Department of Energy & Climate Change

# ANNEX C

## Predicting gas consumption

June 2016

# Contents

# Introduction

This report details some novel analysis carried out by DECC into factors that can predict domestic gas consumption. This work builds upon previously published work that has used NEED variables to model gas consumption[1], with the considerable advantage of pulling together data from the main NEED model with household information from the English Housing Survey (EHS)[2] and from DECC's Fuel Poverty analysis[3].

There is increasing interest in identifying factors that can predict residential energy consumption and much research has been carried out looking at both socio-demographic and psychological factors[4]. DECC is in a unique position to contribute to this debate as it has access to administrative data from a wide range of sources that are linked up through NEED.

The socio-demographic factors can be split into those relating to the people in the household versus those relating to the building itself. The sample size and accurate data held by DECC gives us a powerful starting point and, in order to use some indicator of behaviour, we can use previous gas consumption as a crude proxy to indicate the influence of some of the psychological contributors.

### Summary of results

The modelling presented here indicates that building and demographic characteristics accounted for 44 per cent of the variance in domestic gas consumption. A hierarchical extension of the model, adding previous consumption as a proxy measure for previous behaviour, caused a large increase in model fit (to 80.8 per cent). This suggests that the behaviour of householders explains a significant proportion of the overall variation in the demand for gas.

Floor area was found to be the most important building characteristic for determining demand, and the presence of children the most important demographic characteristic. The latter may reflect a preference of those with young families to heat their home to a higher temperature, and also a tendency for the home to be

[1] NEED Report 2012 Annex E and Annex F https://www.gov.uk/government/statistics/national-energy-efficiency-data-need-report-summary-of-analysis
[2] https://www.gov.uk/government/collections/english-housing-survey
[3] https://www.gov.uk/government/collections/fuel-poverty-statistics
[4] See Frederiks, Stenner & Hobman (2015) for a recent review. *Energies* **2015**, 8, 573-609; doi:10.3390/en8010573

occupied (and heated) a greater proportion of the time.  The age of the oldest resident was also found to be significant, with over the 60s being associated with higher demand.  Again this is likely due to retired individuals spending more time in the home.

# Sample selection

The NEED full dataset contains approximately 26 million records across England, Wales and Scotland, while the EHS sample contains approximately 6,200 records per year (combined to give ~12,000 across each two year period) where interviews and physical surveys were carried out[5]. The EHS dataset for any given year contains data from two combined financial years to boost the same size of those with a physical survey. We aimed to include all households that were present in both NEED and the EHS, where the address could be matched with a high degree of certainty. Note that not all addresses within the EHS sample consent for their data to be used in this way. We only included addresses with a valid gas consumption figure for 2012 and 2013 (i.e., between 2,500 and 50,000 kWh), which were not identical for both years and were based on actual rather than estimated meter readings. This gave a sample of 6,870 records.

[5] https://www.gov.uk/government/publications/english-housing-survey-2012-to-2013-technical-report

