Modelling the likelihood of being fuel poor

Background

This article examines the impact certain household and dwelling characteristics have on the likelihood of a household being classed as fuel poor under the Low Income High Costs (LIHC) indicator. Under this indicator, a household is considered to be fuel poor where:

i) they have fuel costs that are above average (the national median level); and
ii) were they to spend that amount, they would be left with a residual income below the official poverty line (i.e. less than 60 per cent of median income).

The aim of this analysis is to develop a model of the most influential characteristics – that are readily identifiable in published data sources and organisations such as Local Authorities – which help determine the probability of households being fuel poor. The modelling set out in the following section has been reviewed (and approved) by the Office for National Statistics Methodology Advisory Service.

It should be noted that the model provides an indicative probability of the likelihood that a household is living in fuel poverty based on a set of known characteristics. However, this does not lead to a definitive classification of these individual households as fuel poor.

Logistic regression modelling

The logistic regression modelling technique assesses how certain characteristics within a household, such as employment status or the type of boiler they have in the house, may affect the likelihood of that household being fuel poor. For example, will a full-time working couple with dependent children living in a block of flats with a 7 year old heating system, be more likely to be classed as fuel poor compared to their next door neighbours, who also live under very similar circumstances but have recently installed a new heating system?

The advantage of using logistic regression is that it is able to verify whether the patterns commonly seen across fuel poverty are actually associated with single characteristics or a combination of a number of characteristics. For example, households in which the main reference person is unemployed are also more likely to be living in fuel poverty compared to the overall population (36% vs. 11%). By holding household characteristics such as the amount of energy consumption constant and equal, logistic regression helps isolate which of these factors – unemployment or the type of heating system – has a stronger association with an increase in the odds of such households to be fuel poor. The modelled output proceeds to show that unemployment is a factor with a great effect on the odds of such households to be living in fuel poverty.

Table 1 summarises the household and dwelling characteristics considered in the modelling to reliably predict households living in fuel poverty under the LIHC indicator.

The table also details the baseline or ‘reference’ for each variable, and the final set of variables that are retained in the final model. The reference category is the one with which all other categories are compared. For example, the odds of being fuel poor for all family compositions are compared against couples with no dependent children.

Table 1: Variables considered in the modelling

Variables	Reference Category	Low Income High Costs
Family Composition	Couple, no dependent child(ren)	
Household size	Number of persons in the household >=5	●
Age band of youngest person in household	Aged between 16 to 59	●
Individual(s) disabled or with chronic illness	No disabled household members or unknown	
Employment status of household reference person	HRP - Full/Part-time employment	●
Employment status (primary) of partner	Partner - Full/Part-time employment	●
National Statistics Socio-Economic Classification	Higher managerial and professional occupations	●
Household on means tested benefits/tax credits	No	●
Attendance allowance or DLA mobility/care component	No or No Answer	●
Method of payment - electricity	Direct debit	●
Method of payment - gas	Direct debit	
Government office region	South East	
Rurality - morphology (COA)	Urban	
Whether dwelling is on the gas network	On gas network	●
Dwelling type	Flat	●
Dwelling age	Post1964	●
Total no of bedrooms	One bedroom	
Useable floor area	Less than 50 sqm	●
Tenure	Local Authority/RSL	●
Under occupancy	Not under occupying	●
Energy efficiency rating band (SAP 2005)	A, B or C	
Loft insulation thickness	150mm or more	●
Type of wall and insulation	Cavity with insulation	
Age of heating system	Less than 3 years	●
Main heating fuel	Gas	●
Main heating system	Central heating	
Water heating system	With central heating	
Type of boiler	All condensing boiler	●

● Included in the final model

The regression model outputs show the individual effect each characteristic has on the odds of a household to be fuel poor, compared to a household with the baseline set of reference characteristics. Characteristics with an odds ratio greater than 1, implies an increased likelihood that a household with that particular characteristic will be in fuel poverty compared to the reference characteristic; conversely, an odds ratio less than 1 implies a reduced likelihood - holding all other characteristics constant and equal.

This is graphically shown in the Chart 1, where the bars indicate the proportionate effect on the odds for each category compared with the baseline reference category. An increase in odds (odds ratio > 1) is shown with a right hand bar, and a decrease (odds ratio < 1), with a left hand bar. The confidence intervals for the effects of each category are also shown in the charts – where these are shorter in length, the more precise is the estimate of the associated odds ratio. Where a confidence interval spans the value of 1, this indicates that the effect of the category is not significantly different from the baseline category. It should be noted that the scale of the chart is logarithmic rather than linear.

