

Levels of over- indebtedness in the UK

Technical Report

2017 Over-Indebtedness Model

1. Foreword

There are many challenges we need to address if we are to deal more effectively with problem debt in the UK. Which groups of people are most likely to be over-indebted? Are those groups constant over time? How do we allocate resources effectively to deliver the best outcomes?

In order to begin to answer some of those questions, we have estimated levels of over-indebtedness across the UK since 2013 and, for the last three years, commissioned CACI to model those levels in ever-increasing detail. Our focus is to estimate levels of over-indebtedness across different geographic areas alongside relevant demographic information. The granular level of the estimation is remarkable, achieving estimates for Lower Tier Local Authorities and, for the first time this year, to Lower Super Output Area.

At a time when public budgets are tight and resources stretched, timely information is essential for decision making. For our part, we make use of the data by comparing levels of estimated need with local supply, across all channels, to inform our strategic approach to the commissioning of debt advice.

We invite both public and private organisations, small and large, who are passionate about helping people with debt problem in the UK to make use of this data. Whatever the challenge is in your community, the information provides a picture of the potential need which we believe can assist with service design, shape funding decisions and identify opportunities to bring about positive change.

Dr. Jair Muñoz Bugarin
Service Design Manager
The Money Advice Service

2. Background and Objectives

The Money Advice Service has been measuring individuals' levels of over-indebtedness since 2012. As part of The Money Advice Service's strategy, there is a requirement to generate a measure of over-indebtedness that can be applied geographically across the United Kingdom, and updated regularly.

This helps the service to:

- forecast service demand
- map and manage funding for services and the resources required
- understand the factors associated with over-indebtedness

CACI first worked with The Money Advice Service in 2015 to produce a nationwide model, which combined large numbers of survey respondents with CACI's rich consumer data, and resulted in over-indebtedness scores for a range of geographies. The approach was a "bottom-up" methodology, meaning individuals were modelled separately and then aggregated into regions based on their residential postcode. This granular method is not only more robust than modelling "local averages", but is also more flexible and allows other geographic (or indeed individual or household-level) analysis in the future.

In 2016 and 2017, The Money Advice Service collected additional and updated research data, which CACI used to test, validate, and where necessary update the over-indebtedness model. This ensures a current and robust view of existing levels of over-indebtedness, as well as the predicting factors and the characteristics of the over-indebted population.

The 2017 model is a "refresh" of the 2016 model – this means that although the 2016 model failed initial validation, only minor changes to the variables and parameters were required to reach a satisfactory model for 2017.

It is important for the solution, but also the annual update process, to be clear and understandable, taking a transparent approach to the way over-indebtedness is calculated on an annual basis. This report summarises the original approach, and details this year's changes to and findings from the Over-Indebtedness Model. The detailed results and data have been published in an interactive summary dashboard on the Money Advice Website¹.

¹ www.moneyadvice.org.uk/Over-indebtednessLevelsUK

3. Data Sources

3.1. Research Data

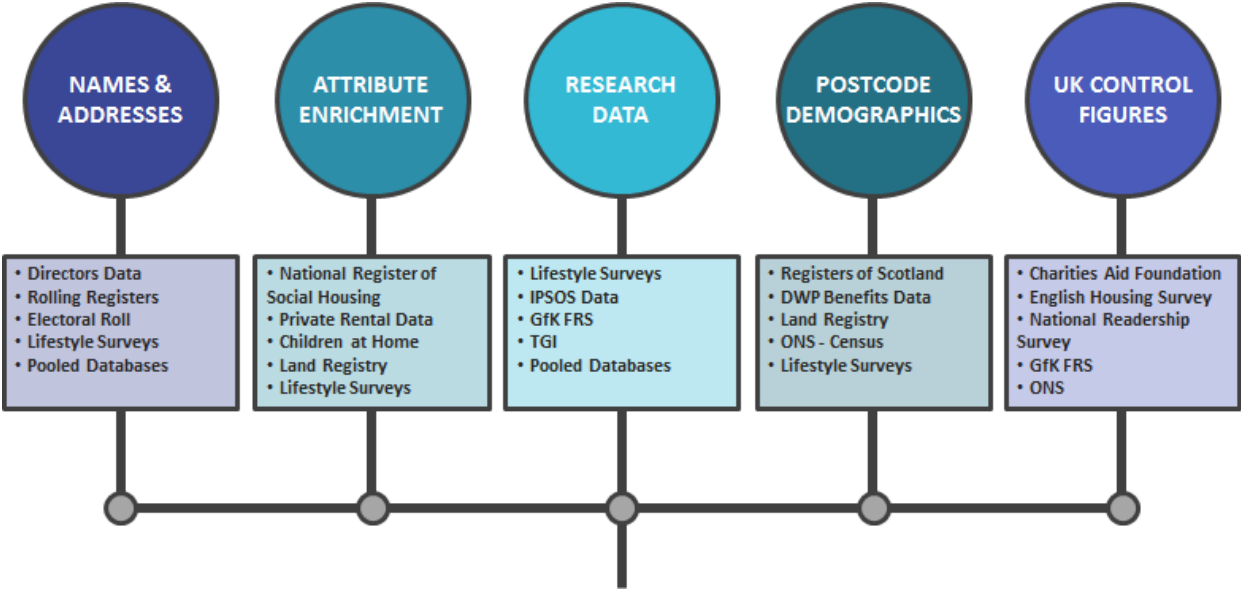
The Money Advice Service provided CACI with research survey data for analysis. The total sample size across two separate surveys (multiple waves) was approximately 20,000 individual respondents. Respondents were selected from both the YouGov Omnibus and the Research Now panel, which are two of the largest and most respected online panel providers in the UK. In terms of regional and demographic representativeness of the UK, and consistent levels of reported over-indebtedness, these two surveys were deemed suitable for combining and using together in the model update.

