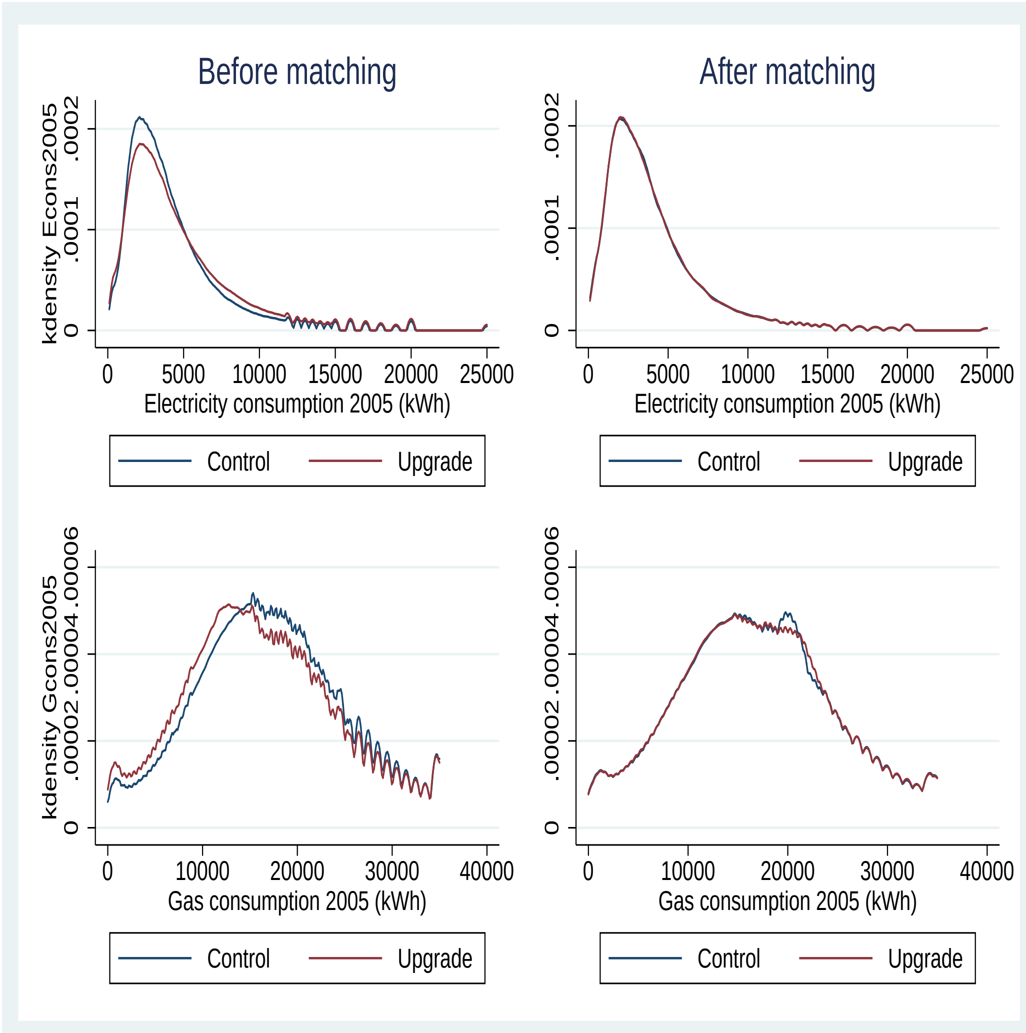

Figure 4: Energy consumption in pre-treatment year 2005, before and after matching. Source: author's calculation based on NEED data

Another important element in assessing the quality of matching is that the parallel paths assumption is not violated. This assumption states that without treatment, the average change for the treated group would have been equal to the observed average change in the control group. Figure 5 demonstrates that this assumption holds for all treatment and control groups.

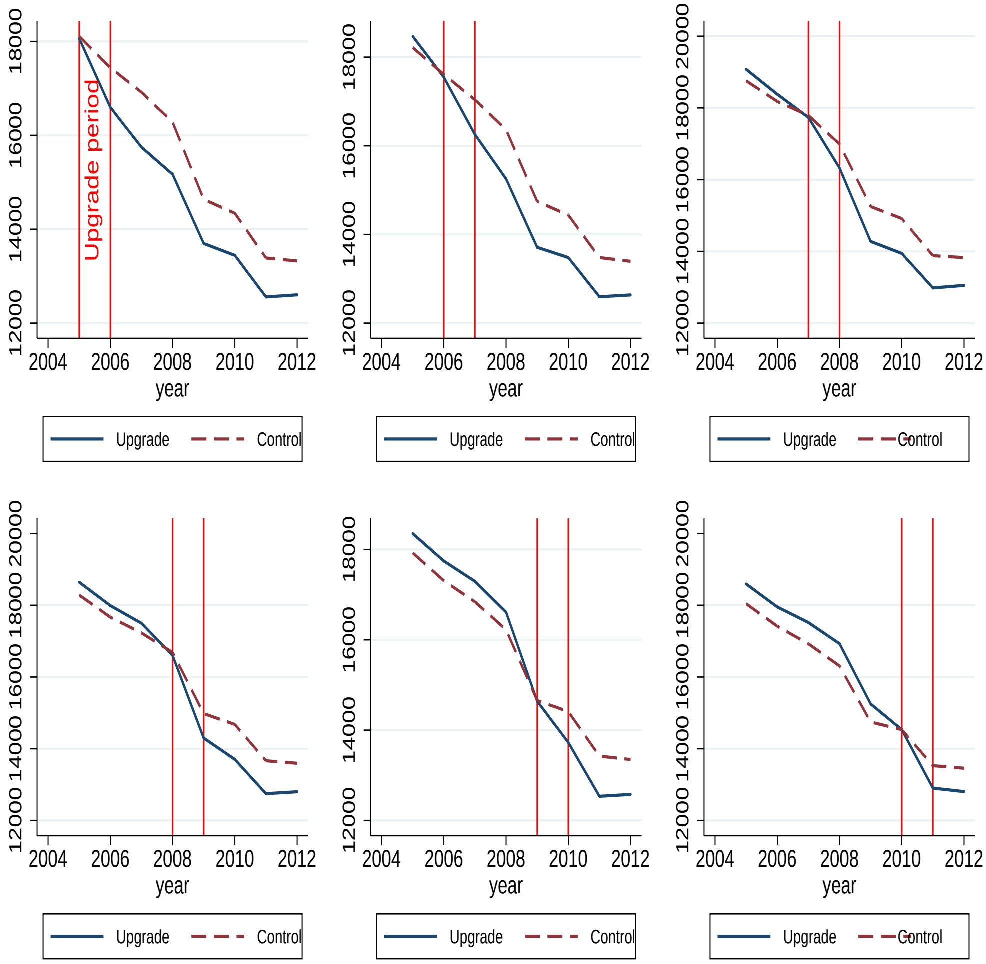

Figure 5: Energy consumption trend in upgrade and control group. Note: For each year in which an upgrade occurs we create separate upgrade and control groups. Source: author's calculation based on NEED data

5 Results

All reported results are estimates of the average treatment effect on the treated (ATT) and can be interpreted as percentage energy savings. Multiple upgrade and control groups are created for the entire period of analysis and for each individual year. This allows us to calculate the average effect and to examine trends over time. Analysis is restricted to households with electricity consumption between 100 and 25,000 kWh, and gas consumption between 3000 and 50,000 kWh. Outliers are excluded to minimise risk of inclusion of invalid consumption readings or non-domestic properties. Following this we create dummy variables to indicate if household energy (either electricity or gas) changed by more than 50, 60 or 70 percent in any given year. These dummy variables are then used in sensitivity analysis to control for any large changes which might haver been the result of unobserved changes in occupancy. For comparison purposes we also calculate the energy savings in kWh for electricity and total energy (Table C1). As these measures primarily impact heating and natural gas is by far the main fuel used for heating the focus of our analysis will be on a subset of dwellings using natural gas for heating. Additional results from a range of sensitivity checks and applying different data trimming are presented in Table D1 and discussed in Section 6.