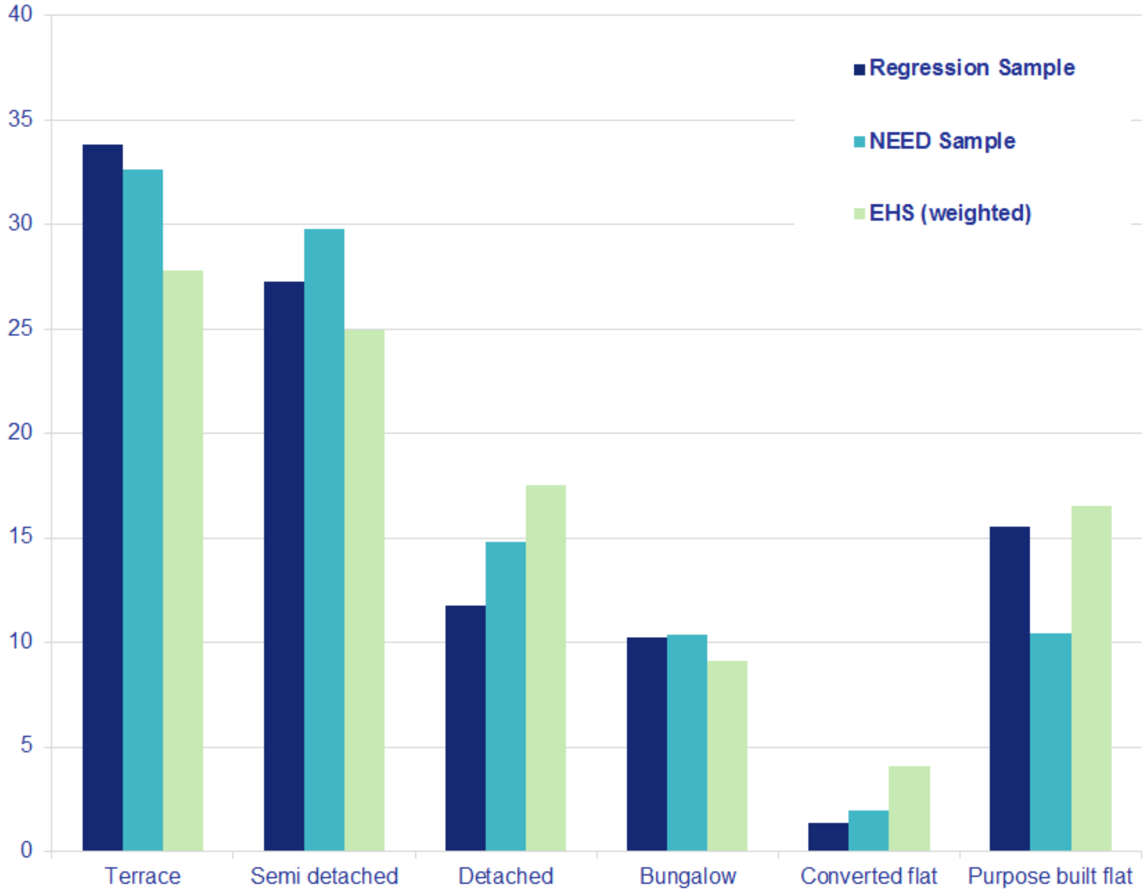
# Limitations

The households included in the analysis here were selected based on having a full set of EHS data, including physical survey. The EHS does not sample representatively across housing stock; it over-samples from the private rented sector to ensure sufficient data is collected for analysis[6]. The rented sector has a disproportionate number of flats and this means that they are over-represented in the sample.

The EHS produces a weighting factor for each property[7], which can be used to weight each household's data in order to make the sample representative of the whole housing stock. When we compared the results of our final sample here against what we would expect from a whole GB sample, we found that, as expected, flats were slightly over-represented compared to the NEED sample. Chart 1 shows the breakdown of counts by dwelling type comparing the regression sample here against the NEED sample and the full weighted EHS sample. The sample used here is, therefore, not completely representative of the full GB housing stock, but does not appear from the chart to be unduly biased towards flats.

[6] See English housing survey 2012 to 2013: technical report (Chapter 1: Sampling) https://www.gov.uk/government/publications/english-housing-survey-2012-to-2013-technical-report
[7] See English housing survey 2012 to 2013: technical report (Chapter 6: Weighting) https://www.gov.uk/government/publications/english-housing-survey-2012-to-2013-technical-report

## Figure C1: Percentage of properties by property type: Comparisons between data sources, 2014



# Predictor variable selection

Using a combination of EHS and NEED variables gives a very large number of possible predictors to use in a regression model. The list of possible predictors we selected to be included in our analysis is shown in Table 1, with a summary of what the code names stand for and the data values in the sample. For full details of the EHS variables, see the EHS data dictionary[8] and for further information on DECC's fuel poverty variables, see the methodology and user manual[9]. Some variables showed a high proportion of missing data, which mainly reflect questions in the EHS that are only applicable to home owners or to tenants, such as equity in home or weekly rent. These variables were not included in any of the analysis to avoid any need to impute mean values where they would not be appropriate.

This wealth of data, based on objective rather than self-report measures, is a significant advance of doing this analysis through NEED, but also represents a challenge; there will be high inter-correlations between many of the variables. We chose to carry out a Principal Components Analysis (PCA) in order to adopt a data-driven approach to predictor section. We included all of the scale variables in the dataset, as well as any nominal variables with five or more levels. This gave us 45 variables to include, where identifying the underlying factors could allow us to select the most appropriate predictors for the later regression model to predict gas consumption.

For the PCA, we wanted to include the different demographic and household characteristic variables in the data to see how they clustered together. We did not include any variables relating to energy use at this stage in order to keep the two steps of the analysis (PCA to select variables then regression to predict gas consumption) independent. This included variables relating to energy costs from the Fuel Poverty dataset.

[8] http://doc.ukdataservice.ac.uk/doc/6923/mrdoc/pdf/6923dictionary_of_derived_variables_2012_14.pdf

[9] https://www.gov.uk/government/publications/fuel-poverty-methodology-handbook-2013

## Table 1: Variables in the sample

| Variable name | Variable description | Source | Data Type | Mean | Min | Max | Missing | Included in PCA? |
|---|---|---|---|---|---|---|---|---|
| adults | Number of adults in HH | Experian | Scale | 2.01 | 0.00 | 9.00 | 0% | Yes |
| ageoldx | Age of oldest resident | EHS | Scale | 52.93 | 17.00 | 98.00 | 0% | Yes |
| agepartx | Age of partner | EHS | Scale | 48.99 | 18.00 | 93.00 | 46% | No |
| AHCIncomeEQ | Equivalised after housing costs annual income (£) | EHS | Scale | 21,120.58 | -24,367.73 | 377,724.00 | 0% | Yes |
| arnatx | Type of area (city to rural scale - 6 categories) | EHS | Ordinal | 2.94 | 1.00 | 6.00 | 0% | Yes |
| attic | Attic present? | EHS | Binary | 1.92 | 1.00 | 2.00 | 0% | Yes |
| basement | Basement present? | EHS | Binary | 1.99 | 1.00 | 2.00 | 0% | Yes |
| bedrqx | Number of bedrooms required (modelled) | EHS | Scale | 1.71 | 1.00 | 7.00 | 0% | Yes |
| BEDS | Number of bedrooms (1 to 5+) | Experian | Ordinal | 2.70 | 1.00 | 5.00 | 0% | Yes |
| boiler | Type of boiler | EHS | Nominal | n/a | n/a | n/a | 1% | No |
| boiler_flag | Has boiler insulation been fitted? | NEED | Binary | 0.26 | 0.00 | 1.00 | 0% | Yes |
| cookcost | Cooking cost | FP | Scale | 41.26 | 25.27 | 96.93 | 0% | Yes |
| cstactbx | Basic repair costs (actual) | EHS | Scale | 1,278.82 | 0.00 | 49,958.52 | 0% | Yes |
| cstactcx | Comprehensive repair costs (actual) | EHS | Scale | 3,575.16 | 0.00 | 75,273.03 | 0% | Yes |
| cstactux | Urgent repair costs (actual) | EHS | Scale | 820.34 | 0.00 | 45,085.22 | 0% | Yes |
| cststdbx | Basic repair costs standardised per square metre | EHS | Scale | 12.56 | 0.00 | 455.29 | 0% | Yes |
| cststdcx | Comprehensive repair costs standardised per square metre | EHS | Scale | 38.85 | 0.00 | 732.01 | 0% | Yes |
| cststdux | Urgent repair costs standardised per square metre | EHS | Scale | 8.00 | 0.00 | 400.00 | 0% | Yes |
| cwi_flag | Has cavity wall insulation been fitted? | NEED | Binary | 0.25 | 0.00 | 1.00 | 0% | Yes |
| draughtproof_flag | Has draughtproofing been fitted? | NEED | Binary | 0.01 | 0.00 | 1.00 | 0% | Yes |
| dwage9x | Date of dwelling construction - 9 categories | EHS | Ordinal | 5.40 | 1.00 | 9.00 | 0% | Yes |
| dwtypenx | Dwelling type - 8 categories | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| EHS_adults | Number of adults in property | EHS (derived) | Scale | 1.81 | 0.00 | 8.00 | 0% | Yes |
| emphrp3x | Employment status of HH reference person - 3 categories | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| EPceir12e | Enviromental impact rating | EHS | Scale | 58.13 | 4.10 | 95.81 | 0% | Yes |
| famnumx | Number of family units in househould | EHS | Scale | 1.09 | 1.00 | 7.00 | 0% | Yes |
| fit_flag | Has a solar PV system under FITs been installed? | NEED | Binary | 0.03 | 0.00 | 1.00 | 0% | Yes |
| floorx | Usable floor area | EHS | Scale | 82.53 | 21.04 | 519.40 | 0% | Yes |

