Applied Energy 159 (2015) 589–600

Contents lists available at ScienceDirect

Applied Energy

journal homepage: www.elsevier.com/locate/apenergy

Explaining domestic energy consumption – The comparative contribution of building factors, socio-demographics, behaviours and attitudes

Gesche M. Huebner [a,*], Ian Hamilton [a], Zaid Chalabi [b], David Shipworth [a], Tadj Oreszczyn [a]

[a] UCL Energy Institute, 14 Upper Woburn Place, London WC1H 0AE, England, United Kingdom
[b] London School of Hygiene and Tropical Medicine, Keppel Street, London WC1E 7HT, England, United Kingdom

HIGHLIGHTS

- Building characteristics explain most variability in domestic energy consumption.
- Controlling for building factors, socio-demographics add little explanatory power.
- Attitudinal variables contribute very little to explaining energy consumption.
- Length of heating season is a significant predictor, even after controlling for region.
- Multicollinearity is a crucial issue in analysis of energy consumption.

ARTICLE INFO

Article history:
Received 31 July 2015
Received in revised form 4 September 2015
Accepted 5 September 2015
Available online 23 September 2015

Keywords:
Energy demand
Buildings
Socio-demographics
Behaviour
Lasso regression
Multicollinearity

ABSTRACT

This paper tests to what extent different types of variables (building factors, socio-demographics, attitudes and self-reported behaviours) explain annualized energy consumption in residential buildings, and goes on to show which individual variables have the highest explanatory power. In contrast to many other studies, the problem of multicollinearity between predictors is recognised, and addressed using Lasso regression to perform variable selection.

Using data from a sample of 924 English households collected in 2011/12, four individual regression models showed that building variables on their own explained about 39% of the variability in energy consumption, socio-demographic variables 24%, heating behaviour 14% and attitudes & other behaviours only 5%. However, a combined model encompassing all predictors explained only 44% of all variability, indicating a significant extent of multicollinearity between predictors. Once corrected for multicollinearity, building variables predominantly remained as significant predictors of energy consumption, in particular the dwelling's size and type. Of the sociodemographic predictors, only the household size remained significant, and of the heating behaviours only the length of heating season was significant. Reported beliefs about climate change were also a significant predictor.

The findings indicate that whilst people use energy, it is physical building characteristics that largely determine how much is used. This finding, together with the relatively greater time-invariant nature of building characteristics underlines their importance when focusing on seeking to understand residential energy consumption at a stock level. Retrofitting and behaviour change initiatives remain important avenues to reduce consumption, as suggested through the lower energy consumption associated with full double-glazing and shorter heating season. However, the dominance of building size also indicates that living in appropriately sized buildings is of great importance for energy consumption. The results also indicate that more than half of the variability in energy consumption cannot be explained, even when using a wide range of predictors. The paper also discusses the need to collect better occupant-related variables to give a correct representation of the impact of behaviour, such as heating demand temperatures. Furthermore, choices about dwelling characteristics could also be seen as a type of behaviour, even though it cannot be modelled in a cross-sectional analysis as used in this study.

* Corresponding author.
E-mail addresses: g.huebner@ucl.ac.uk (G.M. Huebner), i.hamilton@ucl.ac.uk (I. Hamilton), Zaid.Chalabi@lshtm.ac.uk (Z. Chalabi), d.shipworth@ucl.ac.uk (D. Shipworth), t.oreszczyn@ucl.ac.uk (T. Oreszczyn).

http://dx.doi.org/10.1016/j.apenergy.2015.09.028
0306-2619/© 2015 The Authors. Published by Elsevier Ltd.

1. Introduction

Energy use in buildings is one of the largest contributors to energy consumption both locally and globally. In the UK, residential buildings are responsible for about a quarter of total carbon emissions [1]. The 2008 Climate Change Act requires a 34% reduction in 1990 greenhouse gas emissions by 2020, and an 80% reduction by 2050. Dwellings are an important target area for emission reduction: The UK Government established the goal of reducing emissions from homes by 29% by 2020 [2]. Energy efficiency improvements in UK homes form a central part of the decarbonisation plans, with millions of retrofits of domestic homes planned over the next decades [3]. Achieving this improvement in energy efficiency requires having a better understanding of those drivers that have the greatest impact on energy demand.

Previous research has shown that building factors alone explain at least 40% of the variability in energy use, as summarised below. However, not all of the building predictors commonly examined can be impacted on through energy efficiency retrofits. Other factors, such as behaviours, are also widely considered to have a large impact on energy consumption and are likewise the subject of interventions such as feedback and use of social norms.

The aim of this paper is to show in a representative sample of the English housing stock (1) how much of the variability of residential energy consumption can be explained by different categories of predictors, contrasting the explanatory power of building variables, socio-demographics, self-reported heating behaviour, and attitudes towards energy, and (2) which individual predictors have the greatest impact on energy consumption. The findings are important to (1) understand which variables need to be measured to understand energy consumption, and (2) to shape interventions with the greatest potential impact.

The research presented carefully addresses the problem of multicollinearity, which occurs when two or more predictor variables in a multiple regression model are highly correlated. The presence of multicollinearity means that regression coefficients cannot be reliably interpreted. For each regression analysis, variance inflation factors are inspected to see if multicollinearity exists, and if it does, Lasso regression is carried out which sets redundant predictors to zero [4], therefore performing variable selection and removing multicollinearity.

The following review of existing research is almost entirely based on empirical estimates as opposed to modelled data. This is because research has repeatedly shown a gap between modelled and actual energy consumption of dwellings (e.g. [5–7]), and the aim of this paper is to identify factors determining *measured* energy consumption. For research on determinants of modelled and simulated energy consumption, see e.g. Aydinalp et al. [8], Kialashaki and Reisel [9], Koo et al. [10], or Martinaitis et al. [11]. In addition, this review does not include studies looking just at energy consumption excluding space heating (e.g. [12]), given the aim of understanding total energy consumption better which, at least in the UK, is largely driven by space heating.

1.1. Impact of building characteristics on energy use

Several studies have looked at the impact of building variables on energy use (for an excellent summary and overview, see [13]). Building factors were found to explain about 42% [13] and 54% [14] of the variability in energy consumption. Without providing a combined score for the total predictive power of building factors, Steemers and Young Yun [15] found that building factors were more important than occupant characteristics in explaining space heating demand. Generally, predictors that could not be easily changed through energy-efficiency interventions, such as floor area (e.g. [16–19,13]), dwelling type [13,19], climate [20,15] and weather [21] were most important in predicting energy demand. The role of dwelling age, has been shown to have a negative linear relationship with energy consumption in some studies (e.g. [13]) but not in others (e.g. [18]) which might reflect that building technologies or retrofitting programs occurred at different times in different countries. Presence of basement, shed, and garage were all associated with greater energy use [13].

Of those variables that could be targeted by energy-efficiency interventions insulation levels of walls, floors, and windows are associated with energy consumption [16,13,18,15]. Whilst these studies do not cite the joint amount of variability explained by the above-mentioned factors, their respective impact weight (i.e. beta coefficient in regression analysis) is generally lower than those of the more fixed factors of dwelling type and size [13].

1.2. Impact of occupants on energy use

When reviewing studies on the impact of any occupant characteristics, the composition of the study sample needs to be considered. If occupants live in very similar building types in the same location, i.e. there is hardly any variation in building factors, one would expect that the remaining variability is mainly due to non-building factors, e.g. occupant characteristics. Indeed, a number of studies have shown that in similar buildings, energy consumption can vary greatly due to occupant characteristics (e.g. [22,23]). However, in these studies building factors are already accounted for indirectly by the choice of very similar buildings; hence, the remaining variability is more likely to be due to non-building factors. Therefore, studies are reviewed below only if they have *not* restricted the sample to very similar buildings.

Guerra Santin et al. [13] found that when controlling for building characteristics, occupant characteristics explained an additional 4.2% of the variability in domestic energy consumption. For space heating, occupant characteristics account for 20% of the variability in energy use [15], a much higher value, but the authors do not address or discuss the issue of multicollinearity of predictors. Sonderegger [14] concluded that 18% of the variability in gas consumption was due to occupant behaviour; however, this was estimated from changes observed when houses changed occupants which could have also brought other significant changes.