5.1 The effect of energy efficiency upgrades by year of upgrade

Table 5 shows that the energy savings over time are quite consistent for each measure, regardless of when the installation took place. Annual gas savings for cavity wall insulation range from 8-11 percent, loft insulation 2-3 percent, and replacement heating systems 8-10 percent. These results are consistent with Adan and Fuerst (2015) and Hamilton et al. (2016) who perform similar analysis on this dataset and other related data. This gives us confidence in the accuracy of our central estimates and we proceed on that basis.

Table 5: The effect of energy efficiency upgrades on energy consumption

	(1)	**(2)**	**(3)**	**(4)**	**(5)**	**(6)**	**(7)**
	Full sample	**2006 upgrades**	**2007 upgrades**	**2008 upgrades**	**2009 upgrades**	**2010 upgrades**	**2011 upgrades**
Cavity wall insulation	-0.094***	-0.097***	-0.111***	-0.099***	-0.098***	-0.097***	-0.101***
	(0.001)	(0.002)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Loft insulation	-0.030***	-0.026***	-0.031***	-0.028***	-0.027***	-0.039***	-0.035***
	(0.001)	(0.003)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)
Replacement boiler	-0.092***	-0.080***	-0.093***	-0.087***	-0.102***	-0.109***	-0.099***
	(0.001)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)	(0.002)
Control variables	Y	Y	Y	Y	Y	Y	Y
Household fixed effects	Y	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y	Y
Year*region fixed effects	Y	Y	Y	Y	Y	Y	Y
Observations		617022	545627	564756	730447	746573	871379
Number of households		77128	68203	70595	91306	93322	108922
R squared	0.349	0.327	0.353	0.370	0.369	0.386	0.367

Notes: This table reports coefficient estimates and standard errors from eight separate regressions. The dependent variable in all regressions is the logarithm of annual gas consumption in kilowatt hours. Column(1) "All" denotes efficiency upgrades occurring at any time during the sample period. Columns (2-8) relate to upgrades occurring only in the relevant year. Each individual year denotes upgrades occurring solely in that year. For each upgrade group a matched control group is created using coarsened-exact matching. The sample includes billing records from 2005 to 2012. Standard errors are clustered at the household level. Triple asterisks denote statistical significance at the 1% level; double asterisks at the 1% level; single asterisks at the 10% level.

5.2 Heterogeneity and persistence in returns to energy efficiency upgrades

5.2.1 By measure and IMD group

The next set of results, presented in Figure 6, show the interaction of the treatment variable with the variable indicating the socioeconomic characteristics (deprivation level) of the area in which the household resides. Energy savings are much greater for those households living in more affluent areas (IMD = 5), compared to those in less affluent areas (IMD = 1)[8]. This is true for all upgrade types. Combining all measures, the annual savings range from approximately 15 percent for those in the lowest IMD category to approximately 25 percent for those in the highest. This result raises concerns over distributional issues as the costs of these policies were likely applied as a flat-rate tariff on energy bills (Chawla et al., 2013). If savings are concentrated in the higher income groups, this suggest a further loading of policy costs onto those least able to afford it. Particularly as a flat-rate charge is already regressive, disproportionately affecting those on lower incomes.

[8]Further analysis using quantile regression techniques indicates that this result holds for different levels of energy consumption. Results available on request.

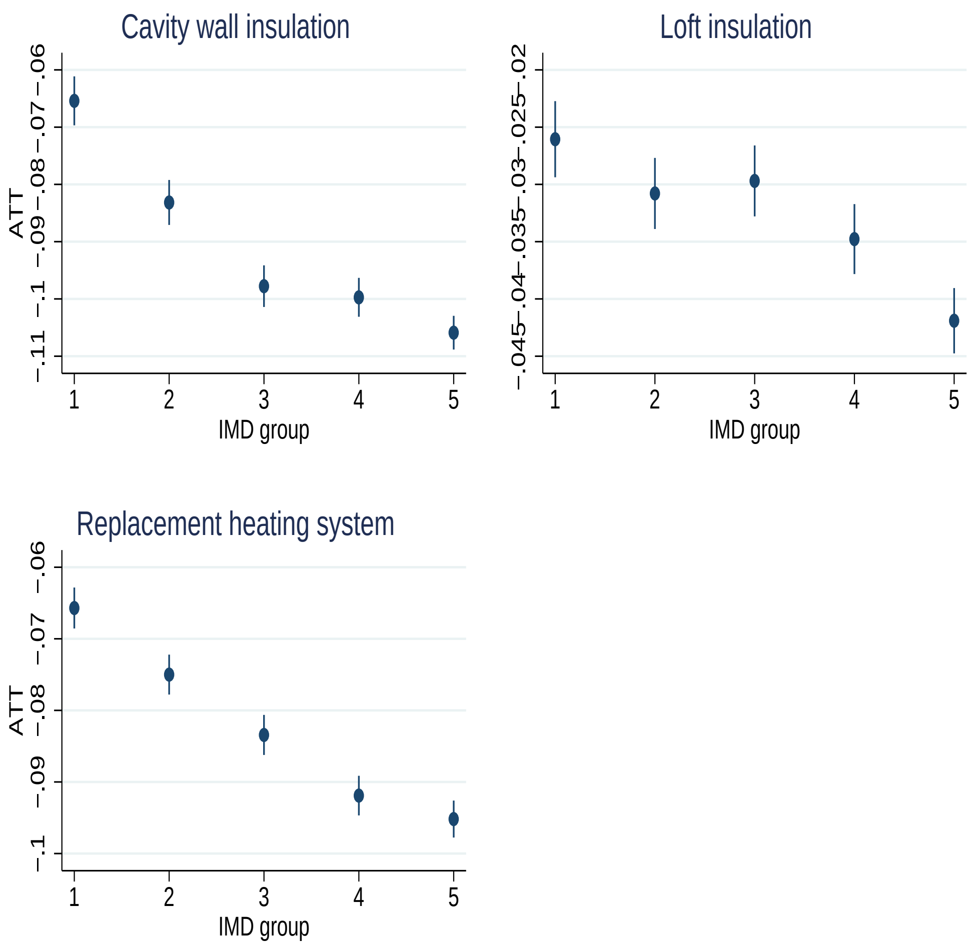

Figure 6: Average Treatment effect on the Treated (ATT) for different IMD groups.
Source: author's calculation based on NEED data

5.2.2 By measure and over time

A key novelty of this research is the ability to examine how measures perform over longer periods of time. Figure 7 presents results of estimations in which the treatment variable is interacted with the year variable to examine the persistence of savings over time. The reported results are for upgrades occurring between 2006 and 2007. This period is chosen as it allows a matched control group to be created using 2005 consumption level, and allows the analyst to observe the longest possible post-upgrade time series. Cavity wall and loft insulation show no clear time trend or degradation. This is not surprising as these measures are expected to last for 30-40 years (Dowson et al., 2012). However, for replacement heating systems the ATT shows a clear decreasing time path. This indicates that energy savings are greatest in the years immediately following installation and decreases thereafter. Given that the estimated lifespan for condensing gas boilers of this vintage was 12 years (Dowson et al., 2012), our results have implications for assessments of both household investment decisions and policy evaluations and require deeper analysis.