| Variable name | Variable description | Source | Data Type | Mean | Min | Max | Missing | Included in PCA? |
|---|---|---|---|---|---|---|---|---|
| FPEER Band | Fuel Poverty Energy Efficiency Ratings | FP | Ordinal | 3.90 | 1.00 | 7.00 | 0% | Yes |
| fpvuln | Fuel poverty vulnerable marker (elderly, disabled, children) | FP | Binary | 0.80 | 0.00 | 1.00 | 0% | Yes |
| Fuel cost (equiv) | Equivalised fuel costs | FP | Scale | 1,226.88 | 491.35 | 4,039.36 | 0% | Yes |
| fuelexpn | Total fuel expenditure | FP | Scale | 1,259.78 | 402.91 | 5,331.96 | 0% | Yes |
| GCons2012 | Gas consumption in 2012 | NEED | Scale | 13,091.01 | 2,522.00 | 47,739.00 | 0% | No |
| GCons2013 | Gas consumption in 2013 | NEED | Scale | 12,644.90 | 2,504.00 | 49,647.00 | 0% | No |
| glazing_flag | Has double glazing been installed? | NEED | Binary | 0.16 | 0.00 | 1.00 | 0% | No |
| heat4x | Main heating system - 4 categories | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| heat7x | Main heating system - 7 categories | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| hhbensx | HH on means tested benefits/tax credits with income below threshold? | EHS | Binary | 1.63 | 1.00 | 2.00 | 0% | Yes |
| hhcompx | HH composition - 7 categories | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| hhltsick | Anyone in HH have illness/disability? | EHS | Binary | 1.62 | 1.00 | 2.00 | 0% | Yes |
| hholdincome | HH income (categories) | Experian | Ordinal | 2.52 | 0.00 | 9.00 | 0% | Yes |
| hhsizex | Number of people in HH | EHS | Scale | 2.48 | 1.00 | 11.00 | 0% | Yes |
| hhvulx | HH vulnerable? (On certain means tested or disability benefits) | EHS | Binary | 1.60 | 1.00 | 2.00 | 0% | Yes |
| HousingCosts | Any rent or mortgage repayments | FP | Scale | 4,761.06 | 0.00 | 68,400.00 | 0% | Yes |
| hpregdis | HH reference person or partner registered disabled? | EHS | Binary | 1.89 | 1.00 | 2.00 | 1% | Yes |
| imd1010 | IMD 2010 decile ranking of areas (deprivation declines) | EHS | Ordinal | 4.82 | 1.00 | 10.00 | 0% | Yes |
| lenres2 | Modelled length of residence (based on 2011 lenres) | EHS | Scale | 13.16 | 0.00 | 41.32 | 0% | Yes |
| litecost | Lighting cost | FP | Scale | 450.39 | -268.50 | 1,535.69 | 0% | Yes |
| loft_flag | Has loft insulation been installed? | NEED | Binary | 0.26 | 0.00 | 1.00 | 0% | Yes |
| loftinsx | Loft insulation thickness | EHS | Scale | 166.52 | 0.00 | 325.00 | 10% | Yes |
| loncoupx | Single HHer or with partner? | EHS | Binary | 0.54 | 0.00 | 1.00 | 0% | Yes |
| NBedsX | Total number of bedrooms | EHS | Scale | 2.68 | 1.00 | 7.00 | 0% | Yes |
| ndepchild | Number of dependent children in HH | EHS | Scale | 0.67 | 0.00 | 8.00 | 0% | Yes |
| olderx | Number of people (HRP plus partner) who are aged above 60 | EHS | Scale | 0.50 | 0.00 | 2.00 | 0% | No |
| othfamlp | Type of additional families in HH | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| propage | Property age banded - 6 categories | Experian | Ordinal | 4.25 | 1.00 | 6.00 | 0% | Yes |
| pyngx | Age of youngest person in HH | EHS | Scale | 36.20 | 0.00 | 95.00 | 0% | Yes |