1.2.1. Household characteristics

Two of the most well-documented household characteristics with known impacts on energy use are income and household size. Generally speaking energy consumption increases with higher income (e.g. [24–26]). Household size has been shown to be positively correlated with total energy usage (e.g. [25,26]). The role of householder age is less clear, with some studies finding a negative relationship between age and energy consumption (e.g. [27]), some found no significant relationship [24], while others found a positive relationship between age and energy consumption (e.g. [28,13]); these differing findings might be due to the fact that studies were not consistent in whose age they measured, e.g. any respondent, household reference person, or eldest occupant. Tenure is related to energy consumption. However, it is also likely to be confounded with building characteristics, e.g. socially rented dwellings tend to be better insulated and privately rented dwelling fare the worst [1]. Working from home has been shown to be significantly associated with gas and electricity usage [16].

1.2.2. Occupant heating behaviour

In the models of domestic energy consumption that are most commonly used in the UK (BREDEM models, for an overview see Kavgic et al. [29]), occupant influence can be modelled using the heating demand temperature and heating pattern, even though generally standard assumptions are used for those variables. A sensitivity analysis on the BREDEM-informed model Community Domestic Energy Model (CDEM) found that heating demand temperature was the variable to which the model was most sensitive, followed by heating pattern [30]. Assuming the model is a good representation of reality, then these two occupant behaviour related variables offer considerable scope for changing energy consumption and could form useful targets for behaviour change interventions. Hence, these two variables might be important predictors of domestic energy consumption and are of a type that could be changed through an intervention.

Empirical studies have confirmed the link between indoor temperature and space heating demand [31] and total energy consumption [32] and between heating set point temperature and space heating demand [15], and between day, night, and evening indoor temperature and energy consumption [13]. The proportion of heated rooms [15] and bed rooms [13] was also positively related to energy consumption. Using a thermostat as a temperature control was associated with higher energy use [13] as well as more frequent heating [33].

1.2.3. Psychological constructs

A range of psychological constructs have been postulated as explanatory variables for energy demand including values; attitudes; habit strength; and others. Values are generally defined as desirable trans-situational goals varying in importance, which serves as a guiding principle in the life of a person [34]. While the relationship between values and specific environmentally significant behaviours has been studied in detail (see [35] for an overview), very little research has been carried out on the relationship between values and measured domestic energy consumption. Vringer and Blok [36] found no relationship between domestic energy requirements and values including problem perception of climate change. Abrahamse and Steg [24] found that psychological variables such as attitudes and perceived behavioural control were not related to energy consumption but only to energy savings in an intervention program. Similarly Brandon and Lewis [25] found that environmental attitudes did not predict historic energy consumption but were related to energy savings in a subsequent intervention program.

Huebner et al. [37] showed that self-reported habit strength was significantly related to self-reported energy consuming behaviours and to actual energy consumption, when controlling for several building factors. However, the sample was restricted to social housing tenants only, and the overall impact of habits were relatively small.

Hence, previous research found no or little impact of psychological variables on domestic energy consumption.

1.3. Our study

The main aim of our study is to calculate and compare the explanatory power of different types of variables on domestic energy consumption. First, the explanatory power is calculated using linear models of (1) building variables only, (2) socio-demographic variables only, (3) heating behaviour variables only, and (4) a mixed category of other "occupant" variables, i.e. self-reported behaviours, and attitudes. Models are tested and corrected for multicollinearity in order to arrive at coefficients for individual predictors that can be interpreted. Then, the latter three models (socio-demographic, heating behaviour, other occupant variables) are subsequently added to the building variable model, to calculate the increase in explanatory power, again testing and correcting for collinearity.

2. Methods

2.1. Data set

The data analysed for this paper formed part of the Energy Follow-Up Survey (EFUS), commissioned by the Department of Energy and Climate Change [38]. An interview survey asked householders about details of their dwelling and their heating practices. Gas and electricity meter readings were obtained in a subsample of homes, and were used to estimate annualised energy consumption. All households in the survey had also participated in the English Housing Survey (EHS) which collects detailed information about the English building stock. The sample size for EFUS was $N = 2616$; meter readings were available for $N = 1345$ households. Of those 1345 households, another 412 were excluded from the sample based on the following five criteria:

(1) there was a positive reply to the question if physical changes to the dwelling had been carried out since the last EHS; as it was not recorded what exactly changed and when the impact on energy consumption could not be assessed,
(2) there was a positive reply to the question if the household composition had been changed since the last EHS; again, as it was not recorded how and when the household changed, the impact on energy consumption could not be assessed,
(3) the annual energy consumption was considered an extreme value (±3 SD from the sample mean of energy consumption),
(4) usage of a heating fuel that was not gas or electricity (to avoid too small subsamples), and
(5) missing data on the attitudinal variables which would have made it necessary to code the variable categorically instead of using them as a continuous predictor and creating a little informative category of "missing data". Hence the total remaining sample size was $N = 924$ households which formed the basis for all the analyses carried out in this paper.

Predictors were broadly categorised into building characteristics, i.e. factors that are pertinent to the building, socio-demographics, and a wider range of occupated-related variables, such as attitudes towards climate change, energy-saving actions, self-reported heating practices. The variables were chosen based on previous research (see Section 1), and limited by what was available in the data set.

2.1.1. Building variables

Table 1 summarises the building variables and their frequencies used in subsequent analysis.

Note that boiler type is not included as it is highly correlated with fuel type: All properties with electricity as their main heating fuel do not have boiler. Similarly, the type of heating system is not included as it is largely identical to fuel type: Those using electric as their main fuel had storage heaters and those using gas had a central heating system, with only nine households using a fixed room heating (3 of an electric type). Finally, loft insulation is not included as a predictor because the category "not applicable" meant that the dwelling in question was a flat; rendering this category uninformative and identical with the dwelling category.

2.1.2. Socio-demographic variables

Table 2 summarises the variables used for the socio-demographic set of predictors. Income was coded as equivalized income, meaning that household incomes were adjusted for household composition and size such that those incomes can reasonably be directly compared with each other. This implies increasing the

Table 1
Overview of building variables and their frequencies.

Variable (abbreviation)	Categories (*N*)
Floor area (FloorArea)	n/a (continuous: $M = 92.07$ m^2, SD = 43.11)
Dwelling type (DwType)	Converted & purpose built flat (133), detached (225), end terrace (113), mid-terrace (170), semi-detached (283)
Number of storeys (NoStorey)	n/a (continuous: $M = 2.10$; SD = 0.76)
Government Office Region (GOR)	East (96), East Midlands (63), London (102), North East (67), North-West (162), South East (127), South-West (91), West Midlands (89), Yorkshire and the Humber (127)
Dwelling age (DwAge)	Pre 1919 (131), 1919–44 (162), 1945–64 (212), 1965–80 (216), 1981–90 (69), post 1990 (134),
Wall type (WallType)	9-in. solid wall (131), cavity uninsulated (278), cavity with insulation (451), other (64)
Double glazing (DblgGlaz)	Entire house (736), more than half (105), less than half (36), no double glazing (47)
Attic (Attic)	Yes (98), no (826)
Conservatory (Conservatory)	Yes (185), no (739)
Main heating fuel (Fuel)	Electrical system (35), gas system (889)
SAP rating (SAP)	B & C (120), D (526), E (233), F & G (45)

Table 2
Socio-demographic variables and their frequencies.