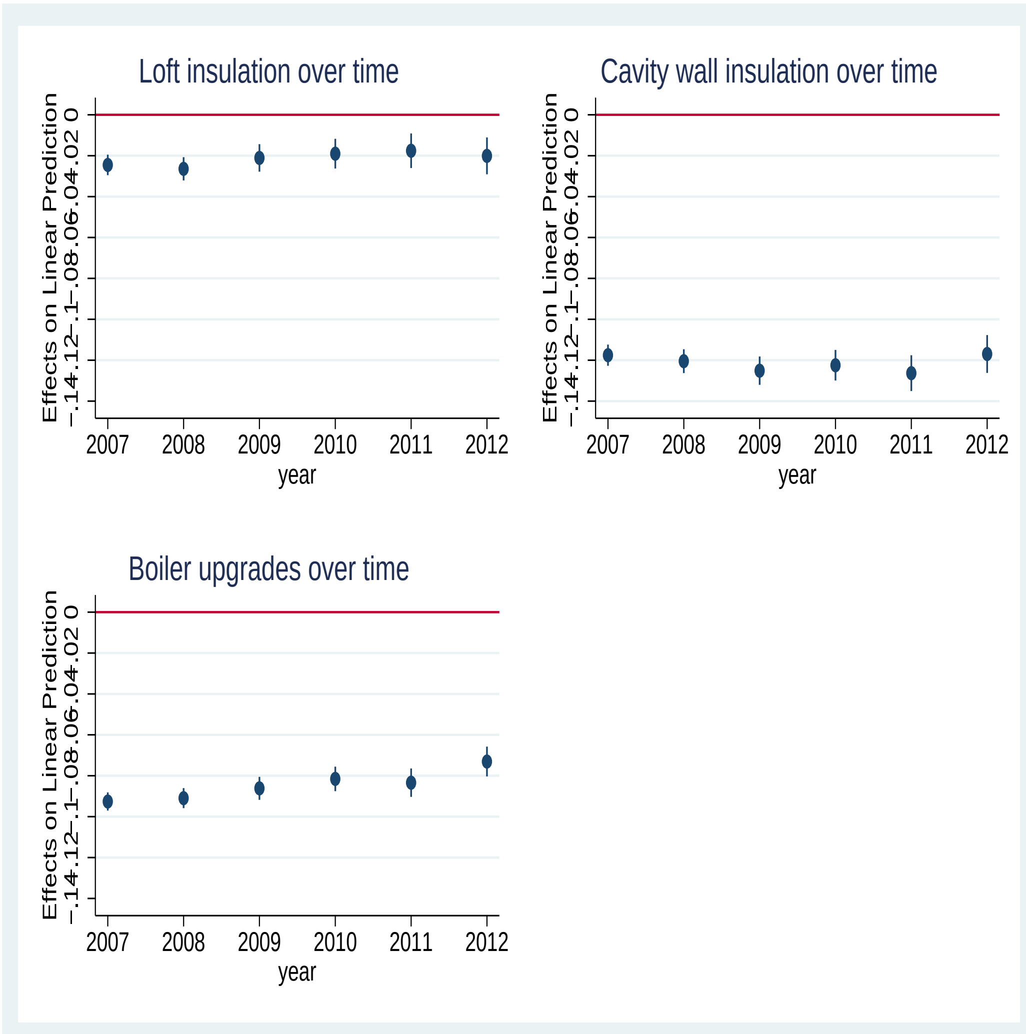

Figure 7: Persistence of Average Treatment effect on the Treated (ATT) over time by measure. Source: author's calculation based on NEED data

5.2.3 Heating system replacements by IMD group over time

The reduction in savings could be due to degradation of equipment or a behavioural response from households. Changes to building regulations in 2005 mandated all replacement boilers to have a minimum of 86% efficiency. Other than this we do not have any detailed product characteristics. However, by decomposing this trend by socioeconomic group it is possible to examine whether this effect varies for different household types. Not only are energy savings less for those in lower income areas, the trend of decreasing savings over time is much more pronounced for these households. The savings for those is the lowest IMD group have halved within five years.

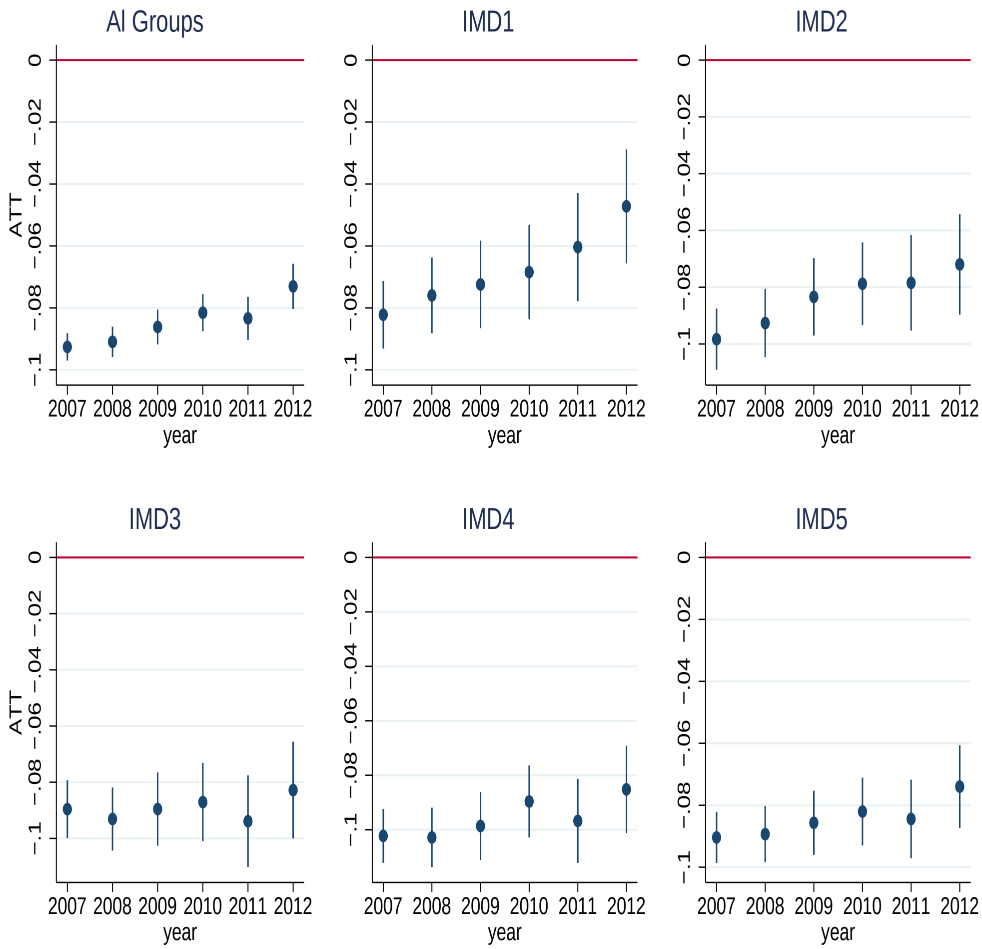

Figure 8: Average Treatment effect on the Treated (ATT) for replacement heating system. Persistence over time and by IMD group. Source: author's calculation based on NEED data

It is not clear what exactly is causing this erosion of savings, particularly whether it is caused by a behavioural response or due to technical factors. The Home Energy Efficiency Database (HEED) sheds some light, however a number of data quality issues exist[9], making firm inference problematic. This database allows us to break down heating system replacements funded through the supplier obligations by tenure type. Information is also provided on whether households had heating control systems installed along with their heating replacement. From Table 6 it is clear that a very low proportion of condensing boiler replacements were accompanied by installation of heating controls during this period. The proportion of households with heating system controls is much higher for "Owner Occupier" than for other categories, apart from "Unknown". Given that many more deprived households live in categories such as "Privately rented", "Rented from a housing association", "Rented

[9]Not least the fact that a considerable proportion of heating control installations are in dwellings of "Unknown" tenure

from Local Authority" and "Social housing" it is quite possible that this is a contributing factor to the smaller and less persistent savings that these households experience.

Table 6: Heating system replacement and heating control installation 2002-2012

Heating system measure	Condensing Boiler	Condensing Boiler (Intelligent Controls)	Percentage with heating controls
Other tenure	1,001	8	0.80%
Owner Occupier	117,737	4,167	3.54%
Privately rented	33,020	102	0.31%
Rented from a housing association	11,272	22	0.20%
Rented from Local Authority	21,455	27	0.13%
Social housing	15,402	25	0.16%
Unknown	14,307	2,941	20.56%
Grand Total	214,194	7,292	3.40%

Source: Home Energy Efficiency Database (HEED). Energy Savings Trust

5.3 Cost effectiveness of measures

5.3.1 Estimated costs to suppliers and private costs of measures

Taking our estimates of both the ATT and the impact over time for different households we can develop more realistic assessments of the cost-effectiveness of measures than have been previously calculated. In order to do this, we need information on the costs. Cost estimates can be difficult to obtain and exhibit a wide degree of variation exists. To that end some assumptions must be made in order to make some back-of-the-envelope calculations. Usefully, a number of published academic and policy papers provide cost-estimates. The estimates we present in Table 7 are based on a range of previous studies, outlined in more detail in Tables F1, F2 and F3. The costs we present are the costs incurred by the energy companies in installing each measure. These may understate the actual costs in some cases, and can thus be considered a lower bound. For example, the cost for replacement heating system used (£200) for past policy evaluations is the assumed additional cost of installing a high efficiency system, over and above a typical lower-efficiency system. We consider this estimate to be a lower bound. The Energy Savings Trust estimate the costs of boiler replacements to range from £700 to £6,000 (EST, 2013). On average the installation of condensing boiler costs around £2,400 per dwelling. In a 2015 review Frontier Economics assumed that the fixed cost of for gas-fired condensing boilers lie between £2,200-3,000 (Economics, 2015). For the purposes of comparison we consider the upper bound for a replacement heating system installation to have been £2000 in 2006.