| Variable name | Variable description | Source | Data Type | Mean | Min | Max | Missing | Included in PCA? |
|---|---|---|---|---|---|---|---|---|
| reslength | Residency length in years (11 categories) | Experian | Ordinal | 7.87 | 0.00 | 11.00 | 0% | Yes |
| rumorph | Rurality - 8 categories | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| sap12 | Energy efficiency (SAP12) rating | EHS | Scale | 62.31 | 1.00 | 93.47 | 0% | Yes |
| spahcost | Space heating costs | FP | Scale | 615.67 | 124.28 | 3,328.28 | 0% | Yes |
| swi_flag | Has solid wall insulation been installed? | NEED | Binary | 0.01 | 0.00 | 1.00 | 0% | Yes |
| tenure4x | Tenure - owner occupied/privated rented/local authority/RSL | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| Unoc | Is the property under-occupied? | FP | Binary | 0.24 | 0.00 | 1.00 | 0% | Yes |
| wallinsy | Type of wall and insulation | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| watersys | Water heating system - 5 categories | EHS | Nominal | n/a | n/a | n/a | 0% | No |
| wathcost | Water cost | FP | Scale | 152.46 | 48.58 | 834.64 | 0% | Yes |

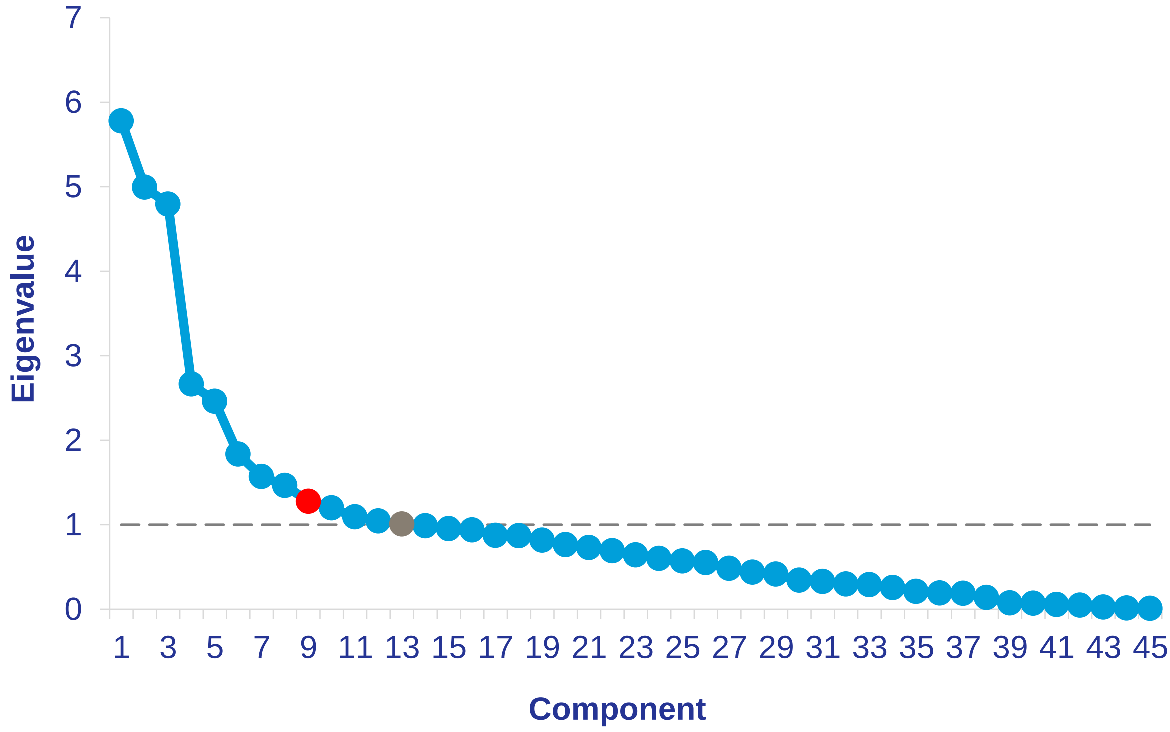
Key: CWI = Cavity wall insulation, EHS = English Housing Survey, FITs = Feed-in Tariffs, FP = Fuel Poverty, HH = Household, HRP = Household reference person, IMD = Index of multiple deprivation, NEED = National Energy Efficiency Data-framework, PV = Photovoltaic, RSL = Registered social landlord, SAP = Standard Assessment Procedure, SWI = Solid wall insulation.

## Principal Components Analysis

The PCA was conducted on the 45 items identified in Table 1, using a non-orthogonal rotation (Promax with Kaiser Normalization). Due to some missing data, the sample size for the analysis was 6110. Sampling adequacy was verified using the Kaiser-Meyer-Olkin measure, where KMO = 0.72, indicating that this is a suitable sample for PCA[10]. All individual KMO values were above 0.5 (lowest value 0.585 for cststdbx). Correlations between items were significantly large to justify the use of PCA, as indicated by Barlett's test of sphericity, $\chi^2 = 208{,}272$, $p < 0.001$. The average communality was 0.693.

When the analysis was run to identify the components in the data, 13 components had eigenvalues greater than 1 and in combination, they could explain 69 per cent of the variance. The scree plot did not show one single clear inflection point, but arguably indicated that nine factors should be included, see Chart 2. Nine components could account for 60 per cent of the variance in the data. The red point on the chart indicates a possible inflection point at component 9 while the orange point indicates component 13, which is the final component with an eigenvalue greater than 1. Given our large sample size, all 13 components were kept.