The size of the effects, the corresponding 95% confidence intervals, and Wald statistics are provided in Table 2. The validation tests for the model are also provided.

Modelled output for households living in fuel poverty

Chart 1 shows the final modelled outcome for predicting households that are fuel poor. The model was created by using a backward elimination procedure, where variables were dropped from the model as they were not found to be statistically significant. In this work, the following variables were dropped as they were not statistically significant or in some cases due to multicollinearity[1] between the variables: family composition, disability, method of payment for gas, region, rurality, the number of bedrooms in the household and the main heating and water heating system.

Other variables – such as SAP[2] rating - were deliberately left out of the model. Whilst we would expect that this would be a strong predictor of household energy costs, we know that SAP rating is determined by a number of other dwelling characteristics (e.g. heating type, level of insulation, size of dwelling) and, as such, is not as useful in determining the specific factors that are driving households to be fuel poor.

The findings from the regression analysis for all LIHC households (Chart 1) are intuitive. Household characteristics associated with higher instances of modelled energy costs and low incomes tend to be the ones that increase the likelihood of being fuel poor. Holding all other characteristics constant and equal, it is apparent that against the baseline characteristic for each group:

- Single one person households have higher odds of being fuel poor compared to larger households with more occupants in – here the odds are almost four times that of households with five or more occupants.

- Households with children aged below 16 also significantly increase the odds of being fuel poor by almost 20%. On the other hand, pension aged households (where the youngest household member is aged 60 or over) have almost half the odds of being fuel poor compared to younger households. This may be due to the fact that such households are likely to have reduced housing costs, and therefore a higher level of equivalised disposable income, compared to younger households.

- The odds of being fuel poor more than double for households in which the main household reference person (HRP) is either unemployed or inactive[3] compared to households where the HRP is employed. Retired HRPs also show a 27% increase in the odds of being fuel poor compared to their employed counterparts.

- In addition, having a retired or unemployed partners' increase the odds of being fuel poor by over three-fold compared to households in which the partners are in some form of employment.

- Households on means-tested benefits also have increased odds of being fuel poor – an almost four-fold increase is seen for households on mean-tested benefits compared to those not on benefits.

- Households that are off the gas grid network are reliant on using alternative main fuel (i.e. other than mains gas) such as electricity or 'other[4]' fuel types. As a result, the effect of being off the gas grid should be viewed in conjunction with the main fuel type consumed. The odds of being fuel poor therefore are 44% higher for households that are off the gas grid and

---

[1] Multicollinearity occurs when two or more predictor variables in the model are highly correlated and so provide redundant additional information about the response variable in the model – in this case, whether or not the household is in fuel poverty.

[2] SAP is the Government's Standard Assessment Procedure for Energy Rating of Dwellings. It used for calculation of the energy performance of buildings.

[3] Economically inactive people include those who are in full time education, the permanently sick or disabled, or those looking after the family or home or engaged any other activity.

[4] 'Other' fuel types include: anthracite nuts, bottled gas, bulk LPG, heating oil, house coal, smokeless fuel and wood.

consuming electricity as their main fuel type, and similarly 38% higher for those consuming 'other' fuel types – compared to households on the mains gas network.

- Of the dwelling types, households living in bungalows or detached properties have the highest odds of being fuel poor (3.52 times that of flats) followed by households living in semi-detached or terraced properties (odds ratios of 3.22 and 2.32 respectively). Also, households living in older properties, generally tend to have increased odds of being fuel poor compared to more recently built properties.

- The odds of being fuel poor increase notably for properties with floor spaces above $50m^2$.Households living in properties larger of $110m^2$ or more, have the largest odds of being fuel poor followed by those living in properties with floor spaces between 90 – 109m2 and 70 – 89m2 (with odds ratios of 17.89, 11.73 and 7.05 respectively).

- Under-occupied[5] households have reduced odds of being fuel poor - around half the odds of households which are not under-occupied.

- Households living in privately rented accommodation have over twice the odds of being fuel poor compared to households in social housing. This is most likely due to the fact that the energy efficiency in the housing stock across both these tenures are in stark contrast to each other – the energy efficiency across the social housing stock is generally better than average, and that in the private rental market is considerably worse.