Consistent to the data used in 2016, and the majority of 2015 data used in the seminal model, the data originated from online panel research amongst UK consumers. The time period covered was July 2017, and respondents were de-duplicated across surveys and waves.

Survey	Conducted by	Sample Size	Mode	Period
Omnibus	YouGov	15,486	Online	July 2017
Panel	Research Now	4,848	Online	July 2017

3.2. CACI Ocean Data

Ocean was used to build the original over-indebtedness model in 2015, and then used each year to validate and refine the model. It is an attribute-rich consumer database for the UK, maintained by CACI and updated quarterly. Hundreds of millions of records from research surveys, open data, government data and many other sources are collated together to create the universe.



Ocean includes:

- Names and addresses of 48 million adults. The name and address base forms the 'spine' of the Ocean database. It is built by merging and de-duplicating names and addresses from multiple different high-volume sources, and selecting the most up to date information.
- 27 million date of birth records. Dates of birth are selected and combined from multiple sources. Where date of birth isn't available, age is imputed from a range of data including forename, household composition, region, product holdings and other lifestyle variables.
- A wide range of variables for each individual. Variables can be supplied as yes/no flags (i.e. indicator variables) where known or where they can be confidently imputed from other knowledge about the individuals.
- Values inferred from modelling based on other known characteristics. Modelled estimates can be provided as estimates of the probability that a person has an attribute (it is these propensity scores that are used for over-indebtedness modelling), as inferred Yes/No flags, or as categorical assignments for appropriate variables such as tenure.
- A set of composite indicators built by combining a number of variables to indicate attributes such as wealth, disposable income, etc.

The real and modelled variables on Ocean cover a wide range of attributes, attitudes and behaviour. They include:

ATTRIBUTES

Age and gender
Number and age of children
Household Income
Household size and composition
Length of residence
Housing: type, tenure, size, value
Occupation
Social Grade
Number, age and type of cars

ATTITUDES

Reading preferences; books and magazines
Charities: which causes supported and how
Newspaper readership
Attitudes to financial products and channels
Intention to switch financial products
Attitudes to online privacy and safety
Lifestyle attitudes
Shopping attitudes
Attitudes to the environment

FINANCIAL BEHAVIOUR

Financial products owned
Savings and Investments value
Credit card patterns of use
Loans and debt
Channel preference
Medical insurance

LIFESTYLE

Technology ownership and use
Holidays: destination, type, spend and booking method
Interests and hobbies
Smokers
Internet usage: frequency, location and technology
Types of goods and services purchased online
Online activities: gambling, dating, gaming etc.
Social networking: which networks and types of activity
Mobile phone: type of phone and how used
Shopping: types of stores visited (premium, mass, value)

4. Defining “Over-Indebtedness”

OVER-INDEBTED

Finds meeting monthly commitments a heavy burden and/or is regularly in arrears with bills

The Money Advice Service first investigated the characteristics of over-indebtedness in the 2013 research, “Indebted Lives: the complexities of life in debt”². The definition of over-indebtedness has remained consistent since then, and the component questions have been coded onto research surveys each year. Over-indebted individuals are those that answer either:

- i. I find meeting my monthly bills/commitments a **heavy burden**
- ii. I have missed bill payments in **three or more months** out of the last six months

Note that the three months in (ii) do not need to be consecutive. Individuals may respond positively to one or both of the above questions to be identified as over-indebted.³ Those that do not respond positively to either question are defined as “not over-indebted”.

These questions feed into a single “Yes/No” binary variable that was modelled to predict over-indebtedness at an individual level.

Within the supplied surveys, the average proportion of respondents finding bills a heavy burden was 12.8%, while 7.2% of respondents had been in arrears in three of the last six months.

The table below shows how these figures vary across the two different survey sources.

Survey Source	Respondents	Paying bills is a burden	Arrears in 3m of last 6m	Over-Indebted
YouGov	15,486	13.08%	6.83%	15.86%
Research Now	4,848	11.96%	8.40%	16.48%
Total	20,334	12.81%	7.20%	16.01%

The overlap between respondents who answered yes to both questions was 4%, thus the overall observed level of over-indebtedness in the surveys was 16%. This figure is comparable to that of The Financial Lives Survey conducted by the FCA in 2017, which reports that an estimated 14.6% of the UK adult population is over-indebted⁴. However it should be noted that the collection methodology differs to that of the MAS research, and includes both online and face-to-face interviews.

² <https://www.moneyadvice.service.org.uk/en/corporate/indebted-lives-the-complexities-of-life-in-debt>

³ “To what extent you feel that keeping up with your bills and credit commitments is a burden?” [A heavy burden; Somewhat of a burden; Not a burden at all; Don’t know].

“In the last 6 months, have you fallen behind on, or missed, any payments for credit commitments or domestic bills for any 3 or more months? These 3 months don’t necessarily have to be consecutive months.” [Yes; No ; Don’t know]

⁴ <https://www.fca.org.uk/publications/research/understanding-financial-lives-uk-adults>

6. Cleaning/De-duping Respondents and Tagging with Ocean Data

The first stage of the model update was to match all of the survey respondents to CACI's database of individuals. This appended the Ocean attributes and characteristics to each respondent, so that the model variables could be retested and validated against the dependent over-indebtedness variable derived from the research.