**Figure C2: Scree plot for PCA**




[10] Hutcheson, G & Sofroniou, N (1999) *The multivariate social scientist: Introductory statistics using generalized linear model.* Thousand Oaks, CA, USA: Sage Publications.

Table 2 shows the factor loadings after rotation in the pattern matrix, as well as suggested names for each component based on which variables clustered together. The structure matrix was highly consistent with the pattern matrix, with some subtle differences noted for factors 6. The full structure matrix is presented in the appendix for comparison. Note that for both factor loadings tables, loadings of less than 0.5 have been suppressed for clarity.

Based on the factor loadings in the pattern and structure matrices, the components appear to represent:

1. Immediate repair costs (required within five years)
2. Household composition
3. Energy efficiency
4. Property size
5. Length of residence
6. Number of adults
7. Marker of vulnerability
8. Urbanity
9. Comprehensive repair costs
10. Number of families
11. Loft insulation
12. Non-lived on floors
13. Presence of solid wall insulation or draught proofing

## Table 2: Pattern matrix for PCA

| Component | 1<br>Imm. Repairs | 2<br>HH composition | 3<br>Energy Efficiency | 4<br>Property Size | 5<br>Length of residence | 6<br># of adults | 7<br>Vulnerable | 8<br>Urbanity | 9<br>Comp. repairs | 10<br># of families | 11<br>Loft insulation | 12<br>Non-lived on floors | 13<br>Insulation installed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cststdux | 0.967 | | | | | | | | | | | | |
| **cstactux** | 0.963 | | | | | | | | | | | | |
| cststdbx | 0.872 | | | | | | | | | | | | |
| cstactbx | 0.857 | | | | | | | | | | | | |
| **ndepchild** | | 0.936 | | | | | | | | | | | |
| hhsizex | | 0.839 | | | | | | | | | | | |
| bedrqx | | 0.811 | | | | | | | | | | | |
| **pyngx** | | -0.702 | | | | | | | | | | | |
| **sap12** | | | 1.006 | | | | | | | | | | |
| FPEER Band | | | -.972 | | | | | | | | | | |
| EPceir12e | | | 0.954 | | | | | | | | | | |
| dwage9x | | | 0.503 | | | | | | | | | | |
| NBedsX | | | | 0.842 | | | | | | | | | |
| **floorx** | | | | 0.828 | | | | | | | | | |
| beds | | | | 0.828 | | | | | | | | | |
| Unoc | | | | 0.639 | | | | | | | | | |
| **lenres2** | | | | | 0.849 | | | | | | | | |
| reslength | | | | | 0.833 | | | | | | | | |
| HousingCosts | | | | | -0.683 | | | | | | | | |
| **ageoldx** | | | | | 0.613 | | | | | | | | |
| loncoupx | | | | | | 0.931 | | | | | | | |
| **EHS_adults** | | | | | | 0.757 | | | | | | | |
| adults | | | | | | 0.598 | | | | | | | |
| **hpregdis** | | | | | | | 0.825 | | | | | | |
| hhltsick | | | | | | | 0.796 | | | | | | |
| hhvulx | | | | | | | 0.683 | | | | | | |
| hhbensx | | | | | | | 0.572 | | | | | | |
| **arnatx** | | | | | | | | 0.736 | | | | | |
| imd1010 | | | | | | | | 0.618 | | | | | |
| **cststdcx** | | | | | | | | | 0.935 | | | | |
| cstactcx | | | | | | | | | 0.896 | | | | |
| **famnumx** | | | | | | | | | | 0.919 | | | |
| loft_flag | | | | | | | | | | | 0.850 | | |
| **loftinsx** | | | | | | | | | | | 0.726 | | |
| **basement** | | | | | | | | | | | | 0.730 | |
| **attic** | | | | | | | | | | | | 0.631 | |
| **swi_flag** | | | | | | | | | | | | | 0.662 |
| **draughtproof_flag** | | | | | | | | | | | | | 0.622 |

Key: Comp = comprehensive, HH = household, Imm = immediate. Dashed line indicates inflection point of scree plot.

## Variables selected for regression

Based on the components from the PCA, we selected variables to use in the regression. Although we could have used factor scores from the PCA, this would not have allowed us to include some of the nominal variables also in the dataset. In addition, using 'raw' predictor variables has the advantage of being directly related to real world measures and therefore being easier to interpret for policy decisions.

Where possible, we prioritised data from EHS over modelled data from Experian. Where there were scale and ordinal variables contributing to a component, we prioritised scale variables or ordinal scales with as many levels as possible. We also selected some nominal variables not included in the PCA where they looked likely to fit well within some components and where there has been evidence in the literature that they are linked to energy consumption. We also selected variables where there was sufficient variability in the data; e.g., the addition of heating system type was considered as 99 per cent of the sample used a boiler system with radiators, this left very few cases to cover the other possible categories. The selected variables are listed below;

1. Urgent repair costs – cstactux (actual repair costs)
2. Household composition – hhcompx and ndepchild
3. Energy efficiency – sap12 (Rating band of dwelling)
4. Property size – floorx (useable floor area)[11]
5. Length of residence – lenres2
6. Age – ageoldx and pyngx (Age of oldest and youngest resident)
7. Number of adults – EHS_adults
8. Vulnerability marker – hpregdis (Household reference person or partner registered disabled)
9. Urbanity – arnatx
10. Comprehensive repair costs – cststdcx (standardised repair costs)
11. Number of family units in household –famnumx
12. Non lived on floors – attic and basement
13. Loft insulation thickness - loftinsx
14. Energy efficiency measures fitted –swi_flag, draughtproof_flag
15. Dwelling type – dwtypenx[12]

[11] Previous work modelling gas consumption using NEED has only included floor areas ranging between 2 and 500 meters squared, based on advice from the VOA (https://www.gov.uk/government/statistics/national-energy-efficiency-data-need-report-summary-of-analysis). As shown in Table 1, our data in the regression sample ranged from 21 to 519 meters squared for floor area (floorx). Only one property had a useable floor area above 500 meters squared and no data was excluded.