Variable (abbreviation)	Categories (N)
Number of occupants (HHSize)	n/a (continuous: $M = 2.41$, SD = 1.27)
Age of youngest dependent children (DepChild)	No dependent children (621), 0–4 years (127), 5–10 years (85), 11–15 (61), older than 16 (30)
AHC (After-Housing-Costs) equivalised income quintiles (Income)	1st quintile – lowest (135), 2nd quintile (201), 3rd quintile (193), 4th quintile (198), 5th quintile – highest (197)
Tenure (Tenure)	Local authority (108), owner occupied (596), private rented (90), Registered Social Landlord RSL (130)
Sex of HRP (SexHRP)	Female (375), male (549)
Age of HRP (AgeHRP)	16–29 yrs (47), 30–44 (228), 45–64 (385), 65 or over (264)
Employment status of household (EmployHH)	1 or more work full time (463), 1 or more work part time (81), none working and none retired (88), none working, one or more retired (292)
Someone in household sick or disabled? (sick/disabled)[a]	No (610), yes (314)
Someone in household over 75 years?	No (816), yes (108)
Length residency (LengthRes)	2 yrs or less (156), 3–4 yrs (109), 5–9 years (187), 10–19 (205), 20–29 (126), 30+ years (141)

[a] 7 households had not answered the question of whether someone sick or disabled was in the household; these were combined with the "no" answers.

incomes of small households and decreasing the incomes of large households and the extent of these increases and decreases is determined by an internationally agreed set of scales. Equivalized income was chosen as it is considered to provide a better indication of household disposable income which should in turn be a predictor of expenditure on energy consuming appliances as well as financial pressure on energy bills. Age of the household reference person (HRP) was coded as a categorical variable, with another dichotomous variable indicating if anyone over 75 years was present in the household.

2.1.3. Heating behaviour variables

Participants had been asked about their heating behaviour with pre-defined answer categories. Table 3 summarises the variables used. Central heating is abbreviated with 'CH'.

2.1.4. Other 'occupant' variables

Whilst EFUS asked a variety of questions on pro-environmental behaviour, energy use, and climate change, only a subset were used in the following analysis. The questions were selected on several grounds.

- Availability of data: If in more than 10% of responses, the chosen option was "not applicable", the item was excluded. This was the case for items such as composting or setting the dishwasher in certain ways (which is only a possible option to those who have the opportunity to do so), and also when asking about the likelihood of investing in new loft insulation (not applicable to those in rented accommodation and those living in flats).
- Relevance to energy consumption: For behavioural items, the only items selected were those related to domestic energy consumption, such as turning off lights, but not general ones such as recycling.
- Demonstrated impact in previous research: Items were included that had been shown to be of importance previously, e.g. habit [37,39] and perceived behavioural control [24].

Table 4 shows mean answers and standard deviation for the variables included in the analysis.

Note that individual items are used as predictors instead of combining them into scales (e.g. construction of a "pro-environmental behaviour" scale). This was done as factor analysis and reliability analysis did not provide evidence for scales underlying the items.

The correlations between items were generally low, e.g. the mean correlation between the four items asking about energy-saving actions in the household was $r = .11$, ranging from $r = .002$ to $r = .222$. The item 'LightsOff' was reverse-coded for the correlation analysis so that positive correlations would be expected between all items.

One item was used as a categorical predictor, asking participants to indicate "Which of these statements best reflects how you currently feel?". The response options and number of participants who chose each option are summarised below; bold shows the abbreviation used later in the paper.

- Climate change is caused by energy use and I'm beginning to think that I **should do something** ($N = 102$).

Table 3
Variables measuring heating behaviour.

Variable (abbreviation)	Categories (*N*)
Operation Heating System (Timer)[a]	CH/timer not used/not present (163), thermostat (68), manual switch (130), timer (563)
Proportion of rooms heated by supplementary heating (SupplHeating)	None (461), up to 20% (417), 20–50% (46)
Proportion of rooms not heated (PropNotHeated)	None (346), up to 10% (110), 10–20% (241), 20–50% (193), Over 50% (34)
Length heating season (HeatingSeason)	Not applicable (45), 1–3 months (60), 4 months (128), 5 months (228), 6 months (255), 7 months (129), 8 months (48), 9–12 months (31)
Heating duration hrs/day (HeatingDuration)	na (247), <4 hrs (458), 4–10 hrs (91), 11–16 hrs (89), >17 hrs (39)

[a] Notes. The category CH or timer not used/not present is relatively broad as it encompasses those not having a timer, not using a timer or responding not applicable.

Table 4
Overview of the variables measuring other occupant variables.

	Variable (abbreviation)	*M* (SD)
Answer scale	*Do you agree that...*	
1 = Agree strongly	The Government is taking sufficient action to tackle climate change? (*Government*)	3.19 (1.03)
2 = Tend to agree	It would embarrass me if my friends thought my lifestyle was purposefully environmentally friendly? (*Embarrass*)	3.06 (1.07)
3 = Neither agree nor disagree	Being green is an alternative lifestyle, it's not for the majority? (*BeingGreen*)	3.05 (1.22)
4 = Tend to disagree	I find it hard to change my habits to be more environmentally-friendly? (*Habit*)	3.32 (1.20)
5 = Disagree strongly	It's not worth me doing things to help the environment if others don't do the same? (*NotWorth*)	3.64 (1.27)
Answer scale	How often, if at all, do you personally...	
1 = Always	Switch off lights when you are not in the room? (*LightsOff*)	1.64 (0.98)
2 = very often	Boil the kettle with more water than you are going to use? (*BoilKettle*)	3.73 (1.31)
3 = Quite often	Leave your TV or PC on standby for long periods of time? (*TVStandby*)	3.57 (1.62)
4 = occasionally	Wash clothes at 30 degrees or lower? (Wash30)	3.35 (1.59)
5 = never		

- Climate change is caused by energy use and I'm **doing a few small things** to help reduce my energy use and emissions ($N = 397$).
- Climate change is caused by energy use and I'm **doing quite a number** of things to help reduce my energy use and emissions ($N = 216$).
- Climate change is caused by energy use and I'm **doing lots of things** to help reduce my energy use and emissions ($N = 44$).
- I **don't believe** there are climate change problems caused by energy use and I'm **not willing or able to change my behaviour** ($N = 52$).
- Whether there are climate change issues or not, I **am not willing or able to change** my behaviour with regards to energy use ($N = 65$).
- Don't know ($N = 48$). – **don't know**.

2.2. Dependent variable: annualized combined energy consumption

The dependent variable used was the annualized energy consumption in kW h. This value reflected either the sum of both gas and electricity data, or just electricity consumption for households that were not connected to the gas grid. The dependent variable was log-transformed (natural log) to achieve greater symmetry of the distribution, in particular of the residuals in the regression analysis. Values that were three standard deviations above or below the mean value were excluded from analysis, i.e. 9 cases. The mean log-transformed energy consumption was $M = 9.78$ kW h with a standard deviation of $SD = 0.56$; i.e. the geometric mean of the non-transformed energy consumption was $M = 17635.98$ kW h, and the arithmetic mean was $M = 20427.07$ kW h.[1]

[1] In regression models where the dependent variable has been log-transformed and the predictors have not, the format for interpretation is that dependent variable changes by 100 ∗ (coefficient) percent on average for a one unit increase in the independent variable while all other variables in the model are held constant (http://www.ats.ucla.edu/stat/sas/faq/sas_interpret_log.htm).

2.3. Statistical analysis

In a first step, linear ordinary least squares (OLS) regression analysis was performed separately for the four classes of variables as presented above, i.e. 'building factors', 'socio-demographic', 'heating behaviour', and 'other occupant variables'. Given the suspected issue of collinearity, the variance-inflation factors (VIF) were then inspected. VIF are a measure of how much the variance of an estimated regression coefficient increases if the explanatory variables are correlated. If uncorrelated, VIF = 1. There is no formal cut-off point for critical values of VIF; in this paper a value of 3.3 was used [40]; this is a compromise value which is slightly higher than a conservative value of 2.5 (e.g. [41]) but below other suggestions of above 5 or even 10 [42].

If VIFs greater 3.3 were found in the OLS regression, then the Lasso regression (least absolute shrinkage and selection operator) was employed. This is based on the linear model but estimates the regression coefficients differently (see [42]; for an excellent description of this procedure, which was originally developed by Tibshirani [4]). Lasso is a penalised regression analysis which encourages model sparsity. It effectively performs variable selection by using a fitting procedure which aims to set some coefficients to zero, making the model sparse. It seeks to minimise the usual sum of squared errors, but constrained with a bound on the sum of the absolute values of the coefficients.