This stage also matches respondents to an individual within the UK consumer database, which allows for accurate de-duping amongst the surveys and waves. 65 records were removed as they appeared across multiple waves, and a further 172 records were removed as they represented individuals present on both of the survey panels. A further 140 records did not match to CACI Ocean universe, and therefore could not be used in the modelling process.

Survey Source	Individual or Household Match	Postcode Match	Duplicate Respondents	Unmatched to Ocean
YouGov	14,116	1,288	15	67
Research Now	4,370	355	50	73
Cross-Survey	-	-	172	-
Total Proportion	90%	8%	1%	1%

The total number of records matched by individual, household or postcode level was 20,269. After de-duping across surveys, the total sample size of usable records taken forward into the 2017 model validation was **19,957**.

This was split into a training sample (n=15,966) and a 20% validation sample (n=3,991) – the latter to independently verify a revised model using respondents that haven't contributed to the recalibration of its parameters.

7. Modelling Process

7.1. 2015 Over-Indebtedness Model

In 2015 CACI worked with the Money Advice Service to produce estimates of over-indebtedness for the UK, and for each local authority. This was based on a logistic regression analysis of 11,279 survey respondents, which modelled each individual's likelihood of being over-indebted. The technical report of this original model is available online⁵. The resulting model consisted of sixteen variables, and an interactive summary dashboard was published on the Money Advice Service website⁶.

In subsequent years, this model has been re-tested and validated against fresh research data.

7.2. Model Update Approach

Money Advice Service commissions new research each year in order to collect up-to-date data on levels of over-indebtedness within the United Kingdom. In both 2016 and 2017 the research came from two separate online panels, and the sample sizes were in the region of 20,000. This data is run through a three-step validation process which tests the suitability of the current over-indebtedness model:

1. Existing Model Validation
2. Model "Refresh"
3. Model "Rebuild"

Pass/fail criteria are set up for each step, and the process terminates if any of the steps pass. If a step fails, the next step is undertaken. In both 2016 and 2017, the existing model validation failed and a model refresh was required.

7.2.1. Existing Model Validation

The existing over-indebtedness model – i.e. its variables, parameters and intercepts – are applied to the latest market research data. The accuracy of the model, as well as the output of some other key statistical tests, is acceptable when the following criteria are met:

- New research data should be representative of the UK population, both demographically and regionally. Some minor variations can be permitted as the model is at individual level (and then aggregated to small areas), and CACI population data is able to take this into account. Each adult is represented within Ocean, and therefore weighting is not required. However large variations may indicate sample bias, and sufficient volumes of each region characteristic are required to ensure statistical significance.
- Within the sample, observed over-indebtedness should be no more than 10% higher or lower than that observed in the previous year. For 2017 this means it should fall between 12.8% and 15.7%.
- When the existing model is applied to the sample, the predicted level of over-indebtedness should be sufficiently close to the observed level amongst the respondents. An error of one standard deviation is permitted (which in practice means an acceptable error of around +/- one percentage point).
- The concordance statistic (*c*, a measure of individual-level accuracy) should be sufficiently close to that of previous models. Given the 2015 and 2016 models returned 71.1% and 69.0%, *c* should be greater than or equal to 70%. However *c*>65% will be permitted if the model performs well elsewhere.
- The fitted model should pass Hosmer-Lemeshow's "Goodness-of-Fit" test. As in previous years this will be assessed by analysing ten deciles and showing there is no evidence to support a lack of fit, with *p* – the probability of finding such lack of fit by chance – greater than or equal to 0.25.

⁵ www.moneyadvice.service.org.uk/Over-indebtednessLevelsUK_Reports

⁶ www.moneyadvice.service.org.uk/Over-indebtednessLevelsUK

7.2.2. Model Refresh

If one or more of the above conditions are not met, then minor adjustments are made to the over-indebtedness model to produce improvements in accuracy or robustness. This is achieved by applying a logistic regression to the new research data, forcing the same variables into the model to elicit new parameters and intercept. Additionally, variables may be removed if statistically insignificant, and variables from previous years' models may be tested in the model for improvements. Note that *new* variables are not tested until the third step, the model rebuild. A model is accepted at this stage if the following criteria are met:

- The concordance statistic (c , a measure of individual-level accuracy) should be sufficiently close to that of previous models. Given the 2015 and 2016 models returned 71.1% and 69.0%, c should be greater than or equal to 70%. However $c > 65\%$ will be permitted if the model performs well elsewhere.
- The fitted model should pass Hosmer-Lemeshow's "Goodness-of-Fit" test. As in previous years this will be assessed by analysing ten deciles and showing there is no evidence to support a lack of fit, with p – the probability of finding such lack of fit by chance – greater than or equal to 0.25.
- All variables retained in the model, as well as the intercept, should be statistically significant at a level greater than 95% confidence.
- All variables retained in the model should act in the same direction as previous years (i.e. positive and negative predictors remain positive and negative).
- A validation sample (20% of the research data) should be retained and tested. The proposed model should predict over-indebtedness within one percentage point and produce a concordance statistic of 70% (>65% if the model performs well elsewhere). This is a new test introduced in 2017, made possible by a larger sample size than previous years.

7.2.3. Model Rebuild

If one or more of the above conditions are not met, then full statistical analysis will be run on the new research data in order to build a new logistic regression model. In this case, additional variables will be sought, and new interactions tested. It should be noted that this stage has not been required thus far, and the 2017 over-indebtedness model remains largely consistent and comparable to the original model in 2015.