These variables can broadly be split into household versus building characteristics;

- Building characteristics
    - Property size
    - Repair costs (urgent and comprehensive)
    - Energy efficiency (SAP plus measures fitted)
    - Dwelling type
    - Non-lived on floors
    - Urbanity
- Household demographics
    - Length of residence
    - Household composition (number of adults/family units and age of youngest/oldest resident, plus nominal composition variable)
    - Vulnerability

[12] This variable has eight categories but some were combined to match those most commonly presented within NEED. Low and high rise purpose built flats were considered as a single category of 'purpose built' while end and mid terrace were combined to a single category of 'terraced houses'.

# Regression analysis

We carried out a hierarchical linear regression with three levels. These levels attempted to group potential predictor variables into three categories:

1. Building characteristics
2. Household demographic characteristics
3. Household behaviour (previous gas consumption)

The variables included at each step are summarised in the previous section, with the exception of gas consumption in 2012, which was added in stage three. This approach means that we could separately identify what proportion of the variance in gas consumption in 2013 could be explained by the characteristics of the building first, then see if adding in household demographic information would improve the model fit. Finally, we could add in a measure of energy use behaviour to see how much additional variance could be accounted for.

For the nominal variables included in the regression model (dwtypenx and hhcompx), dummy variables were created. For dwelling type, purpose built flats were used as the reference group since the main NEED analysis indicated that this property type had the lowest gas consumption. For household composition, we used the category of 'one person under 60' as the reference group, since this group showed the lowest mean gas consumption compared to the other six categories. An initial analysis indicated a concern of multicolinearity with the age of youngest person in household variable (VIF[13]=14.0) and therefore this predictor was excluded from the final analysis. The choice of predictor variables meant that our sample size slightly increased to 6130, due to less missing data being present. Data at each stage of the regression used the forced entry method. We avoided more exploratory methods so that the results can hopefully be extrapolated from more easily and the model can be compared/used with different samples.

The results from the regression are shown in full in Table 3. The highest VIF was 6.22 and the lowest tolerance was 0.16. The Durbin-Watson statistic was 2.02. The model was significant at stage one, $F(15,6115) = 286.07$, $p<0.001$, where the building characteristics accounted for 41 per cent of the variance in the data. Adding in demographic characteristics significantly improved the model fit ($\Delta R^2 = 0.028$, $F(12,6103)=25.79$, $p<0.001$), so that it now accounted for 43 per cent of the variance in gas consumption. Notably, adding in a measure of previous behaviour in stage

[13] Variance inflation factor (VIF)

three caused a large increase in model fit ($\Delta R^2 = 0.369$, $F(1,6102)=11{,}793.28$, $p<0.001$). The final model accounted for 81 per cent of the variance in gas consumption.

Looking to individual coefficients in the final model (stage three), only a limited number of building characteristics were significant predictors. These variables were urgent repair costs (cstactux), energy efficiency rating (sap12), property size (floorx), urbanity (arnatx), the presence of an attic (attic) and having solid wall insulation fitted (swi_flag). Notably, none of the housing type categories were significant. The direction of the coefficients was mostly as expected, for example having a larger floor area led to higher consumption as does the presence of an attic (note that in the EHS, attic is coded '1' for yes and '2' for no, so a negative coefficient here is expected). A higher energy efficiency rating lead to lower gas consumption, as did the installation of solid wall insulation.

Half of the household demographic characteristics were significant predictors: being a couple (over or under 60, with or without children) or being a lone parent or multi person household. The age of the oldest resident was also significant. All types of household apart from single residents over 60 showed increased gas consumption relative to our reference group of single under 60 households.

One perhaps surprising result is that having higher urgent repair costs led to lower gas consumption. This is presumably because having urgent repairs that are not completed is an indicator of financial strain, meaning less energy is consumed over concern over paying bills.

Notably, we found an association between the age of the eldest householder and energy usage, such that as this age increased, so did gas consumption. Age is not always found to be a significant predictor of energy consumption, although the measure used is often age of household head rather than age of eldest resident[14].

The size of the standardised coefficients highlights that previous gas use is by far the biggest predictor of current gas use. The largest contributor from the building characteristics is useable floor space and the largest contributor from the demographic characteristics is being a couple with children.