In order to choose the optimal tuning penalty parameter λ, k-fold cross-validation was used, with 100 values for λ, and the data were randomly split into $k = 10$ groups. For each λ, the cross-validation error was calculated, and then optimal value of λ was chosen which corresponds to the minimum cross-validation error (for details, see [42]). The "one-standard error" rule was applied; choosing as the final optimal value of λ that which gives the most regularized model (most sparse model) such that its error is within one standard error of the minimum error as estimated in cross-validation. After choosing the final value of λ, the model was re-run an all data.

In the data analysed here, most variables were categorical variables which were dummy-coded prior to analysis. Group-lasso was used which discards a categorical variable in total instead of individuals categories within that variable to ease interpretation ([43]; R package SGL).

After identifying which coefficients were set to zero using Lasso, then OLS was repeated omitting those variables.

After building all individual models, models were then successively combined until resulting in a final model encompassing all predictors, tested and adjusted for multicollinearity.

3. Results

3.1. Individual models

3.1.1. Building model

All building factors together explained 39.39% of the variability in domestic energy consumption (adjusted R^2 = 37.29%), $F(31,892)$ = 18.70, $p < .001$. Two VIF were greater than 3.3, making it necessary to run a Lasso regression.

In the Lasso regression, two variables were set to zero: number of storeys and wall type. The OLS building model was then re-run omitting these two variables. The model explained 39.07% of the variability (adjusted R^2 = 37.24%), $F(27,896)$ = 21.28. For details of coefficients, see Appendix A, Table A.1.

3.1.2. Socio-demographic model

The socio-demographic model explained R^2 = 24.77% (adjusted R^2 = 22.59%) of the variability in domestic energy consumption, $F(26,897)$ = 11.36, $p < .001$. However, three variables showed VIF values above the chosen criterion. Hence, Lasso regression was performed on the data. Four variables were set to zero: Gender HRP, Employment Status, Presence of sick/disabled person, Presence of person over 75 years.

The OLS was then rerun excluding those three variables. The resulting model explained 24.48% of the variability in energy consumption, with an adjusted R^2 of 22.81%, $F(20,903)$ = 14.64, $p < .001$. Details can be found in Appendix A, Table A.2.

3.1.3. Heating behaviour model

The overall model of self-reported heating behaviour explained 14.14% of the variability in annual energy consumption (adj. R^2: 12.2%), $F(20,903)$ = 7.44, $p < .001$. All VIF were lower than 3.13, indicating no critical multicollinearity. For details of coefficients, see Appendix A, Table A.3.

3.1.4. Other occupant variables model

The final individual model consisted of testing the explanatory power of other "occupant" variables. The overall model was significant, $F(15,908)$ = 3.02, $p < .001$, with R^2 = 4.75% and adjusted R^2 = 3.18%. Collinearity was not an issue with all VIFs smaller than 1.4. For details of coefficients, see Appendix A, Table A.4.

3.2. Combined models

In the next step, we combined the models successively, and tested if each additional model increased explanatory power:

(1) *Building + socio-demographic:* The model combining building factors and sociodemographic factors (*build_and_socio*) explained 43.11% of the variability, with the adjusted R^2 being 40.20%, $F(45,878)$ = 14.79, $p < .001$. An ANOVA showed that the difference in explanatory power between the model containing building variables only and that containing socio-demographics variables in addition was significant ($p < .001$) implying that inclusion of socio-demographic variables increases the explanatory power of the model.

(2) *Building + socio-demographic + heating behaviour:* Occupant heating behaviour variables were added to the *build_and_socio* model. This new model (*build_socio_heating*) explained 45.06% of the variability in energy consumption, with an adjusted R^2 of 40.90% [$F(65,858)$ = 10.83, $p < .001$]. This increase was marginally significant (p = .066).

(3) *Building + socio-demographic + heating behaviour + otheroccupant:* In a final step, the other 'occupant' variables were added to the previous model. The '*build_socio_heating_other occupant*' model explained 47.05% of the variability (adjusted R^2 = 42.03%), $F(80,843)$ = 9.36, $p < .001$. This increase was significant (p = .007).

Fig. 1a shows the adjusted R^2 for the four individual models and Fig. 1b shows the adjusted R^2 for the combined models.

So far, the analysis showed that building variables explain the largest variation in domestic energy consumption but that adding further variables, in particular socio-demographic variables increases explanatory power further. However, the combined *build_socio_heating_otheroccupant* model showed five critical VIF values; hence, Lasso regression was performed to remove variables with high multicollinearity and then arrive at coefficients that can be reliably interpreted. Table 5 shows which variables were set to zero in the different predictor classes in the Lasso regression.

Excluding these variables, OLS was performed. All VIF were smaller than 2. The total model explained R^2 = 45.57% of the

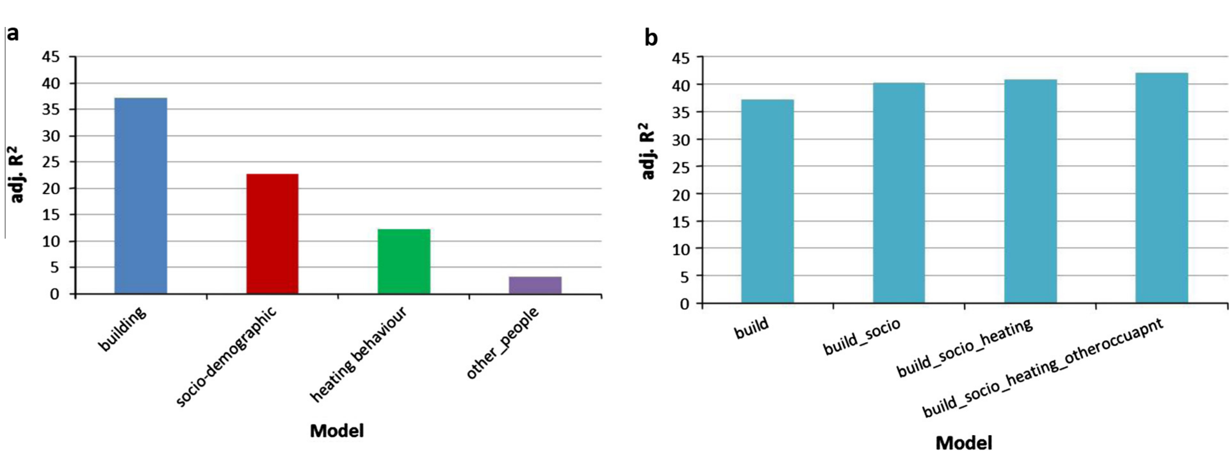

Fig. 1. Adjusted R^2 (%) for the four models (a) and for the combined models (b).

Table 5
Variables set to zero in the Lasso regression combining all models.

Predictor class			
Building	Socio-demographics	Heating behaviour	Other occupant
None	• Presence of dependent children • Income • Tenure • Age HRP • Length residency	• Proportion of rooms with supplementary heating • Proportion of rooms not heated • Heating hours/day	• Government • Embarrass • Being Green • Lights off • Boil kettle • Wash 30 degrees

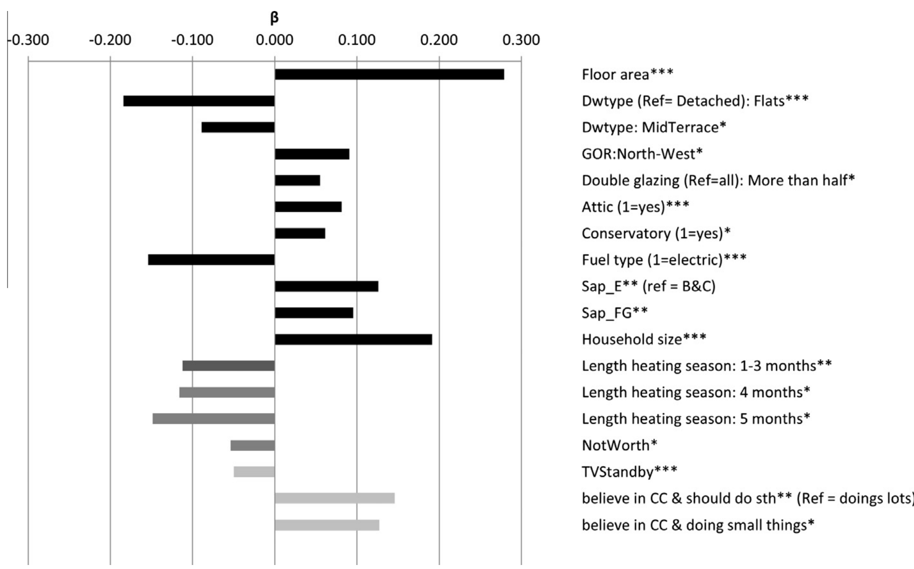

Fig. 2. Standardised regression coefficients for a full model without collinearity, showing only significant predictors.

variability in energy consumption (adjusted $R^2 = 42.65\%$), $F(47,876) = 15.61$, $p < .001$. Table A.5 in the Appendix A shows regression coefficients of the Lasso regression and the subsequent OLS regression.

