Towards Benchmarking English Residential Gas Consumption

J. Morris, D. Allinson, J. Harrison, K.J. Lomas

Abstract—The UK Government has emphasized the role of Local Authorities as a key player in its flagship residential energy efficiency strategies, by identifying and targeting areas for energy efficiency improvements. Residential energy consumption in England is characterized by significant geographical variation in energy demand, which makes centralized targeting of areas for energy efficiency intervention difficult. This paper draws on research which aims to understand how demographic, social, economic, urban form and climatic factors influence the geographical variations in English residential gas consumption. The paper reports the findings of a multiple regression model that shows how 64% of the geographical variation in residential gas consumption is accounted for by variations in these factors. Results from this study, after further refinement and validation, can be used by Local Authorities to identify areas within their boundaries that have higher than expected gas consumption, these may be prime targets for energy efficiency initiatives.

Keywords—UK Housing, Heating Energy, Socio-Economics, Statistical Modelling

I. INTRODUCTION

The UK Government is committed to achieving an 80% reduction in CO₂ emissions relative to 1990 levels by 2050 [1]. The UK’s residential sector accounts for 25% of total UK CO₂ emissions and represents 30% of its total energy use [2],[3]. Reducing the level of residential energy consumption is vital for the UK to meet its long term CO₂ reduction targets [4]. Space heating accounts for approximately 60% of residential energy use [5].

Reducing residential gas consumption can be achieved, without requiring households to change their behaviours or reducing their internal air temperatures, by increasing the energy efficiency of the existing housing stock. The residential sector in the UK has one of the oldest and least efficient housing stocks in Western Europe. A40% reduction in residential CO₂ emissions could be met by improving the thermal properties of UK housing and the introduction of appliances with greater levels of energy efficiency [6],[7].

Successive UK Governments have underlined their commitments to reducing energy consumption through policies to increase the energy efficiency of the residential sector. The 2011 Energy Act sets out plans for new residential energy efficiency policies [8]. The flagship policy for residential energy efficiency of the Coalition Government of the UK is the Green Deal.

The Green Deal allows households to cover the upfront costs of energy efficiency installations to their house through a loan which is attached to the property and is paid back through the assumed reductions to the householder’s energy bills [8], [9]. Alongside the Green Deal, the Energy Company Obligation (ECO) obliges energy companies to provide finance to lower income households to cover the costs of energy efficiency measures [8],[10]. The ECO publications of 2011 outlined plans to research ways in ‘which Government might be able to help energy companies find households’ suitable for the ECO scheme [10].

The Department of Energy and Climate Change (DECC) have indicated that Local Authorities are to be the front line in efforts to reduce UK CO₂ emissions, believing that Local Authorities have an advantage over centralized government in being able to adapt policies to best suit the needs of their communities and residents [11]. The Green Deal consultation emphasizes Local Authorities role in promoting the scheme due to the trust Local Authorities have from their residents [11]. Local Authorities are expected to play the roles of provider, partner, or promoter of the Green Deal. Each of these roles would be helped considerably by being able to identify areas which would benefit from energy efficiency schemes [9],[10].

Since 2007, beginning with the year of 2005, DECC have been producing residential gas and electricity consumption statistics. The stated aim is to ‘allow LAs and other interested bodies to more easily target specific areas as part of the implementation and monitoring of local energy strategies’ [12]. From 2009 DECC began to publish data at lower super output area (LSOA) level, census areas of approximately 500 households, and a minimum of 1000 people. LSOAs are DECC’s chosen unit for spatial dissemination as they are not subject to political boundary changes [12],[13].

Simply using DECC’s LSOA data to identify areas for Green Deal and ECO incentives based on ‘above average consumption’ figures would mask the demographic, social, economic, urban form, and climatic variations that exist between Local Authority areas (and between the LSOAs within Local Authorities) that account for variations in gas consumption. What is needed are benchmarks that account for these variations, and so highlight areas in which the housing is inherently less efficient. These areas are most likely to benefit from energy efficiency interventions, either through changes to the building stock, or to behavioural change campaigns. Thus the benchmarks would highlight areas where energy use is higher than that in other areas purely because of the energy inefficiency of the stock.