The 2017 Model Update process passes the Model Refresh step, and therefore Model Rebuild is not required

1. Existing Model
Validation



2. Model
Refresh



3. Model
Rebuild

8. Model Parameters

8.1. The 2015 and 2016 models

The 2015 over-indebtedness model consisted of sixteen variables, some of which were Ocean variables and some of which were combinations and interactions of variables. Twelve were positive factors (suggesting an increased likelihood of over-indebtedness), and four were negative factors (suggesting a decreased likelihood of over-indebtedness).

The model update in 2016 tested these sixteen variables against new research data. The 2015 model did not prove to satisfy the statistical criteria required – five of the variables were found to no longer be statistically significant and were therefore removed from the model. The remaining eleven parameters (and the intercept) were adjusted accordingly. The table below summarises these changes:

Parameter	2015 Parameter Co-efficients	2016 Parameter Co-efficients
<i>Intercept</i>	-1.925	-2.205
Has Loan for Consolidation	4.584	5.808
Private Renting	0.315	0.365
Social Renting	0.431	0.255
Has 3+ Children	1.050	1.159
Single Parent	0.209	-
Social Grade D or E	1.067	1.770
Northern Ireland	0.527	-
Value of Home <£100k, South East	0.831	-
Value of Home <£100k, London	4.464	-
Unemployed, Wales & West Midlands	1.952	-
Household Income <£10k, Household Size 3+	1.159	1.578
Own Home Outright, Wales	0.670	0.369
Has Savings £10k+	-2.127	-1.933
Aged 65-74	-0.919	-0.809
Aged 75+	-1.211	-1.012
Scotland	-0.259	-0.210

All coefficients are statistically significant at a 95% confidence level.

8.2. The 2017 model

8.2.1. Existing Model Validation

The 2017 research data, collected July 2017 and summarised in sections 2 and 3, presented a reported level of over-indebtedness of 15.9%. This is 1.6 percentage points higher than that reported within the 2016 data, and outside of the 10% range set within the acceptance criteria (i.e. greater than 15.7%). This suggests that either there has been an increase in the underlying level of over-indebtedness in the UK, or that the 2017 research data may not be entirely comparable to the 2016 research data, and that further investigation may be worthwhile.

The research data was, however, reasonably representative of the UK – both in terms of demographics and region – and more so than that of the previous year.

When the 2016 over-indebtedness model was applied to the 2017 research data, the modelled level of over-indebtedness was 14.7%. Despite the concordance statistic (a measure of individual-level predictability) being 71.1%, the overall forecast produced by the model was 1.2 percentage points lower than that reported by the respondents

and outside of the acceptable levels of error, specifically one standard deviation. Furthermore it was apparent the over-indebtedness model performed worse for middle-aged and family demographics.

Therefore the decision was made to reject the 2016 model, and move onto the second step of recalibrating the model with adjusted parameters.

8.2.2. Model Refresh

Running logistic regression with the eleven variables derived in 2016 on the 2017 research data (a 20% sample was reserved for validation testing, see section 8.5) suggested new parameters and a lower intercept – this latter point corresponding to the perceived increase in the underlying level of over-indebtedness amongst survey respondents.

One variable, the indicator for Scotland, saw a change of sign (it was negative in 2015 and 2016), but more importantly became statistically insignificant ($p=0.5398$). The implication being that there was no longer evidence to suggest that the population of Scotland was any less over-indebted than the rest of the UK, above and beyond the influences of other characteristics appearing in the model.

A second model was run that excluded Scotland as a variable, and this produced acceptable levels of statistical significance across all ten variables. Aside from small variations in the intercept and parameters, this second model was comparable to the first, but performed better statistically.

Parameter (2017)	11-variable Model Parameters	Pr > Chi Sq	10-variable Model Parameters	Pr > Chi Sq
<i>Intercept</i>	-1.862	<.0001	-1.850	<.0001
Has Loan for Consolidation	7.528	<.0001	7.449	<.0001
Private Renting	0.452	<.0001	0.446	<.0001
Social Renting	0.416	0.0002	0.410	<.0001
Has 3+ Children	1.079	0.0045	1.065	0.0002
Social Grade D or E	0.869	0.0010	0.873	0.0050
Household Income <£10k, Household Size 3+	1.567	0.0016	1.552	0.0009
Own Home Outright, Wales	0.541	0.0120	0.532	0.0017
Has Savings £10k+	-2.339	<.0001	-2.362	0.0133
Aged 65-74	-0.917	<.0001	-0.916	<.0001
Aged 75+	-1.071	<.0001	-1.071	<.0001
Scotland	0.047	0.5398		

All coefficients except Scotland are statistically significant at a 95% confidence level.

The five variables removed from the model between 2015 and 2016 were tested for re-inclusion. However none were statistically significant, and in fact produced a deterioration in model performance. Further information on the original 2015 model can be found on the Money Advice Service website.⁷

⁷ <https://www.moneyadviceservice.org.uk/en/corporate/a-picture-of-over-indebtedness-2016>

For the final model used, full variable statistics are given in the table below. The sign of the parameter coefficients indicates whether the variable has a positive or negative effect on over indebtedness. To understand the magnitude of a change in the dependent variable to the likelihood to be over indebted we need to look at the marginal probabilities. Presented in the last column of the following table, the average marginal probability describes how the likelihood of over-indebtedness changes given the presence of the variable (with all other things remaining constant). For example, an individual with a loan for consolidation is likely to be 72 percentage points more likely to be over-indebted than the same individual without a loan.

Standardised estimates of the coefficients take into account the distribution (mean and variance) of the independent variables, and so are more useful when interpreting each parameter's true effect and contribution to the prediction. For simplicity, the standardised estimates have been transformed into relative importance scores that indicate the weight of each variable within the model – their absolute values sum to 100, and the sign indicates the direction of their effect.