Fig. 2 shows the standardised regression coefficients β for significant effects only from the OLS (corrected for collinearity).

In the final model, building variables clearly dominate in explaining energy consumption. Both for socio-demographics and heating behaviour, only one variable each remained a significant predictor.

4. Discussion

The detailed analysis clearly indicated that building variables explain by far most of the variability in domestic energy consumption. Whilst all four individual models were significant and explained a significant amount of variability in energy consumption, the impact of socio-demographic, heating behaviour and other occupant variables reduced greatly once controlling for building variables. In terms of the importance of individual predictors, findings were broadly in line with previous research, showing the importance of building size and dwelling type ([17,18,13], climate [15]) which was here indicated by the geographic location of the dwelling, double-glazing, the presence of an attic, and the presence of a conservatory [44]. The finding that the SAP rating is a significant predictor – with the highest ranking, here encompassing B- and C-rated dwellings showing significantly lower energy consumption – is an important finding given the debate about the usefulness of energy performance certification [45]. Household size remained the only significant predictor of energy consumption among the socio-demographic variables once controlling for all other predictors; in line with previous findings [25,26]. Analysis showed that age of HRP did not impact on energy consumption supporting the finding of Abrahamse and Steg [24]. Employment status did also not matter. Heating hours per day and the proportion of not heated rooms were non-significant predictors which might be surprising; on the other hand, these data were self-reported and respondents might differ in their accuracy. Also, which rooms were not heated, e.g. in terms of size, will presumably be important in determining the energy impact; however, this was not possible with the given data. The self-reported variable of heating periods in months per year remained significant in the final model, with those heating 5 months or less, using significantly less energy than those heating year-round. As discussed in the introduction, in building stock models, the heating duration per day and the heating demand temperature were those predictors that the CDEM model was most sensitive to [30]. The standard assessment procedure (SAP) assumes that in three summer months there is no heating at all, but for other months calculates internal and solar gains to see if heating is needed to reach the assumed demand temperature for the assumed hours [46]. BREDEM 12 calculates monthly energy demand; where the fraction of the month heated is a calculated input; hence, in neither model is the length of the heating season determined by the occupant beyond their impact on internal gains [47]. Given that broad climatic differences have been accounted for by modelling the geographical location, the observed variation in heating season is unlikely to arise just from climatic differences, reflecting instead variations in occupant seasonal heating behaviours. It is thus a potential new behavioural parameter to consider in modelling and policy.

Finally, the analysis performed here showed an impact of opinions about climate on energy consumption; with those believing in climate change and doing "lots of actions" to mitigate it, using less energy than those doing "small things" or thinking they should "do something", However, those not believing in climate change were not using more energy; hence, the overall effect is small and somewhat ambiguous.

Table A.1
Results of the Lasso regression and subsequent OLS omitting the predictors set to zero in the Lasso regression. Here and in all subsequent tables, *B* means unstandardised coefficient, β standardised coefficient, and bold font indicates predictors set to zero in the Lasso regression. Buildings model.

Predictor	β_L	B_{OLS}	SE_{OLS}	β_{OLS}
Floor area	6.058	0.004***	0.000	0.338
Dwtype (Ref = Detached): Flats	−2.343	−0.342***	0.070	−0.216
Dwtype: EndTerrace	0.445	0.003	0.059	0.002
Dwtype: MidTerrace	−0.340	−0.114	0.060	−0.079
Dwtype: Semi	0.496	−0.058	0.048	−0.048
Number Storeys	**0.000**	n/a	n/a	n/a
GOR (Ref = East): Midlands	0.064	0.079	0.072	0.036
GOR: London	0.160	0.152*	0.066	0.086
GOR: North East	0.123	0.129	0.071	0.060
GOR: North-West	0.219	0.136*	0.057	0.093
GOR: South East	−0.056	0.033	0.060	0.020
GOR: South West	−0.374	−0.119	0.065	−0.064
GOR: WestMidlands	−0.078	0.018	0.065	0.010
GOR: Yorkshire & Humber	0.057	0.067	0.060	0.042
Dwage (Ref = pre1919): 1919–44	0.186	0.041	0.057	0.028
Dwage: 1945–64	0.206	0.089	0.058	0.068
Dwage: 1965–80	−0.157	0.000	0.059	0.000
Dwage: 1981–90	−0.121	−0.026	0.077	−0.012
Dwage: post1990	−0.259	−0.024	0.071	−0.015
Wall (Ref = Cav. ins): Solid	**0.000**	n/a	n/a	n/a
Wall: Cavity uninsulated	**0.000**	n/a	n/a	n/a
Wall: Other	**0.000**	n/a	n/a	n/a
Double glazing (Ref = all): More than half	0.208	0.088	0.048	0.050
Double glazing: Less than half	0.079	0.049	0.081	0.017
Double glazing: None	0.065	0.055	0.072	0.022
Attic (1 = yes)	0.849	0.172***	0.051	0.096
Conservatory (1 = yes)	0.672	0.090*	0.039	0.065
Fuel type (1 = electric)	−2.355	−0.546***	0.082	−0.188
EPC: D	−0.045	0.077	0.052	0.069
EPC: E	0.425	0.163**	0.063	0.128
EPC: F & G	0.316	0.254**	0.095	0.098
Intercept	n/a	9.255***	0.121	n/a

Table A.2
Results of the Lasso regression and subsequent OLS omitting the predictors set to zero in the Lasso regression. Socio-demographics model.

Predictor	β_L	B_{OLS}	SE_{OLS}	β_{OLS}
Household size***	4.633	0.156***	0.020	0.357
DepChild(Ref = none): 0–4 years	−0.104	0.044	0.073	0.027
DepChild: 5–10 years	0.073	0.066	0.074	0.034
DepChild: 11–15 years*	0.278	0.193	0.075	0.086
DepChild: >16 years*	0.162	0.196*	0.098	0.062
Income	−0.574	0.030	0.055	0.022
Income2	−0.047	0.086	0.057	0.063
Income3**	0.509	0.153**	0.058	0.113
Income4***	1.224	0.260**	0.061	0.191
Tenure (Ref = Owner occ) Local authority***	−1.465	−0.188	0.055	−0.109
Tenure: private landlord	−1.039	−0.043	0.065	−0.023
Tenure: RSL***	−1.964	−0.245	0.052	−0.153
Gender HRP (1 = female)	**0.000**	n/a	n/a	n/a
AgeHRP (Ref: >65 yrs): 16–29 yrs	−0.325	−0.169	0.090	−0.067
AgeHRP: 30–44 yrs	−0.107	−0.091	0.060	−0.071
AgeHRP: 45–64 yrs	0.122	−0.041	0.043	−0.037
Employment (Ref = min 1 full time): at least 1 part time	**0.000**	n/a	n/a	n/a
Employment: none working, none retired	**0.000**	n/a	n/a	n/a
Employment: none working, at least 1 retired	**0.000**	n/a	n/a	n/a
Sick or disabled person (1 = yes)	**0.000**	n/a	n/a	n/a
Person over 75 yrs (1 = yes)	**0.000**	n/a	n/a	n/a
Length residency (Ref ⩽ 2 yrs): 3–4 yrs	−0.202	0.048	0.064	0.028
Length residency: 5–9 yrs	−0.107	0.104	0.058	0.075
Length residency: 10–19 yrs	0.218**	0.168	0.059	0.126
Length residency: 20–29 yrs	0.214**	0.197	0.068	0.122
Length residency: 30 + yrs	0.473***	0.293	0.067	0.190
Intercept	n/a	9.230***	0.085	n/a

The analysis was unable to detect the impact of novel technologies that might have significant implications for energy consumption, such as intelligent façade glazing [48] or technologies for nearly zero energy buildings [49,50]. Those factors were not assessed as they are of extremely low prevalence in the UK; however, their importance will likely increase in the next decades.