- And finally, households that have non-condensing boilers all have increased odds of being fuel poor.

Reviewing the importance of these household and dwelling characteristics on the odds of being fuel poor, the largest and most significant[6] increases in the odds are seen for households living in larger and older properties. Households on means tested benefits are also at a significant risk of being fuel poor, as are households in which the main reference person or their partner is not in active employment.

It is possible to convert the odds effects described above into probabilities of being fuel poor for households with any particular combinations of characteristics from the model[7]. The individual effects (see Table 2) are multiplied together to find an overall effect which is then converted to a probability[8]. Take for example the following households:

[5] Some dwellings are considered excessive in size for the number of occupants that live there. In these cases, the house is assumed to be "under-occupied", that is only a proportion of the dwelling will need heating.
[6] The rank order of significance for a variable is shown by the magnitude of the corresponding Wald statistic. High Wald statistics imply an increased significance to the model
[7] Note, any number of variable combinations can be selected here as shown in the proceeding example.
[8] Probability = odds/(1+odds)

Household A	Odds	Household B	Odds
• 3 person household	0.98	• 2 person household	1.31
• Unemployed HRP	2.59	• Employed HRP	1.00
• Unemployed Partner	3.07	• Employed Partner	1.00
• On income related benefits	3.87	• Not income related benefits	1.00
• Living in a Terraced Property	2.32	• Living in a flat	1.00
• 1940's build	3.45	• 1990's build	1.00
• Property size: 70-89 sqm	7.05	• Property size: 70-89 sqm	7.05
• Not underoccupying	1.00	• Underoccupying	0.50
• With loft insulation	1.00	• No loft	0.97
• No boiler	2.78	• Combination boiler	1.26
The remaining characteristics are the reference characteristics specified in the model			
Multiplied effects	4730.60		5.64

The model gives the odds of being fuel poor for the reference household of 0.0001 or 0.01% (Table 2).

The example household A has 4730.60 times these odds of being fuel poor

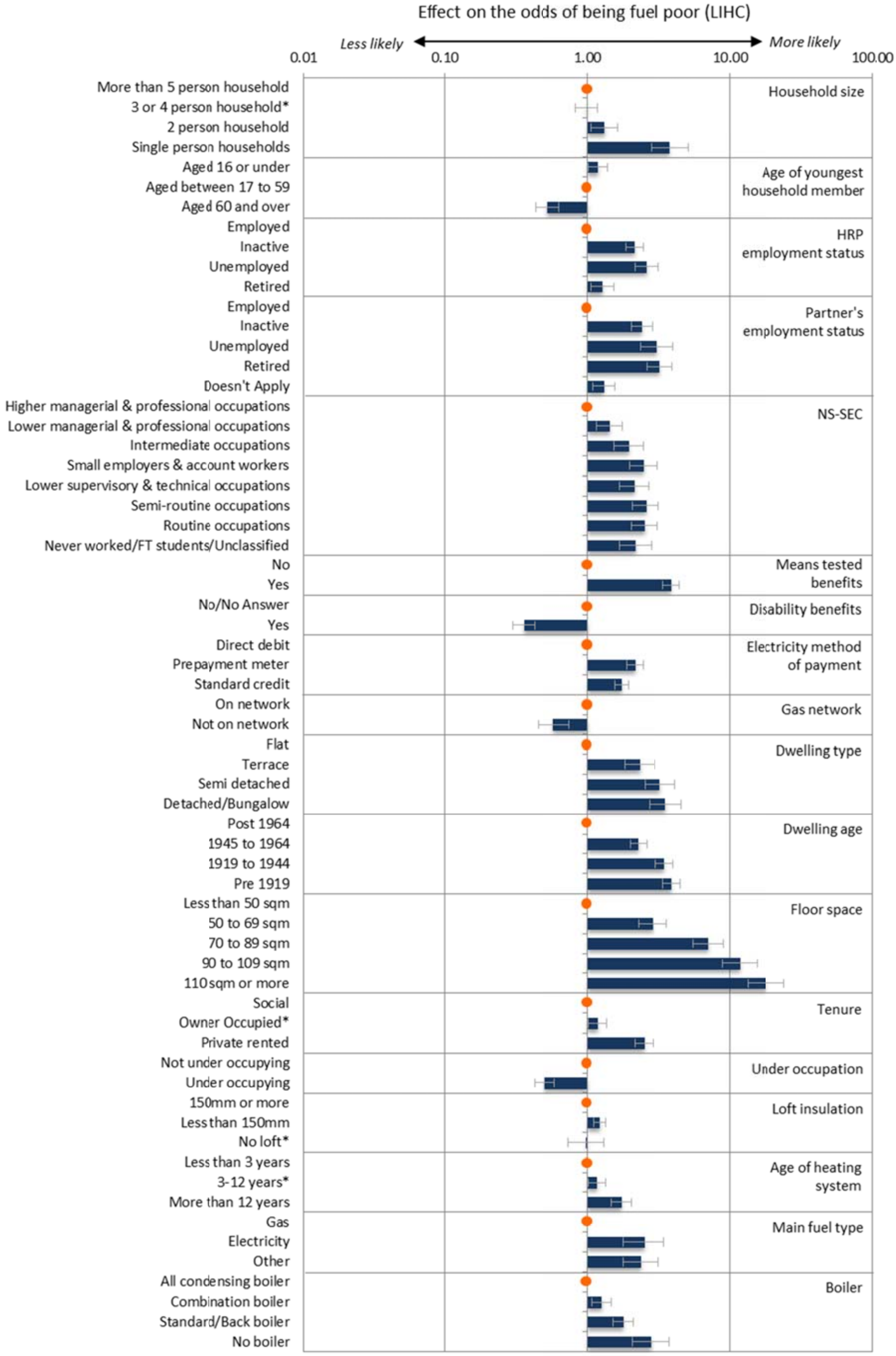
(4730.60*0.0001) = 0.4736, or
(0.4736/[1+0.4736] = 32.1%.