Parameter (2017 final model)	Estimated Parameter Coefficient	Standardised Estimate	Relative Importance Score	Average Marginal Probability
<i>Intercept</i>	-1.850			
Has Loan for Consolidation	7.449	0.075	8.1	72%
Private Renting	0.446	0.067	7.3	6%
Social Renting	0.410	0.066	7.1	5%
Has 3+ Children	1.065	0.030	3.2	13%
Social Grade D or E	0.873	0.075	8.1	11%
Household Income <£10k, Household Size 3+	1.552	0.042	4.5	19%
Own Home Outright, Wales	0.532	0.028	3.0	7%
Has Savings £10k+	-2.362	-0.185	-20.0	-29%
Aged 65-74	-0.916	-0.178	-19.2	-11%
Aged 75+	-1.071	-0.180	-19.5	-13%

Model uses 15,966 observations, of which 2,526 are over-indebted. All coefficients are statistically significant at a 95% confidence level.

The predictor factors that bear the most importance in the 2017 over-indebtedness model are those that suggest the individual is not in the age band 65-74 or 75+, and does not have a significant savings balance (greater than £10,000). These three variables are the same that the 2015 and 2016 models highlighted as having the highest importance score.

8.3. Variable Definitions

Has Loan for Consolidation

The likelihood (ranging from 0 to 1) that an individual has a loan for the purpose of consolidating existing debt.

Private Renting

The likelihood (ranging from 0 to 1) that an individual lives in a home that is rented privately.

Social Renting

The likelihood (ranging from 0 to 1) that an individual lives in a home that is rented through a local authority or housing association.

Has 3+ Children

The likelihood (ranging from 0 to 1) that an individual is aged 25-39 and has three or more children at home. The inclusion of the age criteria ensures that an effect is truly caused by the presence of children and not by other age-related secondary effects. For example the very old and very young are unlikely to have more than two children at

home, and so these individuals should be removed from the set with 3+ children that is being compared against. Other age criteria were examined in 2015, but 25-39 provided the strongest model.

Social Grade D or E

The likelihood (ranging from 0 to 1) that an individual is classified within the National Readership Survey (NRS) social grades D or E.

Household Income <£10k, Household Size 3+

The likelihood (ranging from 0 to 1) that an individual lives in a household of at least three people (adults or children) and that the household income is £10,000 or below.

Other income bands and household sizes were tested in 2015, but this combination produced the best model in terms of effect and significance.

Own Home Outright, Wales

The likelihood (ranging from 0 to 1) that an individual owns their home outright (i.e. without a mortgage) and that they live in Wales. This interaction was found to be a strong predictor in 2015, and has been retained in the model each year.

Has Savings £10k+

The likelihood (ranging from 0 to 1) that an individual has savings with a total value of at least £10,000. All savings products (fixed and variable) are included, but investment products and pension savings are not included.

Investment values and individual product holdings were also tested in 2015, but these introduced multicollinearity into the model and compromised overall fit.

Aged 65-74

The likelihood (ranging from 0 to 1) that an individual is aged between 65 and 74 years old (inclusive).

Other age bands (including broader bands) did not prove significant in any model.

Aged 75+

The likelihood (ranging from 0 to 1) that an individual is aged 75 years or older.

Other age bands (including broader bands) did not prove significant in any model

The source for each variable is given below.

Model Variable	Source of Data
Has Loan for Consolidation	FRS
Private Renting	FRS
Social Renting	FRS
Has 3+ Children	FRS
Social Grade D or E	FRS
Household Income <£10k, Household Size 3+	FRS
Own Home Outright, Wales	FRS, ONS
Has Savings £10k+	FRS
Aged 65-74	FRS
Aged 75+	FRS

FRS = Ocean: Modelled by CACI, based on data from the *Financial Research Survey*, GfK

ONS = Defined boundaries by the UK *Office of National Statistics*

9. Evaluation of Model

9.1. Statistical Significance of Parameters

As demonstrated in 8.2.2 all variables in the model are significant, at the required 95% confidence limit. In fact, the confidence can be increased for the majority of the variables – up to 99% in all variables except “Has Savings £10k+”.

9.2. Hosmer-Lemeshow Test

The Hosmer-Lemeshow Test is a test for goodness-of-fit within a logistic regression model. It is frequently used to evaluate predictive models of this kind by attempting to identify a “lack of fit”.

The test first sorts observations (individual survey respondents) into ten equal-sized groups, based on the modelled probability of each one being over-indebted.

The expected number of over-indebted individuals within each group can be calculated by summing the modelled probabilities. These projections are then compared to the observed values in each group (counts of individuals who actually said they were over-indebted in the research surveys).

These ten pairs of numbers (observed versus modelled) should be close to each other, and they can be statistically tested using a Chi-square test.

Partition	Observed (Survey Data)	Modelled
1	39	38.0
2	65	62.3
3	108	122.0
4	180	179.0
5	227	221.9
6	264	262.4
7	296	305.2
8	393	357.6
9	430	426.7
10	524	551.0

Hosmer and Lemeshow Goodness-of-Fit Test

Chi-Square	8.9586
Degrees of Freedom	8
Pr > ChiSq	0.3458

The test confirmed that there is no lack of fit (a Pr>ChiSq value larger than 0.1), and so it can be concluded that the predicted levels of over-indebtedness within the groups are sufficiently close to observed levels.

9.3. C-Statistic

The c-statistic, or “concordance statistic” is a common test to report on within logistic regression analysis, and is a single measure of the reliability of the predicted levels of over-indebtedness, *at an individual level*. However, because the objective of this model is to provide expected levels of over-indebtedness at a local area level (by summing individual-level probabilities), individual-level predictions are less relevant. The Hosmer-Lemeshow goodness-of-fit test is a more appropriate test for this model.