4.1. Conclusions

One central conclusion is that we are limited in how much of the variability in domestic energy consumption we can explain. Even using all variables measuring a variety of predictor types, we can only explain just under half of the variability in domestic energy consumption. Hence, there is currently a big gap in our understanding of energy consumption, indicating the need for more and better data.

The other central conclusion is the dominance of building variables in explaining domestic energy consumption over socio-demographic, self-reported heating behaviour, and attitudes and values. This conclusion holds true both when looking at the overall explanatory power of models with predictors from different classes of variables, and when looking at the incremental explanatory power when adding more variables to building models. However, the effect of behaviour might have been underestimated for three reasons. Firstly, some potentially crucial variables have not been measured, such as the heating demand temperature even though it was the most important input variable in a sensitivity analysis on a BREDEM-informed model [30]. EFUS measured temperatures in homes but only for a subset of homes, and temperatures themselves do not reveal the heating demand temperature. Whilst demand temperatures can be estimated from temperature measurements (see [39]), such estimates are likely to contain some degree of inaccuracy, e.g. as sensor placements was not entirely standardised and hence measures obtained at different room heights introducing error [51]. Another occupant-related variable that was shown to be related to residential energy consumption, working from home [16] was not measured in this survey.

Secondly, one-off behaviours such as installing loft insulation or double-glazing are not counted as behaviours but as building characteristics even though they originated from the decision, or at a minimum, the consent to install them. One-off behaviours generally have a much larger impact than curtailment behaviours such as turning lights off [52] and should be measured to adequately show the impact of behaviour on energy consumption, for example by using longitudinal surveys to track decision-making on retrofitting over time. However, the current data do not allow modelling them as behaviour, and it also is sensible to consider them as building characteristics, given that they will keep their impact even once occupants move home. Thirdly, as stated in the Methods (Section 2.1.4), factor and reliability analysis showed no underlying scale for presumably related items assessing attitudes and self-reported behaviours. This might reflect inadequate measurement of presumed constructs. Also, the amount of missing data made it necessary to discard a range of variables, such as those asking about composting or flying. Knowing which pro-environmental behaviours are not possible for a significant part of the population, such as composting, is interesting by itself as it might indicate the need for provision of other services. However, these items cannot be meaningfully used in analysis because of the large amount of missing data. Future research might assign a larger role to behaviour, such as through modelling one-off behaviours as a behavioural variable and choosing additional routine behaviours as further variables.

It is worth pointing out that the analysis focused solely on energy consumption as an outcome variable. In case of a different outcome variable, such as home satisfaction, different predictors might be significant and of differential importance. For example, double-glazing only had a small effect on energy consumption. However, it might have a large effect of satisfaction with the home and comfort in the home; as it might make the home feel significantly warmer and occupants consider warmth as the most important aspect of comfort in the home [37].

4.2. Implications

The findings have several implications for energy policy. They indicate that retrofitting homes through, for example, changing glazing, is indeed an important step towards reducing domestic energy demand, and is in line with policy aims [3] and previous research [53]. Also, given the increase in actual energy consumption associated with the presence of a conservatory, lending evidence to the suggestion that conservatories are associated with greater energy use [54] might call for stricter regulation on conservatories. Further research would need to understand in much more detail what type of conservatory is associated with greater energy use, e.g. only heated ones, those of a particular size, etc. Our findings also indicate some scope for behaviour change programs:

Table A.3
Results of OLS regression. Heating behaviour model.

Predictor	B_{OLS}	SE_{OLS}	β_{OLS}
No timer/CH used/present (Ref = Timer)	−0.363***	0.066	−0.250
Thermostat	−0.140	0.079	−0.066
Manual switch	−0.197**	0.064	−0.123
Prop room w/ suppl heating (Ref = none): up to 20%	0.129***	0.036	0.115
Prop room w/ suppl heating: 20–50%	0.169*	0.081	0.066
Prop room not heated (Ref = none): up to 10%	0.249***	0.058	0.145
Prop room not heated: 10–20%	−0.065	0.044	−0.051
Prop room not heated: 20–50%	−0.042	0.048	−0.031
Prop room not heated: over 50%	−0.213*	0.098	−0.072
Length heating season (Ref = 9–12 months): na	−0.056	0.132	−0.022
Length heating season: 1–3 months*	−0.236*	0.116	−0.105
Length heating season: 4 months	−0.171	0.105	−0.106
Length heating season: 5 months	−0.163	0.100	−0.127
Length heating season: 6 months	−0.086	0.100	−0.069
Length heating season: 7 months	−0.046	0.105	−0.028
Length heating season: 8 months	0.023	0.120	0.009
Heating hrs/day (Ref > 16 h): na	−0.068	0.094	−0.054
Heating hrs/day: <4	−0.104	0.093	−0.093
Heating hrs/day: 4–10	−0.038	0.103	−0.021
Heating hrs/day: 11–16	−0.158	0.103	−0.084
Intercept	10.016***	0.129	n/a

Signif.codes: *** 0.001, ** 0.01, * 0.05

Table A.4
Results of OLS regression other 'people variables' model.

Predictor	B_{OLS}	SE_{OLS}	β_{OLS}
Government*	0.033	0.018	0.061
Embarrass	0.002	0.017	0.004
BeingGreen**	0.033	0.015	0.072
Habit	−0.007	0.017	−0.015
NotWorth*	−0.010	0.016	−0.024
LightsOff	−0.010	0.019	−0.017
BoilKettle	0.001	0.015	0.002
TVStandby***	−0.042	0.012	−0.123
Wash30	0.000	0.011	0.000
Believe in CC & should do sth*** (Ref = believe & do lots)	0.380	0.104	0.214
Believe in CC & doing small things**	0.284	0.090	0.253
Believe in CC & quite a number**	0.251	0.091	0.191
Dont know	0.218	0.116	0.087
Don't believe in CC & don't want to change	0.157	0.119	0.065
Don't know about CC & don't want to change*	0.271	0.111	0.125
Intercept	9.530	0.164	n/a

Table A.5
Results of the Lasso regression and subsequent OLS omitting the predictors set to zero in the Lasso regression. Combined model with all predictors.