The benchmarks proposed in this paper are for gas consumption in LSOAs. 87 % of UK homes are connected to the gas grid and have central heating systems to heat the rooms and for hot water and also in some homes for cooking [14].
A statistical model to account for the varying demographic, social, economic, technical and climatic factors is developed here. LSOAs in which more gas is used than indicated by the benchmark figure could be targets for energy efficiency measures, while those in which gas use is lower than the benchmarked figure might be more efficient. Local Authorities could then use this information to direct Green Deal, ECO funding and other resources to those areas which have gas consumption levels in excess of their benchmark figure.

The benchmarking tool developed predicts gas consumption for each LSOA in England, using an ordinary least squares (OLS) multiple linear regression model. OLS multiple regression is a widely used method in economics and social sciences and its familiarity and relative simplicity makes it appropriate for use with a non-statistical audience in Local Government [15]. Given the political structure of the UK, with devolved administrations present in Wales, Scotland and Northern Ireland, each with their own arrangements for Energy Efficiency policy [16], this study focuses on England. England is directly governed by the Central Westminster Government and consists of 32482 LSOAs.

### II. MATERIALS AND METHODS

#### A. Data Sources

The energy use pattern in the domestic sector varies depending on factors such as climate, household composition, family income, cultural background and human behavioural factor [17]. DECC have acknowledged the variations in residential energy consumption and have begun developing the National Energy Efficiency Data-Framework (NEED) [18]. Preliminary results from a study of 4 million households has identified the average number of rooms, household income, tenure type and type of dwelling as important factors influencing the level of gas consumption [18]. DECC intend to use these results to evaluate the impacts of energy efficiency measures in the residential sector.

Other studies of household energy consumption, including electricity consumption, have identified external air temperature and the level of urbanization as key factors in explaining residential energy use [19], along with household income, house size, and house type [20].

The level of urbanization was measured by including variables for dwelling density, population density and the area of the LSOA. It was assumed that every house in each LSOA will have an electricity meter and therefore the number of houses is given by the total number of electricity meters.

Engineering studies investigating the effect of external air temperature on heating demands in buildings have used heating degree days [5], [19], [21] which are a function of the length of time the external air temperature is below a specified base temperature and how far below the base temperature it is [21].

The requirement for data to be used in this study was that it was available at consistent resolution for the whole of England. The data used in the analysis are (see Table I) from a combination of government sources (DECC, the Office of National Statistics, and the 2001 National Census), and commercial organizations (Experian, and the Meteorological Office) and are published at LSOA level, with the exception of heating degree days, which were published at 5x5km grid squares and assigned to the relevant LSOAs using GIS spatial tools. The year of study is 2008, as this was the most recent year for which DECC’s sub-national residential energy consumption figures were available when the study began. All the data were available for 2008 with the exception of the heating degree days, tenure type, house type and the average number of rooms. In the absence of more-up-to-date statistics which provide a comprehensive coverage of LSOAs, the 2001 census was used. The MET Office discontinued regular publication of heating degree days at grid square level in 2006, with no indication if future years might subsequently become available. In the absence of 2008 data, the number of heating degree days for 2006 in each grid square has been used. This is appropriate given an assumption that, while the absolute values of heating degree days for 2008 will differ from 2006, the relative differences in heating degree days for LSOAs across England will be broadly consistent. This is an aspect of this work that will be the subject of future refinement.

The dependent variable in the study is average gas consumption, published by DECC as the per meter residential gas consumption. Of the 32482 LSOAs in England, 31956 have at least one property with a connection to the national gas grid. Those with no connection to the gas grid have been excluded from the study. To prevent possible identification of individual household’s gas consumption, DECC do not publish all the data if there are less than six meters in a LSOA and the data is merged with that from a neighbouring LSOA, the per meter gas consumption is the averaged across the ‘merged’ LSOAs. There are 584 merged LSOAs nationally which have 0.3% of the total residential gas consumption.

Descriptive statistics (see Table II) were calculated for the dependent variable and the independent variables to understand the underlying distributions of the data and make necessary judgments concerning data transformations of the dependent variable before proceeding with regression analysis.