Each “over-indebted” observation (i.e. survey respondents who said they were over-indebted) is paired with every “not over-indebted” observation. In the modelled data set of 15,966 usable observations, the observed number of over-indebted individuals is 2,526. This generates 33,949,440 (13,440 x 2,526) possible pairings of an over-indebted individual with a not over-indebted individual. In each pairing, the predicted likelihoods of being over-indebted can be compared. If the model provided a reliable prediction, then the likelihood for the over-indebted individual should always be greater than the likelihood for the not over-indebted individual (this is known as “concordance”). And if the model is entirely random, it would be expected for this to only occur in half of the pairings.

Percent Concordant	70.6%	Somers' D	0.418
Percent Discordant	28.8%	Gamma	0.42
Percent Tied	0.5%	Tau-a	0.111
Pairs	3,3949,440	c	0.709

The c-statistic for the over-indebtedness model is 70.9%.

In other words, if an over-indebted (A) and a not over-indebted individual (B) were randomly selected from the survey respondents, the model is likely to give (A) a higher likelihood of being over-indebted than (B). If this was done 100 times, the model would correctly give the over-indebted individual a higher probability on 71 occasions.

A model is considered good if $c > 70\%$ and strong when it is $> 80\%$ (Hosmer & Lemeshow, 2013). This is an acceptable result for the modelling objectives as previously explained.

9.4. Multicollinearity

The variables selected in the model should be statistically independent. In other words there should be no strong correlation between any pairs of variables. This can be tested by creating a correlation matrix of the variables. The score (Pearson’s correlation moment) ranges from -1 to 1. A score of -1 indicates a perfect negative correlation, 1 indicates a perfect positive correlation, and scores close to 0 indicate no correlation at all.

Model Variable		1	2	3	4	5	6	7	8	9	10
1	Has Loan for Consolidation	1.00	0.13	-0.03	0.09	-0.04	-0.02	-0.08	-0.38	-0.35	-0.38
2	Private Renting		1.00	-0.14	0.10	0.03	0.07	-0.09	-0.31	-0.15	-0.16
3	Social Renting			1.00	0.11	0.75	0.39	-0.07	-0.40	-0.06	-0.07
4	Has 3+ Children				1.00	0.09	0.19	-0.04	-0.28	-0.14	-0.11
5	Social Grade D or E					1.00	0.46	-0.04	-0.58	-0.04	-0.05
6	Household Income <£10k, Household Size 3+						1.00	-0.04	-0.39	-0.14	-0.13
7	Own Home Outright, Wales							1.00	0.11	0.10	0.07
8	Has Savings £10k+								1.00	0.46	0.32
9	Aged 65-74									1.00	-0.14
10	Aged 75+										1.00

Some moderate multicollinearity is to be expected in logistic regression models, however the model presents only two incidences greater than 0.5, and worthy of further attention.

The strongest correlation (0.75) is between individuals who rent their home socially and those who are classed in social groups D and E. Although this score suggests reasonably strong correlation, both variables are strongly significant, with strong positive effects.

The second-strongest correlation is the negative relationship between social grade and high savings values, with a correlation of -0.58. This in principle should not affect the model, but more importantly social grade E includes retired people – a demographic very likely to have higher value savings. We must include both variables in a model to ensure that we can distinguish pensioners that have no/few savings, in comparison to those with high values of savings.

9.5. Model Validation

The research data was split into two parts: an 80% “training” sample to be used to create a revised over-indebtedness model, and a 20% “validation” sample that can be applied to the model to independently verify the accuracy and suitability. The data was sampled using a 1-in-5 selection method ensuring demographic and regional representativeness.

Of 3,991 records in the validation sample, 659 were classified as over-indebted (16.5%). The 2017 model predicts an over-indebtedness figure of 629 individuals. This represents a 4.5% under-prediction, or a 0.7 percentage point error in the over-indebtedness rate – well within acceptable tolerances.

Model Variable	Count	%
Usable Records	3,991	
Modelled Over-indebtedness:	629	15.8%
Observed Over-indebtedness:	659	16.5%
Modelled Error (%)		-4.5%
Modelled Error (pp)		-0.7%
Concordance Statistic (c)		68.8%

The concordance statistic (an estimate of individual-level predictions) for the validation sample was 69%. This is marginally acceptable, and not dissimilar from the 71% seen within the training sample.

Furthermore, when additional age and region variables were added to the over-indebtedness model, no improvements to model accuracy were achieved. At best the model performances were as good as that of the selected model. The results of these additional tests can be obtained by contacting either the Money Advice Service or CACI using the details in section 12.

9.6. Stability of Prediction

Over the last three years the over-indebtedness model has performed well, and has produced consistently stable estimates. Over-indebtedness forecasts have remained between 15.4% and 16.1% in this time period.

Furthermore, the majority of predictor variables derived in 2015 remains in the model, and at a high level of statistical significance. Those variables that have been removed have not had a negative impact on either model accuracy, or a large effect on any aspect of the over-indebtedness estimates across any part of the country or its demographics.

As part of the 2017 update, the 2016 model was also tested, alongside some full model alternatives. These models included dummy variables for all age bands and all regions, but did not produce such strong levels of prediction, and nor were all of the variables statistically significant. This suggests a great degree of stability within the model and its constituent variables and further validation of the robustness of the model, and the insensitivity of the prediction to additional variables.