[14] See Frederiks, Stenner & Hobman (2015). *Energies* **2015**, 8, 573-609; doi:10.3390/en8010573

## Table 3: Regression Results

| Model | | Unstandardised Coefficients B | Unstandardised Coefficients SE B | Standardised Beta | t | Significance |
|---|---|---|---|---|---|---|
| Step 1 | Constant | 19593.082 | 1406.295 | | 13.932 | < 0.001 |
| | cstactux | -0.063 | 0.039 | -0.018 | -1.603 | 0.109 |
| | sap12 | -151.960 | 7.808 | -0.208 | -19.461 | < 0.001 |
| | floorx | 86.299 | 2.271 | 0.483 | 37.997 | < 0.001 |
| | arnatx | 94.176 | 100.524 | 0.010 | 0.937 | 0.349 |
| | cststdcx | 2.830 | 1.255 | 0.025 | 2.255 | 0.024 |
| | attic | -1441.155 | 249.757 | -0.060 | -5.770 | < 0.001 |
| | basement | -1654.303 | 594.924 | -0.028 | -2.781 | 0.005 |
| | loftinsx | -1.403 | 0.821 | -0.018 | -1.709 | 0.087 |
| | swi_flag | -1163.833 | 624.356 | -0.018 | -1.864 | 0.062 |
| | draughtproof_flag | 295.071 | 535.285 | 0.005 | 0.551 | 0.581 |
| | terrace | 695.834 | 277.595 | 0.051 | 2.507 | 0.012 |
| | semi_detached | 1903.774 | 289.343 | 0.132 | 6.580 | < 0.001 |
| | detached | 2854.807 | 358.081 | 0.144 | 7.973 | < 0.001 |
| | bungalow | 1807.782 | 325.928 | 0.086 | 5.547 | < 0.001 |
| | converted_flat | 998.911 | 866.768 | 0.012 | 1.152 | 0.249 |
| Step 2 | Constant | 16657.628 | 1510.469 | | 11.028 | < 0.001 |
| | cstactux | -0.077 | 0.038 | -0.021 | -1.999 | 0.046 |
| | sap12 | -153.198 | 7.690 | -0.210 | -19.921 | < 0.001 |
| | floorx | 78.553 | 2.280 | 0.440 | 34.460 | < 0.001 |
| | arnatx | 102.065 | 98.544 | 0.010 | 1.036 | 0.300 |
| | cststdcx | 3.393 | 1.227 | 0.030 | 2.765 | 0.006 |
| | attic | -1465.039 | 244.760 | -0.061 | -5.986 | < 0.001 |
| | basement | -1856.559 | 581.540 | -0.032 | -3.192 | 0.001 |
| | loftinsx | -1.513 | 0.809 | -0.019 | -1.869 | 0.062 |
| | swi_flag | -1404.041 | 611.255 | -0.022 | -2.297 | 0.022 |
| | draughtproof_flag | 279.043 | 523.804 | 0.005 | 0.533 | 0.594 |
| | terrace | 76.990 | 275.724 | 0.006 | 0.279 | 0.780 |
| | semi_detached | 1097.300 | 288.018 | 0.076 | 3.810 | < 0.001 |
| | detached | 2126.003 | 352.977 | 0.108 | 6.023 | < 0.001 |
| | bungalow | 1181.994 | 334.913 | 0.056 | 3.529 | < 0.001 |
| | converted_flat | 947.611 | 847.604 | 0.011 | 1.118 | 0.264 |
| | couple_under_60 | 1645.235 | 328.395 | 0.089 | 5.010 | < 0.001 |
| | couple_over_60 | 1454.538 | 335.162 | 0.082 | 4.340 | < 0.001 |
| | couple_children | 2504.952 | 374.401 | 0.159 | 6.691 | < 0.001 |
| | lone_parent | 2125.985 | 350.926 | 0.104 | 6.058 | < 0.001 |
| | multi_person | 1541.453 | 382.004 | 0.064 | 4.035 | < 0.001 |
| | single_over_60 | 68.874 | 323.730 | 0.004 | 0.213 | 0.832 |
| | lenres2 | 18.907 | 6.465 | 0.039 | 2.924 | 0.003 |
| | ageoldx | 47.643 | 7.232 | 0.121 | 6.588 | < 0.001 |
| | EHS_adults | 412.694 | 136.232 | 0.050 | 3.029 | 0.002 |
| | hpregdis | -325.098 | 217.173 | -0.015 | -1.497 | 0.134 |
| | famnumx | 150.366 | 227.418 | 0.009 | 0.661 | 0.509 |
| | ndepchild | 242.749 | 107.439 | 0.039 | 2.259 | 0.024 |
| Step 3 | Constant | 2600.495 | 891.540 | | 2.917 | 0.004 |
| | **cstactux** | **-0.053** | **0.022** | **-0.015** | **-2.360** | **0.018** |
| | **sap12** | **-23.982** | **4.646** | **-0.033** | **-5.162** | **< 0.001** |
| | **floorx** | **12.332** | **1.464** | **0.069** | **8.422** | **< 0.001** |
| | **arnatx** | **136.363** | **57.549** | **0.014** | **2.370** | **0.018** |
| | cststdcx | 0.502 | 0.717 | 0.004 | 0.700 | 0.484 |
| | **attic** | **-362.293** | **143.297** | **-0.015** | **-2.528** | **0.011** |
| | basement | -638.305 | 339.796 | -0.011 | -1.878 | 0.060 |
| | loftinsx | -0.358 | 0.473 | -0.005 | -0.758 | 0.448 |
| | **swi_flag** | **-906.545** | **356.994** | **-0.014** | **-2.539** | **0.011** |
| | draughtproof_flag | 150.819 | 305.897 | 0.003 | 0.493 | 0.622 |
| | terrace | -9.904 | 161.021 | -0.001 | -0.062 | 0.951 |
| | semi_detached | 105.529 | 168.446 | 0.007 | 0.626 | 0.531 |
| | detached | 321.412 | 206.802 | 0.016 | 1.554 | 0.120 |
| | bungalow | 96.098 | 195.840 | 0.005 | 0.491 | 0.624 |
| | converted_flat | -94.595 | 495.082 | -0.001 | -0.191 | 0.848 |
| | couple_under_60 | 871.016 | 191.911 | 0.047 | 4.539 | < 0.001 |
| | couple_over_60 | 823.046 | 195.816 | 0.047 | 4.203 | < 0.001 |
| | couple_children | 1041.067 | 219.060 | 0.066 | 4.752 | < 0.001 |
| | lone_parent | 637.138 | 205.394 | 0.031 | 3.102 | 0.002 |
| | multi_person | 641.222 | 223.239 | 0.027 | 2.872 | 0.004 |
| | single_over_60 | 200.321 | 189.058 | 0.011 | 1.060 | 0.289 |
| | lenres2 | -0.093 | 3.780 | 0.000 | -0.025 | 0.980 |
| | ageoldx | 10.959 | 4.237 | 0.028 | 2.587 | 0.010 |
| | EHS_adults | -97.994 | 79.697 | -0.012 | -1.230 | 0.219 |
| | hpregdis | 60.355 | 126.876 | 0.003 | 0.476 | 0.634 |
| | famnumx | 257.809 | 132.813 | 0.015 | 1.941 | 0.052 |
| | ndepchild | 14.958 | 62.778 | 0.002 | 0.238 | 0.812 |
| | GCons2012 | 0.810 | 0.007 | 0.818 | 108.597 | < 0.001 |