Predictor	β_L	B_{OLS}	SE_{OLS}	β_{OLS}
Floor area***	5.497	0.004	0.000	0.279
Dwtype (Ref = Detached): Flats***	−2.218	−0.291	0.068	−0.184
Dwtype: EndTerrace	0.436	0.003	0.057	0.002
Dwtype: MidTerrace*	−0.445	−0.127	0.059	−0.089
Dwtype: Semi	0.373	−0.073	0.046	−0.060
GOR (Ref = East): Midlands	0.098	0.108	0.070	0.049
GOR: London	0.119	0.114	0.064	0.065
GOR: North East	0.146	0.136	0.070	0.064
GOR: North-West*	0.204	0.132	0.056	0.091
GOR: South East	−0.037	0.053	0.058	0.033
GOR: South West**	−0.400	−0.090	0.063	−0.048
GOR: WestMidlands	−0.062	0.057	0.063	0.030
GOR: Yorkshire & Humber	0.094	0.102	0.058	0.063
Dwage (Ref = pre1919): 1919–44	0.176	0.020	0.056	0.014
Dwage: 1945–64	0.254	0.070	0.055	0.053
Dwage: 1965–80	−0.134	−0.009	0.057	−0.007
Dwage: 1981–90	−0.141	−0.053	0.074	−0.025
Dwage: post1990	−0.341	−0.084	0.069	−0.053
Double glazing (Ref = all): More than half*	0.268	0.097	0.046	0.055
Double glazing: Less than half	0.116	0.109	0.079	0.038
Double glazing: None	0.100	0.082	0.069	0.032
Attic (1 = yes)***	0.753	0.147	0.050	0.081
Conservatory (1 = yes)*	0.619	0.085	0.038	0.062
Fuel type (1 = electric)***	−2.237	−0.447	0.089	−0.154
Sap_D	−0.055	0.083	0.050	0.074
Sap_E**	0.524	0.161	0.061	0.126
Sap_FG**	0.371	0.246	0.091	0.095
Household size***	2.478	0.083	0.012	0.191
Dep	0.000	n/a	n/a	n/a
Dep	0.000	n/a	n/a	n/a
Dep	0.000	n/a	n/a	n/a
Dep4	0.000	n/a	n/a	n/a
Income**	0.000	n/a	n/a	n/a
Income2	0.000	n/a	n/a	n/a
Income3	0.000	n/a	n/a	n/a
Income4	0.000	n/a	n/a	n/a
Tenure (Ref = Owner occ): Local authority***	0.000	n/a	n/a	n/a
Tenure: private landlord	0.000	n/a	n/a	n/a
Tenure: RSL***	0.000	n/a	n/a	n/a
AgeHRP	0.000	n/a	n/a	n/a
AgeHRP2	0.000	n/a	n/a	n/a
AgeHRP3	0.000	n/a	n/a	n/a
Length residency (Ref ⩽ 2 yrs): 3–4 yrs	0.000	n/a	n/a	n/a
Length residency: 5–9 yrs	0.000	n/a	n/a	n/a
Length residency: 10–19 yrs***	0.000	n/a	n/a	n/a
Length residency: 20–29 yrs***	0.000	n/a	n/a	n/a
Length residency: 30+ yrs***	0.000	n/a	n/a	n/a
Timer1	−0.171	−0.083	0.051	−0.057
Timer2	−0.050	−0.041	0.055	−0.019
Timer3	−0.100	−0.064	0.043	−0.040
Prop room w/ suppl heating (Ref = none): up to 20%	0.000	n/a	n/a	n/a
Prop room w/ suppl heating: 20–50%	0.000	n/a	n/a	n/a
Prop room not heated (Ref = none): up to 10%	0.000	n/a	n/a	n/a
Prop room not heated: 10–20%	0.000	n/a	n/a	n/a
Prop room not heated: 20–50%	0.000	n/a	n/a	n/a
Prop room not heated: over 50%	0.000	n/a	n/a	n/a
Length heating season (Ref = 9–12 months): na	−0.042	−0.138	0.106	−0.053
Length heating season: 1–3 months**	−0.098	−0.252	0.095	−0.112
Length heating season: 4 months*	−0.035	−0.186	0.086	−0.116
Length heating season: 5 months*	−0.060	−0.191	0.083	−0.148
Length heating season: 6 months	0.026	−0.150	0.083	−0.120
Length heating season: 7 months	0.065	−0.098	0.086	−0.061
Length heating season: 8 months	0.112	−0.009	0.099	−0.004
Heating hrs/day (Ref > 16 h): na	0.000	n/a	n/a	n/a
Heating hrs/day: <4	0.000	n/a	n/a	n/a
Heating hrs/day: 4–10	0.000	n/a	n/a	n/a
Heating hrs/day: 11–16	0.000	n/a	n/a	n/a
Government*	0.000	n/a	n/a	n/a
Embarrass	0.000	n/a	n/a	n/a
BeingGreen**	0.000	n/a	n/a	n/a
Habit	−0.156	−0.010	0.013	−0.022
NotWorth*	−0.260	−0.024	0.012	−0.054
LightsOff	0.000	n/a	n/a	n/a
BoilKettle	0.000	n/a	n/a	n/a

(*continued on next page*)

Table A.5 (*continued*)

Predictor	β_L	B_{OLS}	SE_{OLS}	β_{OLS}
TVStandby***	−0.420	−0.017	0.009	−0.050
Wash30	0.000	n/a	n/a	n/a
Believe in CC & should do sth** (Ref = doings lots)	0.070	0.258	0.080	0.146
Believe in CC & doing small things*	0.009	0.143	0.070	0.127
Believe in CC & quite a number	−0.018	0.135	0.071	0.103
Dont know	0.003	0.143	0.091	0.057
Don't believe in CC & don't want to change	−0.025	0.030	0.092	0.013
Don't know about CC & don't want to change	0.009	0.145	0.086	0.067
Intercept	n/a	9.342	0.167	n/a

Given that the length of the heating season (corrected for geographic location and building factors) has a significant impact on energy consumption, designing an intervention to shorten the heating season would be an option; however, it will need further research to understand what determines the length of the heating season in a household. Also, the climate-independent length of heating season could be included in models predicting energy consumption in buildings.

Finally, given the dominance of building size in explaining domestic energy consumption, another option might be to promote choosing a building size that corresponds to what is deemed necessary in a standard for the number of occupants and their needs. Around 8.0 million households of the roughly 23.4 million households in the UK were estimated to be under-occupying their accommodation in 2011–2012, i.e., they had at least two bedrooms more than they needed according to the bedroom standard [55]. A further 7.7 million households had one bedroom more than they needed under the bedroom standard. It is noteworthy that the bedroom standard sets out minimum criteria: A separate bedroom is allowed for each married or cohabiting couple, any other person aged 21 or over, each pair of adolescents aged 10–20 of the same sex and each pair of children under 10 [56]. However, a bedroom converted into other uses, i.e. a study, does not count as a spare bedroom. If only those with two or more spare bedrooms downsized their home, significant energy savings could be expected. The problem is that not enough accommodation with fewer bedrooms is available. Overcrowding only affects 1.1 million of households [56] so it is not a question of redistribution of existing housing. Promoting solutions such as turning houses into multiple flats might be an opportunity to ensure living in dwellings with a more appropriate, i.e. most often fewer, number of bedrooms which might be preferential to building new properties with fewer bedrooms as the latter would still leave a tremendous number of too large buildings. We acknowledge however, that implementing such a change through policy would present both political challenges and potential social resistance but recommend not to rule downsizing out as a way of reducing carbon emissions from homes.

A non-content related but equally important implication of the presented analysis is the need to ensure correlations between predictors are tested and accounted for by choosing appropriate analysis techniques.

Acknowledgement

Funding was provided by: Research Councils UK (RCUK) Centre for Energy Epidemiology (EP/K011839/1).

Appendix A

See Tables A.1–A.5.