#### B. Correlation Analysis

Correlation coefficients between the independent variables and the dependent variable were calculated using SPSS. Only those independent variables with a correlation coefficient with the dependent variable of $|r|>0.25$ were considered for entry.
into the regression model. The scatter plots of these correlations were examined to ensure that a linear relationship sufficiently described the association between the independent variables and dependent variable. Given the large data source \((n=31956)\), most effects are likely to be declared statistically significant at the 0.05 level without offering any practical or theoretical justification for inclusion in the regression model \([30]\). This correlation analysis is a necessary step for reducing the volume of data prior to model building.

Correlation coefficients were also calculated between the independent variables to minimize co-linearity in the regression model. Independent variables included in the model should not have a correlation coefficient of \(|r|>0.7\) with another independent variable in the model. The value of \(|r|>0.7\) is emphasized by \([29]\) and \([30]\) to ensure the independence assumption of multiple regression is not violated.

### C. Regression Analysis

The first stage of building the regression model involved entering all the independent variables with correlation values \(|r|>0.25\) against per meter gas consumption by the stepwise entry method in SPSS. This gave an assessment of each value’s relative contribution to the model. Tolerance tests were also conducted to establish the degree of independence between the independent variables, with 0.1 being stated by \([30]\) as the minimum tolerance level acceptable for a multiple regression model to be specified correctly.

- Where two independent variables had correlation coefficients with \(|r|>0.7\), the variable with the strongest change in \(r^2\) in the original model was retained, with the others being removed. This process was repeated until none of the independent variables in the model had correlation coefficients of \(|r|>0.7\) with other independent variables in the model. Tests of statistical significance were run on both the model and the coefficients of the model to ensure that the model was valid and significant at the 0.01 level. Plots of the residuals were examined to ensure the model did not violate the assumptions of multiple regression, that the residuals were normally distributed and not correlated with the predicted variables.

### III. RESULTS

#### A. Descriptive Statistics for the Data Sources

Table III outlines the descriptive statistics and units of measurement for the variables included in the study. As can be seen from Fig.1, the distribution of per meter gas consumption is slightly skewed with a longer tail above the mean. The shape of the distribution, and the skew (1.03) and kurtosis (2.54) values suggest this distribution is sufficiently close to a normal distribution that there is no need for a transformation to the dependent variable.

![Fig. 1 Distribution of LSOA per meter gas consumption (2008)](image)

<table>
<thead>
<tr>
<th>Variable</th>
<th>Mean</th>
<th>Std. Dev</th>
<th>Skew</th>
<th>Kurtosis</th>
</tr>
</thead>
<tbody>
<tr>
<td>Per Meter Gas Consumption (kWh per year)</td>
<td>17072</td>
<td>3842</td>
<td>1.03</td>
<td>2.54</td>
</tr>
<tr>
<td>Number of Households</td>
<td>695</td>
<td>173</td>
<td>4.28</td>
<td>63.37</td>
</tr>
<tr>
<td>Population of LSOA</td>
<td>1584</td>
<td>308</td>
<td>4.84</td>
<td>72.27</td>
</tr>
<tr>
<td>Median Household Income (£ per year)</td>
<td>28693</td>
<td>9734</td>
<td>0.94</td>
<td>1.17</td>
</tr>
<tr>
<td>Area of LSOA (km²)</td>
<td>3.51</td>
<td>11.35</td>
<td>9.33</td>
<td>185.11</td>
</tr>
<tr>
<td>Number of Heating Degree Days (2006)</td>
<td>1869</td>
<td>193</td>
<td>0.36</td>
<td>0.94</td>
</tr>
<tr>
<td>Proportion of Detached Houses (%)</td>
<td>23</td>
<td>22</td>
<td>1.04</td>
<td>0.05</td>
</tr>
<tr>
<td>Proportion of Semi-Detached Houses (%)</td>
<td>32</td>
<td>20</td>
<td>0.60</td>
<td>-0.21</td>
</tr>
<tr>
<td>Proportion of Terraced Houses (%)</td>
<td>26</td>
<td>20</td>
<td>0.93</td>
<td>0.16</td>
</tr>
</tbody>
</table>

![TABLE III DESCRIPTIVE STATISTICS OF VARIABLES IN THE MODEL](image)
B. Correlation Analysis

Correlation analysis was carried out with the independent variables against the dependent variable. The results for those independent variables where |r| > 0.25 with per meter gas consumption are shown in Table IV, giving nine variables to be carried into the model building process. Correlation coefficients were then calculated for all the combinations of these nine independent variables to check the correlations between the independent variables for co-linearity as a basis for decision making over which variables to exclude from the model building. There were six pairs of independent variables with a correlation co-efficient of |r| > 0.7 (Table V).