And household B has 5.64 times these odds of being fuel poor

(5.64*0.0001) = 0.000564, or
(0.000564/[1+0.000564]) = 0.1%.

So household A’s probability of being fuel poor is 32.1% compared to household B’s 0.1% probability and the overall population average of 11.1%.

Chart 1: Effect of characteristics on the odds being fuel poor, 2010



*Not statistically significant

Source: English Housing Survey

Table 2: Regression results

Low Income High Costs - Variable	Variable catagories	Effect on the odds	95% Confidence interval Lower limit	95% Confidence interval Upper limit	Sig. (0.05)	Wald	B	S.E.
Household size	More than 5 person household	1.00						
	3 or 4 person household*	0.98	0.83	1.16	0.81	0.06	-0.02	0.09
	2 person household	1.31	1.06	1.63	0.01	6.12	0.27	0.11
	Single person households	3.77	2.78	5.12	0.00	72.55	1.33	0.16
Age band of youngest person in household	Aged 16 or under	1.18	1.02	1.37	0.03	4.67	0.17	0.08
	Aged between 17 to 59	1.00						
	Aged 60 and over	0.52	0.43	0.63	0.00	47.02	-0.65	0.10
Employment status of household reference person	Employed	1.00						
	Inactive	2.13	1.84	2.46	0.00	102.44	0.76	0.07
	Unemployed	2.59	2.13	3.13	0.00	94.14	0.95	0.10
	Retired	1.27	1.06	1.52	0.01	6.71	0.24	0.09
Employment status (primary) of partner	Employed	1.00						
	Inactive	2.39	2.01	2.83	0.00	98.41	0.87	0.09
	Unemployed	3.07	2.36	4.00	0.00	69.28	1.12	0.13
	Retired	3.21	2.62	3.95	0.00	123.56	1.17	0.11
	Doesn't Apply	1.30	1.09	1.54	0.00	8.70	0.26	0.09
National Statistics Socio-Economic Classification	Higher managerial & professional occupations	1.00						
	Lower managerial & professional occupations	1.42	1.15	1.75	0.00	10.68	0.35	0.11
	Intermediate occupations	1.94	1.53	2.46	0.00	29.83	0.66	0.12
	Small employers & account workers	2.46	1.96	3.09	0.00	60.47	0.90	0.12
	Lower supervisory & technical occupations	2.12	1.68	2.69	0.00	39.53	0.75	0.12
	Semi-routine occupations	2.56	2.06	3.16	0.00	74.08	0.94	0.11
	Routine occupations	2.51	2.02	3.12	0.00	68.16	0.92	0.11
	Never worked/FT students/Unclassified	2.16	1.67	2.80	0.00	33.93	0.77	0.13
Household on means tested benefits/tax credit	No	1.00						
	Yes	3.87	3.41	4.41	0.00	424.20	1.35	0.07
Attendance allowance or DLA mobility/care component	No/No Answer	1.00						
	Yes	0.35	0.30	0.42	0.00	139.65	-1.04	0.09
Method of payment - electricity	Direct debit	1.00						
	Prepayment meter	2.14	1.87	2.46	0.00	117.64	0.76	0.07
	Standard credit	1.73	1.54	1.93	0.00	88.95	0.55	0.06
Whether dwelling is on the gas network	On network	1.00						
	Not on network	0.57	0.45	0.74	0.00	18.80	-0.56	0.13
Dwelling type	Flat	1.00						
	Terrace	2.32	1.83	2.96	0.00	47.50	0.84	0.12