References

[1] Palmer J, Cooper I. United Kingdom energy housing factfile 2013; 2014. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/345141/uk_housing_fact_file_2013.pdf> [accessed 01.04.15].
[2] DECC. The UK low carbon transition plan: national strategy for climate and energy. London: Department of Energy and Climate Change, The Stationary Office; 2009.
[3] UK CCC. Fourth carbon budget. London (UK): UK Committee on Climate Change; 2010.
[4] Tibshirani R. Regression shrinkage and selection via the lasso. J Royal Statist Soc B 1996;58(1):267–88.
[5] Bordass B, Cohen R, Standeven M, Leaman A. Assessing building performance in use 3: energy performance of probe buildings. Build Res Inform 2001;29(2):114–28.
[6] Forouzanfar M, Doustmohammadi A, Menhaj M, Hasanzadeh S. Modeling and estimation of the natural gas consumption for residential and commercial sectors in Iran. Appl Energy 2010;87:268–74.
[7] Menezes A, Cripps A, Bouchlaghem D, Buswell R. Predicted vs. actual energy performance of non-domestic buildings: using post-occupancy evaluation data to reduce the performance gap. Appl Energy 2012;97:355–64.
[8] Aydinalp M, Ugursal VI, Fung AS. Modeling of the space and domestic hot-water heating energy-consumption in the residential sector using neural networks. Appl Energy 2004;79(2):159–78.
[9] Kialashaki A, Reisel JR. Modeling of the energy demand of the residential sector in the United States using regression models and artificial neural networks. Appl Energy 2013;108:271e80.
[10] Koo C, Park S, Hong T, Park HS. An estimation model for the heating and cooling demand of a residential building with a different envelope design using the finite element method. Appl Energy 2014;115:205–15.
[11] Martinaitis V, Zavadskas EK, Motuzienė V, Vilutienė T. Importance of occupancy information when simulating energy demand of energy efficient house: a case study. Energy Build 2015;101:64–75.
[12] Jones RV, Lomas KJ. Determinants of high electrical energy demand in UK homes: socio-economic and dwelling characteristics. Energy Build 2015;101:24–34.
[13] Guerra Santin O, Itard L, Visscher H. The effect of occupancy and building characteristics on energy use for space and water heating in Dutch residential stock. Energy Build 2009;41:1223–32.
[14] Sonderegger RC. Movers and stayers: the resident's contribution to variation across houses in energy consumption for space heating. Energy Build 1978:313–24.
[15] Steemers K, Young Yun G. Household energy consumption: a study of the role of occupants. Build Res Inform 2009;37(5–6):625–37.
[16] Baker KJ, Rylatt RM. Improving the prediction of UK domestic energy-demand using annual consumption-data. Appl Energy 2008;85(6):475–82.
[17] Kelly S. Do homes that are more energy efficient consume less energy?: a structural equation model of the English residential sector. Energy 2011;36(9):5610–20.
[18] Theodoridou I, Papadopoulos AM, Hegger M. Statistical analysis of the Greek residential building stock. Energy Build 2011;43(9):2422–8.
[19] Yohanis YG, Mondol JD, Wright A, Norton B. Real-life energy use in the UK: how occupancy and dwelling characteristics affect domestic electricity use. Energy Build 2008;40(6):1053–9.
[20] French LJ, Camilleri MJ, Isaacs NP, Pollard AR. Temperatures and heating energy in New Zealand houses from a nationally representative study—HEEP. Energy Build 2007;39(7):770–82.
[21] Fikru MG, Gautier L. The impact of weather variation on energy consumption in residential houses. Appl Energy 2015;144:19–30.
[22] Gill ZM, Tierney MJ, Pegg IM, Allan N. Low-energy dwellings: the contribution of behaviours to actual performance. Build Res Inform 2010;5:491–508.
[23] Gram-Hanssen K. Residential heat comfort practices: understanding users. Build Res Inform 2010;38(2):175–86.
[24] Abrahamse W, Steg L. How do socio-demographic and psychological factors relate to households' direct and indirect energy use and savings? J Econ Psychol 2009;30:711–20.

[25] Brandon G, Lewis A. Reducing household energy consumption: a qualitative and quantitative field study. J Environ Psychol 1999;19:75–85.
[26] Druckman A, Jackson T. Household energy consumption in the UK: a highly geographically and socio-economically disaggregated model. Energy Policy 2008;36(8):3177–92.
[27] Gatersleben B, Steg L, Vlek C. Measurement and determinants of environmentally significant consumer behaviour. Environ Behav 2002;34 (3):335–62.
[28] Liao HC, Chang TF. Space-heating and water-heating energy demands of the aged in the US. Energy Econ 2002;24:267–84.
[29] Kavgic M, Mavrogianni A, Mumovic D, Summerfield A, Stevanovic Z, Djurovic-Petrovic M. A review of bottom-up building stock models for energy consumption in the residential sector. Build Environ 2010;45:1683–97.
[30] Firth SK, Lomas KJ, Wright AJ. Targeting household energy-efficiency measures using sensitivity analysis. Build Res Inform 2010;38:25–41.
[31] Haas R, Auer H, Biermayr P. The impact of consumer behavior on residential energy demand for space heating. Energy Build 1998;27(2):195–205.
[32] Palmborg C. Social habits and energy consumption in single-family homes. Energy 1986;11(7):643–50.
[33] Kelly Scott. Do homes that are more energy efficient consume less energy?: a structural equation model of the English residential sector. Energy 2011;36 (9):5610–20.
[34] Schwartz SH. Universals in the content and structure of values: theoretical advances and empirical tests in 20 countries. In: Zanna M, editor. Advances in experimental social psychology. Orlando (FL): Academic Press; 1992.
[35] Nordlund AM, Garvill J. Value structures behind pro-environmental behavior. Environ Behav 2002;34:740–56.
[36] Vringer K, Blok T. Household energy requirement and value patterns. Energy Policy 2007;35:553–66.
[37] Huebner GM, Cooper J, Jones K. Domestic energy consumption – what role do comfort, habit, and knowledge about the heating system play? Energy Build 2013;66:626–36.
[38] DECC. Energy follow-up survey 2011. Report 9: methodology; 2013. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/274780/11_Methodology.pdf> [accessed 08.03.15].
[39] Huebner GM, McMichael M, Shipworth D, Shipworth M, Durand-Daubin M, Summerfield A. Heating patterns in English homes: comparing results from a national survey against common model assumptions. Build Environ 2013;70:298–305.
[40] Roberts N, Thatcher J. Conceptualizing and testing formative constructs: tutorial and annotated example. Database Adv Inform Syst 2009;40(3):9–39.
[41] Chan SH, Chen JH, Li YH, Tsai LM. Gly1057Asp polymorphism of insulin receptor substrate-2 is associated with coronary artery disease in the Taiwanese population. J Biomed Sci 2012;19:100.
[42] James G, Witten D, Hastie T, Tibshirani R. An introduction to statistical learning with applications in R. 1st ed. New York, Heidelberg, Dordrecht, London: Springer; 2013.
[43] Simon N, Friedman J, Hastie T, Tibshirani R. A sparse-group lasso. J Comput Graph Stat 2012;22(2):231–45.
[44] Oreszczyn T. The energy duality of conservatory use. In: Proceedings of the 3rd European conference on architecture: solar energy in architecture and planning, Florence, Italy, 17–21 May; 1993.
[45] Majcen D, Itard L, Visscher H. Theoretical vs. actual energy consumption of labelled dwellings in the Netherlands: discrepancies and policy implications. Energy Policy 2013;54:125–36. http://dx.doi.org/10.1016/j.enpol.2012.11.008.
[46] BRE. Standard assessment procedure (SAP 2012); 2012. <http://www.bre.co.uk/sap2012/page.jsp?id=2759> [accessed 07.96.15].
[47] Anderson BR, Chapman PF, Cutland NG, Dickson CM, Henderson G, Henderson JH, et al. BREDEM-12-model description: 2001 update. Watford (UK): BRE; 2002.
[48] Liu M, Wittchen KM, Heiselberg PK. Control strategies for intelligent glazed façade and their influence on energy and comfort performance of office buildings in Denmark. Appl Energy 2015;145:43–51.
[49] Baglivo C, Congedo PM, D'Agostino D, Zacà I. Cost-optimal analysis and technical comparison between standard and high efficient mono-residential buildings in a warm climate. Energy 2015;83(1):560–75.
[50] Zacà I, D'Agostino D, Congedo PM, Baglivo C. Assessment of cost-optimality and technical solutions in high performance multi-residential buildings in the Mediterranean area. Energy Build 2015;102:250–65.
[51] Gauthier S, Shipworth D. Variability of thermal stratification in naturally ventilated residential buildings. In: Conference proceedings: 2014 building simulation and optimization conference 2014, London, UK; 2014.
[52] Stern PC, Gardner GT. Psychological research and energy policy. Am Psychol 1981;36(4):329–42.
[53] Hamilton IG, Steadman PJ, Bruhns H, Summerfield AJ, Lowe R. Energy efficiency in the British housing stock: energy demand and the homes energy efficiency database. Energy Policy 2013;60:462–80. http://dx.doi.org/10.1016/j.enpol.2013.04.004.
[54] Pathan A, Summerfield AJ, Oreszczyn T, Lowe RJ, Young A. Trends in conservatory use: a comparison between the 1991 conservatory association survey and the UCL 2004 survey. London: University College London (UCL) for the Energy Saving Trust; 2007.
[55] DCLG. English housing survey – headline report 2011/12; 2013. <https://www.gov.uk/government/uploads/system/uploads/attachment_data/file/211288/EHS_Headline_Report_011-2012.pdf> [accessed 12.03.15].
[56] ONS. Overcrowding and Under-occupation in England and Wales; 2014. <http://www.ons.gov.uk/ons/dcp171776_360082.pdf> [accessed 12.03.15]