### TABLE V

**CORRELATION BETWEEN INDEPENDENT VARIABLES (|r| > 0.7)**

<table>
<thead>
<tr>
<th>Variable 1</th>
<th>Variable 2</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Population Density</td>
<td>Housing Density</td>
<td>0.961</td>
</tr>
<tr>
<td>Average Number of Rooms</td>
<td>Proportion of Detached Households</td>
<td>0.751</td>
</tr>
<tr>
<td>Average Number of Rooms</td>
<td>Proportion of Owner Occupied Households</td>
<td>0.744</td>
</tr>
<tr>
<td>Average Number of Rooms</td>
<td>Proportion of Flats</td>
<td>-0.715</td>
</tr>
<tr>
<td>Proportion of Owner Occupied Households</td>
<td>Proportion of Socially Renting Households</td>
<td>-0.899</td>
</tr>
</tbody>
</table>

### TABLE IV

**CORRELATION OF INDEPENDENT VARIABLES AGAINST PER METER GAS CONSUMPTION (|R| > 0.25)**

<table>
<thead>
<tr>
<th>Variable</th>
<th>r</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Rooms</td>
<td>0.724</td>
</tr>
<tr>
<td>Median Household Income</td>
<td>0.637</td>
</tr>
<tr>
<td>Proportion of Owner Occupied Households</td>
<td>0.555</td>
</tr>
<tr>
<td>Proportion of Detached Houses</td>
<td>0.546</td>
</tr>
<tr>
<td>Number of Heating Degree Days</td>
<td>0.253</td>
</tr>
<tr>
<td>Population Density</td>
<td>-0.344</td>
</tr>
<tr>
<td>Proportion of Terraced Houses</td>
<td>-0.349</td>
</tr>
<tr>
<td>Proportion of Flats</td>
<td>-0.364</td>
</tr>
<tr>
<td>Housing Density</td>
<td>-0.380</td>
</tr>
<tr>
<td>Proportion of Socially Renting Households</td>
<td>-0.547</td>
</tr>
</tbody>
</table>

### TABLE VI

**REGRESSION STATISTICS OF VARIABLES IN THE MODEL**

<table>
<thead>
<tr>
<th>Independent Variable</th>
<th>$\Delta r^2$</th>
<th>Cumulative $r^2$</th>
<th>Co-Efficient</th>
<th>Standard Error</th>
<th>p</th>
<th>Tolerance</th>
</tr>
</thead>
<tbody>
<tr>
<td>Average Number of Rooms</td>
<td>0.524</td>
<td>0.524</td>
<td>2544</td>
<td>28.517</td>
<td>&lt;0.001</td>
<td>0.358</td>
</tr>
<tr>
<td>Median Household Income</td>
<td>0.090</td>
<td>0.614</td>
<td>0.155</td>
<td>0.002</td>
<td>&lt;0.001</td>
<td>0.473</td>
</tr>
<tr>
<td>Number of Heating Degree Days</td>
<td>0.027</td>
<td>0.641</td>
<td>3.588</td>
<td>0.073</td>
<td>&lt;0.001</td>
<td>0.782</td>
</tr>
<tr>
<td>Proportion of Terraced Houses</td>
<td>0.002</td>
<td>0.643</td>
<td>-10.20</td>
<td>0.689</td>
<td>&lt;0.001</td>
<td>0.735</td>
</tr>
<tr>
<td>Proportion of Socially Rented Houses</td>
<td>0.001</td>
<td>0.644</td>
<td>-5.430</td>
<td>0.935</td>
<td>&lt;0.001</td>
<td>0.510</td>
</tr>
</tbody>
</table>
C. Model Building