Table 7: Cost assumptions for each measure

Measure	Cost assumptions (GBP)
Cavity wall insulation	350
Loft insulation	285
Replacement boiler lower (policy cost)	200
Replacement boiler upper (private cost)	2000

Source: Based on Lees (2006, 2008); Shorrock et al. (2005)

5.3.2 Internal rate of return (IRR)

The internal rate of return (IRR) on a project is the discount rate (r) that yields a net-present value of zero, or the discount rate at which the average value of avoided discounted future energy costs equals the upfront investment cost. Formally, this can be calculated using the below formula:

$$NPV = \sum_{t=1}^{T} \frac{C_t}{(1+r)^t} - C_0 \tag{8}$$

Where T is the estimated lifespan of the measure, C_t are the avoided energy costs in year t, C_0 is the upfront investment cost and r is the IRR which we solve for. The IRR is calculated based on the econometric estimates we observe, for varying estimated lifespans of measures and assuming constant future energy prices. While these measures were largely funded by the energy companies, it is useful to estimate the private returns as this would be a critical factor in determining uptake in the absence of such schemes.

Table 8 presents estimated IRRs for each measure, calculating the IRR for 10, 20 and 30 years. Our preferred estimated lifespan is 10 years for replacement heating system and 30 years for both types of insulation. Assuming a lifespan of 30 years or more, cavity wall insulation is an attractive investment yielding a return of 16 percent. Loft insulation is less attractive yielding 6 percent. Whether to invest in a heating system depends greatly on whether one uses the estimated policy cost or private cost. The low and highly negative returns in some cases suggest that there may not be much of an energy efficiency gap to explain with regard to replacement heating systems.

Table 8: IRR for each measure

	Cavity wall	Loft	BoilerL (policy cost)	BoilerU (private cost)
IRR 10	7%	-8%	17%	-24%
IRR 20	15%	4%	23%	-7%
IRR 30	16%	6%	23%	-2%

Source: Author's calculations

The next set of results, presented in Table 9 presents IRR estimates for each measure and each IMD group, along with the sample average for comparison purposes. A considerable degree of variation exists around the sample average, with households living in more deprived areas experiencing much lower returns than those in more affluent areas.

Table 9: IRR for each measure and IMD group

	Sample Average	IMD1	IMD2	IMD3	IMD4	IMD5
Cavity wall	16%	11%	14%	16%	17%	18%
Loft	6%	5%	5%	5%	7%	9%
BoilerL	17%	12%	14%	17%	20%	22%
BoilerU	-24%	-26%	-25%	-24%	-23%	-22%

Source: Author's calculations

The final set of IRR estimates we present, adjusts the future energy savings from a heating system replacement to correspond with the observed estimates in Figure 8. In this case, savings erode more quickly over time for households living in more deprived areas. Taking this in account results in a further reduction in the IRR for lower income households.

Table 10: IRR for each measure and IMD group adjusting for time-path of energy savings

	IMD1	IMD2	IMD3	IMD4	IMD5
BoilerL	6%	12%	16%	20%	19%
BoilerU	-29%	-26%	-23%	-21%	-22%

Source: Author's calculations

5.3.3 Other measures of cost-effectiveness

To broaden the perspective somewhat we also consider two other measures of cost-effectiveness: the cost per tonne of CO2 removed and the cost per kWh of energy saved. These are calculated at the sample average and allow a comparison of the overall cost of these policies with other similar initiatives. The estimated cost per kWh of energy saved is calculated by summing up the annual estimated savings over the expected lifetime of the measure. To calculate the cost per tonne of CO2 removed, we convert our kWh estimates based on the estimated CO2 produced in consuming one kWh of gas and electricity based on DEFRA/DECC greenhouse gas conversion factors. These are reported in Table 11, detailed information on the underlying assumptions is presented in Table G1.

Cavity wall insulation is the most cost-effective measure, followed by loft insulation and replacement heating systems. Relative to the estimated social cost of carbon and natural gas prices, insulation and the lower estimate for replacement heating systems seem relatively cost effective. However, at the upper bound of replacement heating system cost it does not represent an attractive investment.

Table 11: Cost-effectiveness of each measure

	GBP per tonne of CO2	**GBP per kWh**
Cavity wall insulation	36	0.0072
Loft insulation	90	0.0171
BoilerL	60	0.0141
BoilerU	600	0.1412

Source: See Table G1 for detail on underlying cost assumptions

Converting the cost per kWh saved in 2000 GBP to 2015 USD it is possible to compare the cost of these interventions with a wide range of other initiatives, such as behavioural programmes, building code changes, subsidies and information provision. These are based on a recent review paper by Gillingham et al. (2018). Table H1 demonstrates that these measures, and in particular cavity wall insulation, compare quite favourably with most other initiatives.

6 Sensitivity analysis and robustness

This section presents the results of a range of sensitivity analyses and robustness tests. All results are presented in Tables D1 and E1 in the Appendices. The aim is to present information on the stability of the results across a range of sample splits and alternative specifications.

Considering Table D1 first, Column 1 presents results when the entire sample is included without using a matched control group. This is useful as it provides quantification of the degree of bias that might exist should selection not be controlled for. The coefficients for both type of insulation remain consistent with our central estimates, however heating system replacement is considerably lower. Column 2 performs the same estimation including only dwellings in which natural gas is the primary source of cental heating. This reduces the sample size slightly and results in a slight improvement in model fit (as measured by R^2), and an increase in size of the coefficient on replacement heating system. Columns 3 and 4 repeat this analysis however we now introduce a matched control group in both cases. Column 5 presents results using a matched control group, where natural gas is the primary source of central heating. However, in this specification we also include electricity consumption as an explanatory variable. This is likely endogenous as households could substitute between both types of heating, and should be treated with caution. However, this variable contains much information about household behaviour, appliance usage and occupancy that is otherwise omitted in our estimations, and is interesting to consider for that reason - particularly if it has an effect on our main results. Inclusion of this variable does not affect our central estimates giving us confidence in their accuracy.

Taking the sample used to estimate Column 4 results (matched upgrade and control group, primary heating is gas), we now estimate a number of additional specifications in which we exclude households with large annual changes in energy consumption. This is to control for any large changes that might be as a result of changes in occupancy. Column's (6-8) present these results in which we exclude dwellings with any annual changes in either gas or electricity greater than ± 50, 40 and 30 percent respectively of the previous year's consumption. When this sensitivity check is included, sample size drops significantly and model fit improves considerably ($R^2 = 0.398$). The coefficients on both insulation measures do not change much, but the effect of heating system replacement increases. The results in column's (6-8) are otherwise consistent with each other and for this reason we favour Column 8 as the preferred specification as this maximises sample size while also accounting for potential unobserved occupancy changes. Finally, in Column 9 we again include electricity consumption as an explanatory variable and results remain largely unchanged.

Finally, Table E1 provides a comparison of our preferred estimates using Coarsened-Exact Matching with results generated using both Nearest Neighbour and Kernel Matching techniques. The results are stable across all matching procedures.

7 Discussion and conclusions

The aim of this research was to estimate the extent to which heterogeneity and persistence effect the returns to commonly installed energy efficiency measures, and how this ultimately impacts household incentives and the cost-effectiveness of policies. By combining statistical matching and a range of panel econometric estimators we control for unobserved heterogeneity and selection into various government schemes which funded the upgrades. Leveraging an extremely large database of energy efficiency measures and metered consumption allows us to systematically explore heterogeneity. The database includes the universe of households entering into energy efficiency schemes administered by energy suppliers in the UK, mitigating site-selection bias.