	Semi detached	3.22	2.51	4.12	0.00	85.99	1.17	0.13
	Detached/Bungalow	3.52	2.72	4.55	0.00	92.44	1.26	0.13
Dwelling age	Post 1964	1.00						
	1945 to 1964	2.26	1.98	2.58	0.00	143.71	0.82	0.07
	1919 to 1944	3.45	3.00	3.98	0.00	295.96	1.24	0.07
	Pre 1919	3.89	3.38	4.48	0.00	356.47	1.36	0.07
Floor area	Less than 50 sqm	1.00						
	50 to 69 sqm	2.86	2.27	3.60	0.00	78.83	1.05	0.12
	70 to 89 sqm	7.05	5.51	9.01	0.00	242.98	1.95	0.13
	90 to 109 sqm	11.73	8.87	15.49	0.00	299.64	2.46	0.14
	110 sqm or more	17.89	13.31	24.06	0.00	364.67	2.88	0.15
Tenure	Social	1.00						
	Owner Occupied	1.17	1.01	1.36	0.03	4.49	0.16	0.08
	Private rented	2.49	2.15	2.87	0.00	156.71	0.91	0.07
Under occupancy	Not under occupying	1.00						
	Under occupying	0.50	0.42	0.58	0.00	76.23	-0.70	0.08
Loft insulation thickness	150mm or more	1.00						
	Less than 150mm	1.21	1.10	1.34	0.00	14.63	0.19	0.05
	No loft*	0.97	0.73	1.30	0.86	0.03	-0.03	0.15
Age of heating system	Less than 3 years	1.00						
	3-12 years	1.16	1.01	1.34	0.04	4.32	0.15	0.07
	More than 12 years	1.72	1.46	2.02	0.00	43.49	0.54	0.08
Main heating fuel	Gas	1.00						
	Electricity	2.48	1.78	3.45	0.00	28.92	0.91	0.17
	Other	2.35	1.77	3.13	0.00	34.23	0.85	0.15
Type of boiler	All condensing boiler	1.00						
	Combination boiler	1.26	1.07	1.47	0.00	8.15	0.23	0.08
	Standard/Back boiler	1.77	1.51	2.07	0.00	50.90	0.57	0.08
	No boiler	2.78	2.05	3.77	0.00	43.27	1.02	0.16
Model constant	Constant	0.0001			0.00	1390.81	-9.08	0.24

* Not statistically significant

Model Validation

The Hosmer and Lemeshow test provides an overall fit of the logistic regression model and tests whether the difference between the observed and expected values are statistically significant. A finding of non-significance implies that the model adequately fits the data. At a 5% level of significance, this test is found to be insignificant (p-value: 0.380) and therefore the logistic regression model above is valid.

Hosmer and Lemeshow Test

Step	Chi-square	df	Sig.
1	8.568	8	0.380

The accuracy of the model to discriminate between fuel poor and non-fuel poor households is evaluated using the Receiver Operating Characteristic (ROC) curve. The Area under this curve (AUC), known as the c-statistic, can range from 0.5 (no predictive ability) to 1 (perfect discrimination). The statistically significant value of 0.851 shows this model offers a very good level of discrimination.

Area Under the Curve

Area	Std. Error	Asymptotic Sig	Asymptotic 95% Confidence Interval	
			Lower Bound	Upper Bound
0.851	0.004	0.000	0.844	0.858

Masuma Ahmed
Fuel Poverty statistics
Tel: 0300 068 5922
E-mail: Masuma.Ahmed@decc.gsi.gov.uk