The results from the correlation analysis were used to build the model, ensuring that the model did not contain two or more independent variables that had strong correlations with each other. From the first regression analysis, the average number of rooms was the strongest variable and was included in the model. Therefore the variables with which it was strongly correlated (proportion of detached houses, proportion of owner occupied households, and proportion of flats) were excluded. The second stage of modelling showed that housing density was a stronger predictor of residential gas consumption than population density, so the population density variable was removed from the model. This left six values in the modelling process: average number of rooms, median household income, number of heating degree days, proportion of social rented households, and the proportion of terraced houses. Building the regression model using the stepwise entry method indicated that housing density was not statistically significant at the 0.01 level and was not included in the final model. This left five predictor variables of average number of rooms, median household income, number of heating degree days, proportion of terraced houses, and proportion of socially renting households. The multiple regression analysis proceeded using these five variables, with the results of the analysis shown in table VI.

These results were used to generate benchmarks for per meter gas consumption (y) using equation (1):

\[ y = -7033 + 2544a + 0.16b + 3.59c - 10.20d - 5.43e \]

Where \( y \) = LSOA per meter gas consumption (kWh), \( a \) = average number of rooms, \( b \) = median household income (£ per year), \( c \) = number of heating degree days (in 2006), \( d \) = percentage of terraced houses, \( e \) = percentage of social rented houses.

The model constructed had a \( R^2 \) value of 0.644 and was statistically significant at the 0.01 level (\( F_{6,31949} = 9697, p<0.001 \)). The regression equation with the five variables explains 64.4% of the geographical variation in per meter gas consumption in English LSOAs.

D. Examination of Residuals

The distribution of the residuals from the regression model was approximately normal with a mean of zero (Fig.2). These results satisfy the assumptions of multiple linear regression, that the residuals are random and not correlated with the predicted value, satisfying the assumption of constant variance (Fig.3). The plot of predicted values against actual values (Fig.4) indicates that the predictions from the model are well correlated with the actual recorded values, although there is a grouping of LSOAs with relatively high levels of household gas consumption that are under-predicted by the model, causing a shift away from the \( y=x \) line. The results for these LSOAs will be investigated in further study as part of the checks on the accuracy and precision of the multiple regression analysis.

![Fig. 2 Histogram of residuals from regression model](image1.png)

![Fig. 3 Plot of model residuals against predicted values](image2.png)
Future work to improve upon the results presented here is intended to revise the heating degree day variable by using the same base years as those used for weather correcting the DECC gas consumption data. This will allow results to be compared over time without the need to update the values of heating degree days in subsequent years. Comparisons of the numbers of meters and number of households from the various sources will be undertaken to understand the number of commercial gas meters may have been misclassified as residential meters, which is essential to ensure statistics on residential energy consumption are not skewed by wrongly classifying industrial consumers. Such misclassification may partly explain the skewing of the data shown in Fig.1. There will also be an investigation into those LSOAs where the predicted consumption values deviate from the measured consumption (Fig.4), this may be due to data misclassification.

Looking forward, it is intended that the model will be redeveloped to incorporate the 2011 Census results to account for changes in the English housing stock and tenure proportions since 2001. The results of a refined model will be compared against models developed using the 2009 LSOA data and the forthcoming 2010 releases to see if a strategy for tracking improvements in residential energy efficiency can be developed. Finally, a similar model to the one presented here will be calculated for LSOA per meter electricity consumption, which would further aid Local Authorities in reducing overall energy consumption from the residential sector.

V. CONCLUSION

DECC have promoted Local Authorities as their preferred scale for implementing policies such as Green Deal and ECO that will lead to large scale renovations of the existing UK housing stock to reduce residential gas consumption and aid the UK in meeting its 2050 CO₂ reduction targets. Local Authorities responsibilities to these schemes are to identify areas for intervention and to promote uptake for Green Deal schemes and direct ECO funding. The preliminary analysis presented here demonstrates that in the UK geographical variations in the size of house, median household income, external air temperature, proportion of terraced houses and proportion of socially renting households accounts for 64% of the geographical variation in average residential gas consumption at LSOA level using a multiple regression model.

Local Authorities could use the information generated from the multiple regression model and compare against the actual recorded figures from DECC. LSOAs which have higher than expected residential gas consumption figures could be targeted for energy efficiency policies, through either Green Deal or ECO.

REFERENCES