Our results indicate that cavity wall insulation and heating system replacement (installation of a condensing gas boiler) result in an energy saving of about 10 percent of annual consumption, while loft insulation results in approximately a three percent reduction. These savings are consistent regardless of when the measures were installed over the sample period. Households living in more deprived areas observe less savings (both in absolute and percentage terms) than those in more affluent areas. This result is true for all measures examined. In addition to this, savings from heating system replacements erode quickly over time for the most deprived households, but remain stable for more affluent households. As far as we are aware this result has not been shown before. It is not entirely clear what is causing this result. Due to changes in the UK building regulations in 2005, all boiler replacements we observe are required to be of 86% or higher efficiency. Therefore, this finding would not appear to be as a result of differences in quality of the system. However, heating system controls appear to have been installed less frequently for lower income households, and this could be a contributing factor. Condensing boilers require regular servicing. If servicing rates differed systematically across the population this might also have an impact on the results. Along with potential technical explanations, behavioural factors might also be driving some of this effect. We could be observing a delayed rebound effect as households become accustomed to warmer internal temperatures over time, and the reduced cost of heating services allow those previously income constrained households greater thermal comfort.

The results of this analysis provide both academic and policy insights. Our academic contribution is to provide new evidence on the performance of energy efficiency measures over time allowing us to better quantify the size of the energy-efficiency gap. As far as we are aware this is the first paper to examine longer run effects and how they vary by levels of household deprivation. This adds to a growing literature examining longer run effects of building energy codes (Kotchen, 2017) and information stimuli (Allcott and Rogers, 2014).

While we cannot identify the precise source of the savings erosion, we can quantify how it affects the financial return on investment. This provides important insights for both policy design and policy evaluation. Pay-as-you-save financing mechanisms are becoming increasingly popular for energy efficiency. For example, the Green Deal was a recent policy initiative in the UK (2011-2015) which provided households with loans in order to finance energy efficiency measures at interest rates of approximately eight percent. This was widely considered to have been a spectacular failure. The National Audit Office conducted an independent audit of the Green Deal scheme, finding that during it's lifespan the scheme only funded one percent of energy efficient measures installed nationally. It also found that the scheme saved negligible amounts of CO2 and that households did not see the loans as an attractive proposition. Concerns were raised prior to the Green Deal policy that it would not have sufficient appeal for householders. These relate to a range of factors, including uncertainty regarding energy savings, limited financial appeal, and limited awareness of the scheme (Dowson et al., 2012). A key factor in limiting its appeal were the high rates of interest charged on loans (Rosenow and Eyre, 2016). Given the results we observe, it is clear that this rate is not sufficiently low to provide incentives for many households to partake in this scheme. In particular, low income households would actually lose money by making these improvements unless energy prices rise significantly. Market-based interventions will only work for certain segments of the population and policy needs to take this into account.

A question one might ask is why lower income households experience lower rates of return? A body of literature on the rebound effect identifies changes in energy service consumption that might reduce the expected savings (Sorrell et al., 2009a). It has also been shown that significant heterogeneity in the rebound effect is determined by household income and wealth (Aydin et al., 2017).[10]. A naive assessment might suggest that environmental policy is more effective when focused on better-off households. However, this is true only in the narrow sense that upgrades to such households will be more effective at reducing energy use and carbon emissions. Total welfare gains from upgrades may well be greater for upgrades to low income households, particularly if lower income households trade-off lower energy savings with increased internal temperatures and this results in improved well-being and even health outcomes. Evidence suggests this might be the case with regard to health (Wilkinson et al., 2009; Hamilton et al., 2015) and poverty alleviation (Hills, 2012; Watson and Maitre, 2015).

Ultimately, this research helps to improve our understanding of the incentives faced by households when making energy efficiency investments, provides a methodology for better evaluating the cost-effectiveness of public policies and raises new concerns over the how the costs and benefits of policies are distributed. It also suggests that the energy-efficiency gap requires less explanation than some would suggest. At an individual household

[10]This has also been shown in an inter-temporal sense by Fouquet et al. (2018)

level, the private benefits of energy efficiency investments need to be re-considered with a greater focus on the non-financial benefits. While at a societal level a greater focus on carbon emissions reduction, as opposed to cost-savings is required.

References

Adan, H. and Fuerst, F. (2015). Do energy efficiency measures really reduce household energy consumption? a difference-in-difference analysis. *Energy Efficiency*, 9(5):1207–1219.

Alberini, A. and Towe, C. (2015). Information v. energy efficiency incentives: Evidence from residential electricity consumption in maryland. *Energy Economics*, 52:S30–S40.

Alberini, A., Towe, C., and Gans, W. (2014). Energy efficiency incentives: Do they work for heating and cooling equipment? evidence from maryland homeowners. *The energy journal*.

Allcott, H. (2011). Social norms and energy conservation. *Journal of public Economics*, 95(9-10):1082–1095.

Allcott, H. (2015). Site selection bias in program evaluation. *The Quarterly Journal of Economics*, 130(3):1117–1165.

Allcott, H. and Greenstone, M. (2012). Is there an energy efficiency gap? *Journal of Economic Perspectives*, 26(1):3–28.

Allcott, H. and Greenstone, M. (2017). Measuring the welfare effects of residential energy efficiency programs. Technical report, National Bureau of Economic Research.

Allcott, H. and Rogers, T. (2014). The short-run and long-run effects of behavioral interventions: Experimental evidence from energy conservation. *American Economic Review*, 104(10):3003–3037.

Angrist, J. D. and Pischke, J.-S. (2009). *Mostly harmless econometrics: an empiricist's companion*. Princeton University Press, Princeton.

Austin, P. C. (2009). Balance diagnostics for comparing the distribution of baseline covariates between treatment groups in propensity-score matched samples. *Statistics in medicine*, 28(25):3083–3107.

Aydin, E., Kok, N., and Brounen, D. (2017). Energy efficiency and household behavior: the rebound effect in the residential sector. *The RAND Journal of Economics*, 48(3):749–782.

Ayres, I., Raseman, S., and Shih, A. (2013). Evidence from two large field experiments that peer comparison feedback can reduce residential energy usage. *The Journal of Law, Economics, and Organization*, 29(5):992–1022.

BEIS (2013). Fuel poverty statistics.

Bertoldi, P. and Rezessy, S. (2008). Tradable white certificate schemes: fundamental concepts. *Energy efficiency*, 1(4):237–255.

Blackwell, M., Iacus, S. M., King, G., and Porro, G. (2009). cem: Coarsened exact matching in stata. *The Stata Journal*, 9(4):524–546.

Blumstein, C., Krieg, B., Schipper, L., and York, C. (1980). Overcoming social and institutional barriers to energy conservation. *Energy*, 5(4):355–371.

Brounen, D., Kok, N., and Quigley, J. M. (2012). Residential energy use and conservation: Economics and demographics. *European Economic Review*, 56(5):931–945.

Brucal, A., Roberts, M., et al. (2017). Do energy efficiency standards hurt consumers? evidence from household appliance sales. Technical report, Grantham Research Institute on Climate Change and the Environment.

Burlig, F., Knittel, C., Rapson, D., Reguant, M., and Wolfram, C. (2017). Machine learning from schools about energy efficiency. Technical report, National Bureau of Economic Research.

Caliendo, M. and Kopeinig, S. (2005). Some Practical Guidance for the Implementation of Propensity Score Matching. *Journal of Economic Surveys (2008)*, 22(1):31–72.

Chawla, M., Pollitt, M. G., et al. (2013). Energy-efficiency and environmental policies & income supplements in the uk: evolution and distributional impacts on domestic energy bills. *Economics of Energy & Environmental Policy*, 2(1).

Cohen, F., Glachant, M., and Söderberg, M. (2017). Consumer myopia, imperfect competition and the energy efficiency gap: Evidence from the uk refrigerator market. *European Economic Review*, 93:1–23.

Davis, L. W., Fuchs, A., and Gertler, P. (2014). Cash for coolers: evaluating a large-scale appliance replacement program in mexico. *American Economic Journal: Economic Policy*, 6(4):207–38.

DECC (2016). Domestic NEED: Methodology Note. Technical report.

Dowson, M., Poole, A., Harrison, D., and Susman, G. (2012). Domestic uk retrofit challenge: Barriers, incentives and current performance leading into the green deal. *Energy Policy*, 50:294–305.