10. Resulting Output

10.1. Applying the Model to the UK Population

The over-indebtedness model was built on 19,957 survey respondents from across the UK. Because the independent variables are all available within CACI's Ocean database, the model could be applied to the 52 million⁸ adults in the UK at an individual level. For the purpose of these counts, "adults" are defined to be individuals aged 18 years or older.

Over-indebtedness scores (the likelihood of being over-indebted) were first built at individual level from the count of Ocean adults. These were then applied to the latest 2017 population estimates (at unit postcode level) to produce a definitive projection of over-indebted individuals for each postcode. Where required, over-indebted counts were adjusted to the latest and most accurate population estimates at a unit postcode level, before being aggregated to areas.

CACI are the sole data provider to the Joint Industry Committee for Population Standards (JICPOPS), which ensures comparable population statistics across the advertising and media industry. These population estimates are very much seen as the standard across a wide range of industries, and are considered the most robust current year estimates available. The over-indebtedness scores have been applied to these figures to ensure the resultant area statistics are as up-to-date and accurate as possible.

10.2. UK Over-Indebtedness Figure

The headline figure for the number of over-indebted adults in the United Kingdom is 8.27 million. This equates to 15.9% of the adult population who are regularly missing monthly payments or finding meeting commitments a heavy burden. The over-indebtedness figure has shown a small increase since 2016, when the model produced a figure of 15.4%.

	Adults	Over-Indebted %	Over-Indebted Adults
United Kingdom	52,108,329	15.9%	8,268,055

10.3. Over-Indebtedness by UK Region

There is a range of over-indebtedness across the twelve regions and countries of the United Kingdom, with average levels of over-indebtedness ranging from 13.3% in the South East to 17.7% in the North East.

Region	Adults	Over-Indebted %	Over-Indebted Adults
North East	2,115,868	17.7%	373,963
Wales	2,491,338	17.7%	439,779
London	6,919,069	17.2%	1,190,557
North West	5,699,848	16.9%	964,373
Yorkshire & The Humber	4,290,217	16.9%	724,995
West Midlands	4,547,384	16.6%	756,390
East Midlands	3,760,559	16.2%	608,956
Scotland	4,376,847	15.9%	696,859
Northern Ireland	1,437,885	15.8%	226,516
South West	4,457,700	14.4%	641,607
East of England	4,862,102	14.3%	693,851
South East	7,149,512	13.3%	950,209

⁸ CACI's Ocean database of individuals contains 48million adults, and these were used within the analysis. The model output was then applied to the full UK adult population of 52million.

10.4. Other Geographies

Over-indebtedness has been aggregated from postcode level into three small-area geographies⁹.

Lower Tier Local Authority

Non-metropolitan districts, metropolitan boroughs, London boroughs and unitary authorities of England. Includes all districts of Scotland (32), Wales (22) and Northern Ireland (11).

There are 391 lower tier local authorities across the United Kingdom.

The 10 most over-indebted and 10 least over-indebted local authorities (lower level) are:

Rank	Local Authority	Over-Indebted %	Rank	Local Authority	Over-Indebted %
1	Newham	22.7%	382	Wokingham	10.7%
2	Tower Hamlets	22.7%	383	Hart	10.7%
3	Sandwell	22.1%	384	Epsom and Ewell	10.5%
4	Nottingham	21.9%	385	Waverley	10.5%
5	Barking and Dagenham	21.8%	386	Wealden	10.4%
6	Blaenau Gwent	21.6%	387	South Bucks	10.1%
7	Kingston upon Hull, City of	21.5%	388	Elmbridge	10.0%
8	Manchester	21.5%	389	Chiltern	9.9%
9	Leicester	21.0%	390	Mole Valley	9.9%
10	Hackney	20.9%	391	East Dorset	9.7%

Upper Tier Local Authority

Non-metropolitan counties, metropolitan counties, Inner & Outer London and unitary authorities.

Includes all districts of Scotland (32), Wales (22) and Northern Ireland (11).

There are 156 upper tier local authorities across the United Kingdom.

The 10 most over-indebted and 10 least over-indebted local authorities (upper level) are:

Rank	Local Authority	Over-Indebted %	Rank	Local Authority	Over-Indebted %
1	Nottingham	21.9%	147	Rutland	12.4%
2	Blaenau Gwent	21.6%	148	West Sussex	12.4%
3	Kingston upon Hull, City of	21.5%	149	East Sussex	12.3%
4	Leicester	21.0%	150	Buckinghamshire	12.2%
5	Merthyr Tydfil	20.7%	151	East Renfrewshire	12.1%
6	Stoke-on-Trent	20.0%	152	East Dunbartonshire	12.1%
7	Blackburn with Darwen	19.8%	153	Dorset	12.0%
8	Middlesbrough	19.7%	154	Surrey	11.3%
9	Rhondda Cynon Taf	19.5%	155	Windsor and Maidenhead	11.1%
10	Peterborough	19.3%	156	Wokingham	10.7%

⁹ CACI's Ocean database of individuals contains 48million adults, and these were used within the analysis. The model output was then applied to the full UK adult population of 52million, which can be allocated to various geographic definitions.