Key: See Table 1 for variable definitions. SE = standard error

# Summary and next steps

By combining data from NEED with that of the EHS, we have managed to create a model of domestic gas consumption for 2013 data that can account for approximately 80 per cent of the variance in our dataset.  We have used a data driven method to select predictor variables, selecting building and demographic characteristics to include based on results from a principal components analysis.  Just using these socio-demographic predictors allowed us to produce a model more successful than previous attempts using NEED (44 per cent of variance accounted for here, compared to 37 per cent in 2008).

Most notably, a third step in our hierarchical regression added in gas consumption in 2012 as a proxy for previous behaviour.  We do not have any measures in NEED that can directly address the psychological factors that contribute to household energy consumption, which will be varied and complex, covering personality, motivation, sense of responsibility etc.  Using a previous measure of behaviour in this way, in a hierarchical analysis, demonstrates that once household and building characteristics are accounted for, previous behaviour can significantly increase model fit over and above what you would expect based on the socio-demographic characteristics.  More information on attitudes and other psychological influences would be of great interest to break down the reasons behind behaviour, particularly with respect to targeting policy interventions surrounding energy consumption.

## Appendix Table 1: Structure matrix from PCA

| Component | 1<br>Imm.<br>Repairs | 2<br>HH<br>composition | 3<br>Energy<br>Efficiency | 4<br>Property<br>Size | 5<br>Length of<br>residence | 6<br># of<br>adults | 7<br>Vulnerable | 8<br>Urbanity | 9<br>Comp.<br>repairs | 10<br># of<br>families | 11<br>Loft<br>insulation | 12<br>Non-lived<br>on floors | 13<br>Insulation<br>installed |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| cststdux | 0.936 | | | | | | | | | | | | |
| cstactux | 0.935 | | | | | | | | | | | | |
| cststdbx | 0.920 | | | | | | | | | | | | |
| cstactbx | 0.910 | | | | | | | | 0.508 | | | | |
| bedrqx | | 0.870 | | | | | | | | | | | |
| hhsizex | | 0.869 | | | | | | | | | | | |
| ndepchild | | 0.868 | | | | | | | | | | | |
| pyngx | | -0.789 | | | 0.620 | | | | | | | | |
| sap12 | | | 0.944 | | | | | | | | | | |
| EPceir12e | | | 0.941 | | | | | | | | | | |
| FPEER Band | | | -0.900 | | | | | | | | | | |
| dwage9x | | | 0.652 | | | | | | | | | | |
| PROPAGE | | | 0.538 | | | | | | | | | | |
| floorx | | | | 0.846 | | | | | | | | | |
| NBedsX | | | | 0.845 | | | | | | | | | |
| BEDS | | | | 0.824 | | | | | | | | | |
| Unoc | | -0.509 | | 0.538 | | | | | | | | | |
| lenres2 | | | | | 0.806 | | | | | | | | |
| ageoldx | | -0.507 | | | 0.769 | | | | | | | | |
| Reslength | | | | | 0.711 | | | | | | | | |
| HousingCosts | | | | | -0.661 | | | | | | | | |
| loncoupx | | | | | | 0.811 | | | | | | | |
| EHS_adults | | | | | | 0.794 | | | | .592 | | | |
| adults | | | | | | 0.609 | | | | | | | |
| hholdincome | | | | 0.500 | | 0.557 | | | | | | | |
| hhvulx | | | | | | 0.538 | 0.815 | | | | | | |
| hhbensx | | | | | | 0.566 | 0.742 | | | | | | |
| hhltsick | | | | | | | 0.695 | | | | | | |
| hpregdis | | | | | | | 0.645 | | | | | | |
| fpvuln | | | | | | | -0.520 | | | | | | |
| imd1010 | | | | | | | | 0.657 | | | | | |
| arnatx | | | | | | | | 0.602 | | | | | |
| cstactcx | 0.535 | | | | | | | | 0.862 | | | | |
| cststdcx | 0.501 | | | | | | | | 0.852 | | | | |
| famnumx | | | | | | | | | | 0.851 | | | |
| loft_flag | | | | | | | | | | | 0.808 | | |
| loftinsx | | | | | | | | | | | 0.697 | | |
| attic | | | | | | | | | | | | 0.567 | |
| swi_flag | | | | | | | | | | | | | 0.549 |
| draughtproof_flag | | | | | | | | | | | | | 0.535 |

Key: Comp = comprehensive, HH = household, Imm = immediate