Duffy, A. (2013). The final report of the community energy saving programme (cesp) 2009-2012. *Ofgem, UK.*

Economics, F. (2015). Energy efficiency: an infrastructure priority. *London: Frontier Economics. http://www. frontier-economics. com/documents/2015/09/energy-efficiency-infrastructurepriority. pdf.*

EST (2013). Review of potential for carbon savings from residential energy efficiency. *Final report for The Committee on Climate Change., Cambridge.*

Ferraro, P. J. and Miranda, J. J. (2017). Panel data designs and estimators as substitutes for randomized controlled trials in the evaluation of public programs. *Journal of the Association of Environmental and Resource Economists*, 4(1):281–317.

Fouquet, R. et al. (2018). Consumer surplus from energy transitions. *The Energy Journal*, 39(3).

Fowlie, M., Greenstone, M., and Wolfram, C. (2018). Do energy efficiency investments deliver? evidence from the weatherization assistance program. *The Quarterly Journal of Economics*, 133(3):1597–1644.

Gerarden, T., Newell, R. G., and Stavins, R. N. (2015). Deconstructing the energy-efficiency gap: conceptual frameworks and evidence. *The American Economic Review*, 105(5):183–186.

Gibbons, C. E., Urbancic, M. B., Anderson, M., Auerbach, A., Andrews, R., Bitler, M., Bohn, H., Buchinsky, M., Bugni, F., Cameron, C., Dobkin, C., Khan, S., Kasy, M., Kline, P., Kodrzycki, Y., Lozano, F., Masten, M., Maurel, A., Miller, D., Montoy, J. C., Moretti, E., Oaxaca, R., Raphael, S., Rosen, A., Shapiro, J., and Sekhon, J. (2014). Broken or fixed effects? NBER Working Paper No. 20342. *NBER Working Paper No. 20342.*

Gillingham, K., Keyes, A., and Palmer, K. (2018). Advances in evaluating energy efficiency policies and programs. *Annual Review of Resource Economics*, (0).

Giraudet, L.-G., Bodineau, L., and Finon, D. (2012). The costs and benefits of white certificates schemes. *Energy Efficiency*, 5(2):179–199.

Goldberg, P. K. (1998). The effects of the corporate average fuel efficiency standards in the us. *The Journal of Industrial Economics*, 46(1):1–33.

Golove, W. H. and Eto, J. H. (1996). Market barriers to energy efficiency: a critical reappraisal of the rationale for public policies to promote energy efficiency. *LBL-38059. Berkeley, CA: Lawrence Berkeley National Laboratory.*

Hamilton, I., Milner, J., Chalabi, Z., Das, P., Jones, B., Shrubsole, C., Davies, M., and Wilkinson, P. (2015). Health effects of home energy efficiency interventions in england: a modelling study. *BMJ Open*, 5(4):e007298–e007298.

Hamilton, I. G., Shipworth, D., Summerfield, A. J., Steadman, P., Oreszczyn, T., and Lowe, R. (2014). Uptake of energy efficiency interventions in english dwellings. *Building Research & Information*, 42(3):255–275.

Hamilton, I. G., Summerfield, A. J., Shipworth, D., Steadman, J. P., Oreszczyn, T., and Lowe, R. J. (2016). Energy efficiency uptake and energy savings in english houses: A cohort study. *Energy and Buildings*, 118:259–276.

Harold, J., Lyons, S., and Cullinan, J. (2015). The determinants of residential gas demand in Ireland. *Energy Economics*, 51:475–483.

Hausman, J. A. (1979). Individual discount rates and the purchase and utilization of energy-using durables. *Bell Journal of Economics*, 10(1):33–54.

Hills, J. (2012). Getting the measure of fuel poverty: Final report of the fuel poverty review.

Hough, D. and Page, R. (2015). Eco, the energy company obligation. *Briefing Paper: Number CBP*, 6814:1–32.

Hunt, L. C. and Ryan, D. L. (2015). Economic modelling of energy services: Rectifying misspecified energy demand functions. *Energy Economics*, 50:273–285.

Iacus, S. M., King, G., and Porro, G. (2008). Matching for causal inference without balance checking.

Iacus, S. M., King, G., and Porro, G. (2012). Causal inference without balance checking: Coarsened exact matching. *Political analysis*, 20(1):1–24.

Imai, K., King, G., and Stuart, E. A. (2008). Misunderstandings between experimentalists and observationalists about causal inference. *Journal of the royal statistical society: series A (statistics in society)*, 171(2):481–502.

Jaffe, A. B. and Stavins, R. N. (1994). The energy paradox and the diffusion of conservation technology. *Resource and Energy Economics*, 16(2):91–122.

Kotchen, M. J. (2017). Longer-run evidence on whether building energy codes reduce residential energy consumption. *Journal of the Association of Environmental and Resource Economists*, 4(1):135–153.

Lancaster, K. J. (1966). A new approach to consumer theory. *Journal of political economy*, 74(2):132–157.

Lees, E. (2006). Evaluation of the energy efficiency commitment 2002-05. *DEFRA, Wantage (Oxfordshire).*

Lees, E. (2008). Evaluation of the energy efficiency commitment 2005-08. *DEFRA, Wantage (Oxfordshire).*

Nordhaus, W. D. (1996). Do real-output and real-wage measures capture reality? the history of lighting suggests not. In *The economics of new goods*, pages 27–70. University of Chicago Press.

Novan, K., Smith, A., and Zhou, T. (2017). Residential building codes do save energy: Evidence from hourly smart-meter data. *UC Davis.*

Ofgem (2013). The final report of the Community Energy Saving Programme (CESP) 2009-2012. *Ofgem.*

ONS (2011). Office of National Statistics. Technical Report: Welsh Index of Multiple Deprivation 2011. Technical report.

Ori, S. (2017). Why Government Energy-Efficiency Programs Sound Great But Often Dont Work. *The Wall Street Journal.*

Payne, R. and Abel, G. (2012). UK indices of multiple deprivation - a way to make comparisons across constituent countries easier. *Health Statistics Quarterly, ONS*, 53:22–38.

Quigley, J. M. and Rubinfeld, D. L. (1989). Unobservables in consumer choice: residential energy and the demand for comfort. *The Review of Economics and Statistics*, pages 416–425.

Rosenow, J. (2012). Energy savings obligations in the uka history of change. *Energy Policy*, 49:373–382.

Rosenow, J. and Eyre, N. (2016). A post mortem of the green deal: Austerity, energy efficiency, and failure in british energy policy. *Energy Research & Social Science*, 21:141–144.

Ryan, L. and Campbell, N. (2012). Spreading the net: the multiple benefits of energy efficiency improvements.

Shorrock, L., Henderson, J., and Utley, J. (2005). Reducing carbon emissions from the uk housing stock. *BRE Report BR480.*

Sorrell, S., Dimitropoulos, J., and Sommerville, M. (2009a). Empirical estimates of the direct rebound effect: A review. *Energy policy*, 37(4):1356–1371.

Sorrell, S., Harrison, D., Radov, D., Klevnas, P., and Foss, A. (2009b). White certificate schemes: Economic analysis and interactions with the eu ets. *Energy Policy*, 37(1):29–42.

Watson, D. and Maitre, B. (2015). Is fuel poverty in ireland a distinct type of deprivation? *The Economic and Social Review*, 46(2, Summer):267–291.

Wilkinson, P., Smith, K. R., Davies, M., Adair, H., Armstrong, B. G., Barrett, M., Bruce, N., Haines, A., Hamilton, I., Oreszczyn, T., et al. (2009). Public health benefits of strategies to reduce greenhouse-gas emissions: household energy. *The Lancet*, 374(9705):1917–1929.

Wooldridge, J. (2010). *Econometric analysis of cross section and panel data.* MIT Press, Cambridge, Mass., 2nd ed. edition.