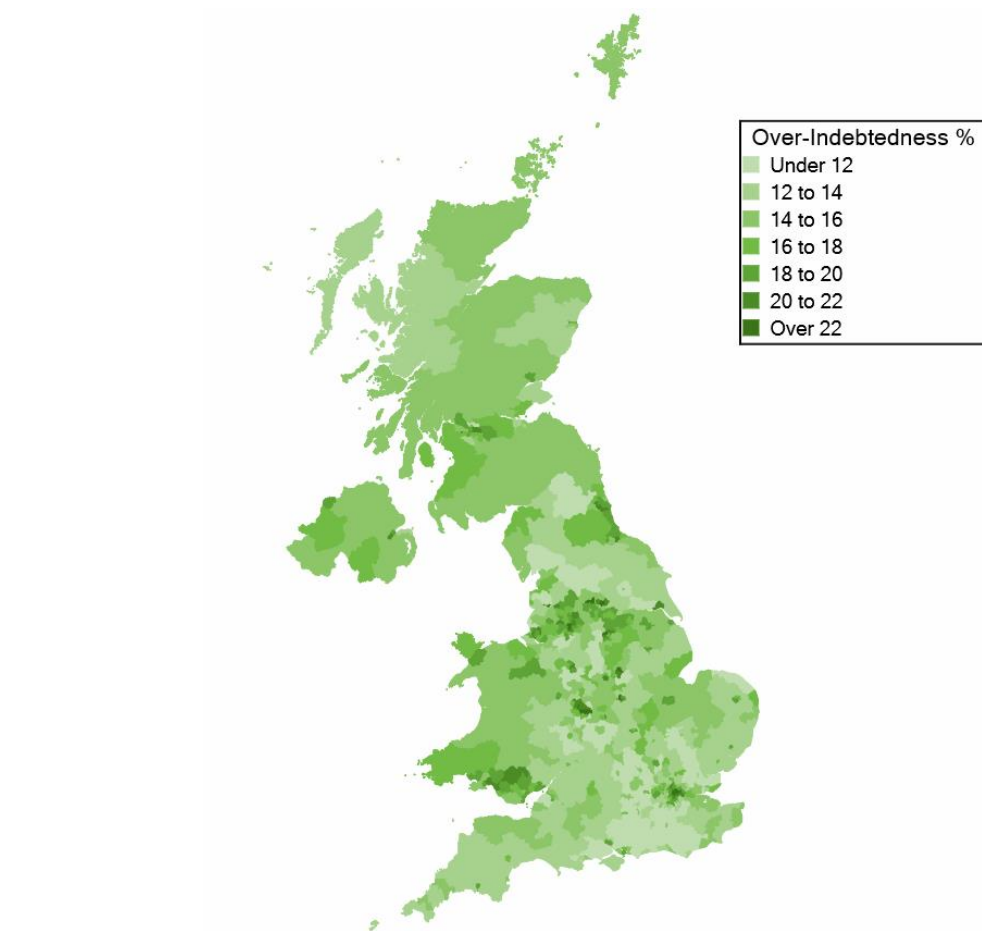
Parliamentary Constituency

2013 Westminster parliamentary constituency boundaries, derived for the 2015 general election. There are 650 parliamentary constituencies across the United Kingdom.

The 10 most over-indebted and 10 least over-indebted constituencies are:

Rank	Constituency	Over-Indebted %	Rank	Constituency	Over-Indebted %
1	Birmingham, Ladywood	24.8%	641	North East Hampshire	10.4%
2	Leeds Central	23.8%	642	Beaconsfield	10.2%
3	Manchester Central	23.2%	643	Henley	10.1%
4	Nottingham North	23.2%	644	Esher and Walton	10.1%
5	West Ham	23.1%	645	Epsom and Ewell	10.1%
6	West Bromwich West	22.9%	646	Chesham and Amersham	9.9%
7	Poplar and Limehouse	22.8%	647	Arundel and South Downs	9.9%
8	Barking	22.7%	648	New Forest West	9.9%
9	Nottingham East	22.6%	649	Christchurch	9.8%
10	Bethnal Green and Bow	22.5%	650	Mole Valley	9.8%

Parliamentary constituencies, mapped by over-indebtedness



An interactive map, which depicts over-indebtedness at different geographies including local authority and parliamentary constituency, can be found by visiting <https://www.moneyadvice.service.org.uk/en/corporate/a-picture-of-over-indebtedness-in-the-uk>

11. References

- Alan Agresti (2012). *Categorical Data Analysis*. Hoboken: John Wiley and Sons. ISBN 978-0-470-46363-5
- Delwiche Lora D and Susan J. Slaughter (2003). *The Little SAS Book: A Primer, Third Edition*, Cary, NC: SAS Institute Inc. ISBN 978-1-59047-333-7
- FCA (2017). *Understanding the Financial Lives of the UK*. Available at: <https://www.fca.org.uk/publications/research/understanding-financial-lives-uk-adults>
- Hosmer, David W.; Lemeshow, Stanley (2013). *Applied Logistic Regression*. New York: Wiley. ISBN 978-0-470-58247-3
- Money Advice Service (2013). *Indebted lives: the complexities of life in debt*. Available at: <https://www.moneyadviceservice.org.uk/en/corporate/indebted-lives-the-complexities-of-life-in-debt>
- Money Advice Service and CACI (2015). *A picture of over indebtedness (technical report)*. Available at: <https://www.moneyadviceservice.org.uk/en/corporate/a-picture-of-over-indebtedness-2016>
- Peduzzi, P; Concato, J; Kemper, E; Holford, TR; Feinstein, AR (1996). *A simulation study of the number of events per variable in logistic regression analysis*. *Journal of Clinical Epidemiology* 49 (12)
- SAS Institute Inc. *SAS/Stat User's Guide (2015)*. Available at: <http://support.sas.com/documentation/93/index.html>

12. Contacts

CACI

Technical Author Jamie Morawiec 020 7605 6035 jmorawiec@caci.co.uk

Account Manager Henry Steenstra 020 7605 6201 hsteenstra@caci.co.uk

The Money Advice Service

Service Design Manager Jair Munoz-Bugarin 020 7943 0524 jair.munoz-bugarin@moneyadviceservice.org.uk