8 Appendix

8.1 Overview of NEED dataset

Table A1: Description of variables in NEED

Variable	Description
HH_ID	Household identifier. A unique value for each record. Created specifically for these datasets.
REGION	Region code - formally Government Office Region. See ONS website for more details: http://www.ons.gov.uk/ons/guide-method/geography/beginner-s-guide/administrative/england/government-office-regions/index.html.
IMD_ENG	English Index of multiple deprivation 2010. Households are allocated to five groups (quintiles) based on the deprivation rank of the 2001 Lower Layer Super Output Area (LSOA) they are located in. Households in the 20 per cent most deprived LSOAs are in the bottom quintile (1) and households in the 20 per cent least deprived LSOAs are in the top quintile (5).
IMD_WALES	Welsh Index of multiple deprivation 2011. Households are allocated to one of five bands based on the deprivation rank of the LSOA (2001) they are located in. 1, most deprived, 5 least deprived.
FP_ENG	EUL only. Fuel Poverty Indicator for England. Households are allocated to one of five bands based on the estimate of the proportion of households in fuel poverty in the LSOA (2001) they are located in. Uses the 2011 estimates of fuel poverty low income high cost definition.
EPC_INS_DATE	EUL Only. Provides information on the date of the EPC inspection (based on lodgement date).
GconsYEAR	Weather corrected annual gas consumption. Based on meter point data from Xoserve and independent gas transporters. Readings relate to October to September each year (e.g. 2012 consumption is October 2011 to September 2012). See here for more information on this source: https://www.gov.uk/government/publications/regional-energy-data-guidance-note.
GconsYEARValid	Flag indicates records with valid consumption and households off the gas network.
EconsYEAR	Annual electricity consumption in kWh - values relate to end January to end January each year (e.g. 2012 consumption is end January 2012 to end January 2013). See here for more information on this source: https://www.gov.uk/government/publications/regional-energy-data-guidance-note.
EconsYEARValid	Valid electricity consumption (between 100 and 25,000 inclusive)
E7Flag2012	Shows whether the electricity meter is an E7 (profile 2) meter - this does not necessarily mean the household has an E7 tariff, some households will have an E7 meter without an E7 tariff.
MAIN_HEAT_FUEL	Main fuel used to heat the property, based on information from Energy Performance Certificate
PROP_AGE	Banded year of construction based on EPC data.
PROP_TYPE	Type of property (based on combination of EPC built form and property type).
FLOOR_AREA_BAND	Banded floor area based on EPC (m2).
EE_BAND	Energy Efficiency Band Based on EPC (A and B grouped).
LOFT_DEPTH	Amount of loft insulation as assessed by EPC (all properties with loft insulation recorded as installed through a Government scheme are assigned 2 irrespective of EPC information). No information could occur where the information is missing from the EPC or where the property does not have a loft.
WALL_CONS	Wall construction as recorded on EPC.
CWI	Cavity wall insulation installed through Government schemes. This includes measures recorded as installed on HEED, including, Energy Efficiency Commitment, Community Energy Savings Programme and Carbon Emissions Reduction Target.
CWI_YEAR	Year of CWI installation
LI	Loft insulation installed through Government schemes. This includes measures recorded as installed on HEED, including, Energy Efficiency Commitment, Community Energy Savings Programme and Carbon Emissions Reduction Target.
LI_YEAR	Year of LI installation
BOILER	This includes boilers installed through Government schemes, and those registered by CORGI (up to 2009) and Gas Safe (2009 onwards).
BOILER_YEAR	Year of Boiler installation
WEIGHT	EUL Only. Weighting based on Region, property age, property type and floor area band. Summing all weights gives (approximate) total number of households in England and Wales 2011.

Table A2: Summary statistic

Variable	Category	n untreated	n treated	Total n	Mean Gas	Mean Electr.
Regions	North East	83,698	115,415	199,113	15510.84	3462.747
	North West	236,794	266,874	503,668	15208.41	3900.542
	Yorkshire (a.t.H.)	188,290	201,006	389,296	15493.32	3813.363
	East Midlands	174,550	179,017	353,567	15177.16	4003.458
	West Midlands	199,366	185,984	385,350	15007.05	4146.137
	East of England	246,651	184,685	431,336	15068.45	4371.629
	London	389,548	185,003	574,551	15372.6	4135.165
	South East	363,751	270,953	634,704	15507.96	4382.642
	South West	229,357	176,074	405,431	13845.36	4457.526
	Wales	105,071	104,361	209,432	14676.42	3789.692
FP	1	515,635	359,281	874,916	13103.34	3998.862
	2	409,940	374,097	784,037	15036.81	4164.775
	3	396,893	363,103	759,996	15661.29	4219.243
	4	375,062	336,290	711,352	16214.88	4197.805
	5	414,474	332,241	746,715	16104.2	4078.469
IMD	1	450,616	455,048	905,664	12817.86	3603.685
	2	469,230	377,913	847,143	14104.08	3890.909
	3	466,882	353,377	820,259	15181.31	4246.657
	4	431,621	338,563	770,184	16325	4429.783
	5	398,727	344,471	743,198	17867.73	4495.282
Age	before 1930	606,589	391,713	998,302	16716.21	4303.14
	1930-1949	251,402	314,856	566,258	16863.97	4047.436
	1950-1966	290,718	431,623	722,341	14515.74	3873.33
	1967-1982	369,192	436,771	805,963	13831.01	3931.536
	1983-1995	299,008	201,981	500,989	13410.3	4250.797
	1996 onwards	400,167	92,428	492,595	13966.06	4287.513
Type	Detached	295,374	251,537	546,911	21843.34	5448.197
	Semi-detached	408,987	514,204	923,191	16433.14	4138.248
	End-Terrace	191,830	186,240	378,070	14938.36	3952.917
	Mid-Terrace	450,951	385,582	836,533	13771.39	3729.568
	Bungalow	138,088	238,298	376,386	15926.08	3978.159
	Flat	731,846	293,511	1,025,357	9899.633	3788.039
Floor	1 to 50 m2	361,530	167,965	529,495	8516.688	3344.52
	51-100 m2	1,334,455	1,259,914	2,594,369	13586.52	3790.722
	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]	[illegible]

B Balance tables for matched variables

Table B1: Unmatched sample

Unmatched sample		Treated			Control			Balance	
	Variable	Mean	Variance	Skewness	Mean	Variance	Skewness	Std-diff	Var-ratio
Variables used in matching	prop_age	2.96	1.98	0.16	3.00	3.00	0.31	-0.03	0.66
	imd_both	2.85	2.11	0.15	2.96	2.01	0.05	-0.08	1.05
	region	5.34	7.33	0.03	5.81	6.31	-0.29	-0.18	1.16
	fuel_type	0.98	0.02	-7.43	0.98	0.02	-6.33	0.04	0.74
	Gcons2005	18124	78900000	0.65	17394	86200000	0.73	0.08	0.92
Variables not used in matching	prop_type	3.33	2.63	0.22	3.56	2.92	0.08	-0.14	0.90
	floor_area	2.20	0.40	0.89	2.20	0.46	0.77	-0.01	0.86
	loft_depth	2.03	0.28	0.04	2.08	0.53	-0.13	-0.08	0.52
	wall_cons	0.73	0.20	-1.02	0.59	0.24	-0.36	0.29	0.82
	FP_ENG	2.95	1.97	0.06	2.89	2.12	0.10	0.04	0.93
	Econs2005	3903	7653561.00	2.16	3998.54	8374713	2.14	-0.03	0.91

Table B2: Matched sample all years

All years matched sample		Treated			Control			Balance	
	Variable	Mean	Variance	Skewness	Mean	Variance	Skewness	Std-diff	Var-ratio
Variables used in matching	prop_age	2.91	2.32	0.24	2.91	2.31	0.24	0.00	1.00
	imd_both	2.92	2.07	0.09	2.92	2.07	0.09	0.00	1.00
	region	5.62	6.83	-0.15	5.62	6.82	-0.15	0.00	1.00
	fuel_type	0.98	0.02	-7.45	0.98	0.02	-7.47	0.00	1.01
	Gcons2005	18020	84200000	0.67	18017	84300000	0.67	0.00	1.00
Variables not used in matching	prop_type	3.36	2.68	0.19	3.48	2.87	0.14	-0.07	0.93
	floor_area	2.21	0.42	0.86	2.21	0.45	0.80	0.00	0.93
	loft_depth	2.04	0.30	0.03	2.05	0.52	-0.08	-0.02	0.57
	wall_cons	0.67	0.22	-0.71	0.63	0.23	-0.52	0.09	0.95
	FP_ENG	2.95	2.04	0.06	2.96	2.09	0.05	-0.01	0.98
	Econs2005	3945.89	7999389.00	2.15	4028.94	8182317.00	2.09	-0.03	0.98

Table B3: Percentage of observations matched

Dwellings receiving upgrades	Count
Full database	1,869,372
2005 or unknown upgrade date	416,994
Remaining sample	1,452,378
Matched sample	1,286,419
Unmatched	165,959
Matched as a percentage of eligible	89%

C Results in kWh

Table C1: Results in units of energy saved (kWh)

	(1) Gas	(2) Elec	(3) Tot Energy
Cavity wall insulation	-1254.909***	-23.002	-1277.911***
	(41.989)	-14.009	(46.216)
Loft insulation	-375.539***	-8.901	-384.440***
	(36.250)	-11.268	(39.533)
Replacement boiler	-1229.965***	-98.273***	-1328.238***
	(32.918)	-10.796	(35.722)
Control variables	Y	Y	Y
Household fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
Year*region fixed effects	Y	Y	Y
Observations	549072	549072	549072
Number of households	68,634	68634	68634
R squared	0.337	0.092	0.357

Notes: This table reports coefficient estimates and standard errors from three separate regressions. The dependent variable in all regressions is annual energy consumption. Column(1) is Gas, (2) Electricity and (3) Total Energy. For each upgrade group a matched control group is created using coarsened-exact matching. The sample includes billing records from 2005 to 2012. Standard errors are clustered at the household level. Triple asterisks denote statistical significance at the 1% level; Double asterisks at the 1% level; single asterisks at the 10% level.

D Results from alternative specifications

Table D1: Results from alternative specifications

	(1) Full sample	(2) Only gas	(3) Matched sample	(4) Only gas and matched	(5) Only gas, matched and elec	(6) 50 drop	(7) 60 drop	(8) 70 drop	(9) 70 drop, elec
Cavity wall insulation	-0.092***	-0.092***	-0.083***	-0.084***	-0.083***	-0.095***	-0.096***	-0.094***	-0.092***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Loft insulation	-0.025***	-0.026***	-0.018***	-0.019***	-0.020***	-0.029***	-0.030***	-0.030***	-0.029***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
Replacement boiler	-0.055***	-0.062***	-0.038***	-0.045***	-0.049***	-0.090***	-0.092***	-0.092***	-0.091***
	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)	(0.001)
					0.179***				0.138***
					(0.000)				(0.001)
Control variables	Y	Y	Y	Y	Y	Y	Y	Y	Y
Household fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Year*region fixed effects	Y	Y	Y	Y	Y	Y	Y	Y	Y
Observations	20,180,976	19,886,106	14,090,155	13,912,535	13,912,535	3,516,155	4,530,404	5,502,936	5,502,936
Number of households	2,527,073	2,489,345	1,764,246	1,741,574	1,741,574	439,545	566,339	687,925	687,925
Number of households	0.115	0.118	0.115	0.118	0.168	0.398	0.375	0.349	0.369

Notes: This table reports coefficient estimates and standard errors from nine separate regressions. The dependent variable in all regressions is the logarithm of annual gas consumption in kilowatt hours. For each upgrade group a matched control group is created using coarsened-exact matching. The sample includes billing records from 2005 to 2012. Standard errors are clustered at the household level. Triple asterisks denote statistical significance at the 1% level; Double asterisks at the 1% level; single asterisks at the 10% level.

E Results from alternative matching procedures

Table E1: Results under different matching procedures

	(1) CEM Matching	(2) Nearest neighbour matching	(3) Kernel matching
Cavity wall insulation	-0.094***	-0.094***	-0.096***
	(0.001)	(0.002)	(0.003)
Loft insulation	-0.030***	-0.029***	-0.031***
	(0.001)	(0.003)	(0.003)
Replacement boiler	-0.092***	-0.093***	-0.097***
	(0.001)	(0.002)	(0.002)
Control variables	Y	Y	Y
Household fixed effects	Y	Y	Y
Year fixed effects	Y	Y	Y
Year*region fixed effects	Y	Y	Y
Observations	5,502,936	3,341,747	3,263,178
Number of households	687,925	417,751	407,928
R squared	0.349	0.363	0.411

Notes: This table reports coefficient estimates and standard errors from three regressions using alternative matching procedures. The dependent variable in all regressions is the logarithm of annual gas consumption in kilowatt hours. Column(1) reports the results from the preferred CEM procedure, Column (2) from Nearest Neighbour matching and Column (3) reports results from a Kernel matching procedure. Triple asterisks denote statistical significance at the 1% level; Double asterisks at the 1% level; single asterisks at the 10% level.

F Cost assumptions

Table F1: Sources of cost assumptions

	Low	High
Cavity wall (pre 1976)	300	325
Cavity wall (post 1976)	300	325
Loft 300mm (currently none)	138	273
Loft 300mm (currently 100mm)	86	211
Loft 300mm (currently 200mm)	35	170
Condensing boiler	100	300

Source: Shorrock et al. (2005)

Note: All costs in GBP

Table F2: Sources of cost assumptions

	EESOP1 (1994)	EESOP2	EESOP3	EEC1 (2005)
Cavity wall insulation	223	219	261	261
Condensing boiler	450	270	165	114

Source: Lees (2006)

Note: All costs in GBP

Table F3: Sources of cost assumptions

	Defra EEC1	Defra EEC2	Defra CERT	Lees 2005	Lees 2008
Cavity wall insulation	268	313	380	274	350
Loft insulation (top up)	213	260	286	217	275
Loft insulation (virgin)	213	260	286	252	295
A and B boiler	145			120	
A and B boiler and heating control	217			190	
All boilers			50		45

Source: Lees (2006, 2008)

Note: All costs in GBP

G Carbon saving assumptions

Table G1: Assumptions for CO2 saving calculation

	(1) Cavity wall insulation		(2) Loft insulation		(3) Replacement heating system	
	Gas	**Elec**	**Gas**	**Elec**	**Gas**	**Elec**
kg CO2 per (kWh)	0.18	0.52	0.18	0.52	0.18	0.52
Total annual saving (kWh)	1551.46	67.51	546.22	10.74	1363.85	52.29
Total annual saving (kgCO2)	284.85	35.42	100.29	5.63	250.40	27.43
Lifespan	30	30	30	30	12	12
Total lifetime savings (kgCO2)	8545.41	1062.54	3008.56	168.99	3004.82	329.16

Notes: Source: Based on DEFRA/DECC Greenhouse gas conversion factors 2011. Available at https://www.gov.uk/government/publications/2011-guidelines-to-defra-decc-s-greenhouse-gas-conversion-factors-for-company-reporting-methodology-paper-for-emission-factors

H Comparison of cost-effectiveness

Table H1: Comparison of cost-effectiveness

Intervention type	Reference	Evaluation type	Relevant subset	Percent reduction in energy usage	Engineering estimates of percent reduction in energy usage	Cost effectiveness (cents per kWh saved, 2015 USD)
Behavioral programs	Allcott (2011)	RCT	NA	2		3.6
	Allcott and Rogers (2014)	RCT	One-shot intervention			4.4
			Two-year intervention			1.1 to 1.8
			Four-year intervention			1.2 to 1.8
	Ayres et al. (2013)	RCT	Sacramento, California	2		5.5
			Puget Sound, Washington	1.2		2
Building codes	Novan et al. (2017)	RD analysis	NA	1.3	20	24.4
Efficient equipment or energy savings subsidy	Alberini and Towe (2015)	Matching	NA	5.3		3.9
	Alberini et al. (2014)	DID	Rebate of $1,000 or more	0		
			Rebate of $450	5.5		47.9
			Rebate of $300	6.2		28.2
	Burlig et al. (2017)	Machine learning	NA	2.9 to 4.5	11.6 to 18	
	Davis et al. (2014)	DID regression	Refrigerators	8		27.2
			Air conditioners	plus 1.7		4.5
Information provision	Alberini and Towe (2015)	Matching		5.5		
Supplier Obligation (TWC)	McCoy & Kotsch (2018)	Matching, FE regression	Cavity wall insulation	9.4	20.0	1.54 to 2.31
			Loft insulation	3	5.2	3.65 to 5.47
			Replacement heating system	9.2	24.9	3.02 to 30.19
Previous estimate of UK Supplier obligation (Lees, 2008))						1.92

Adapted from Gillingham et al. (2018)