

Income estimates for small areas in England and Wales, technical report: financial year ending 2023

Methods used to produce small area income estimates in England and Wales, including information on model methods, data sources, and estimate quality.

Contact:
Admin and Modelled Income
Estimates (AMIE) team
AMIE@ons.gov.uk
+44 1633 988062

Release date:
11 March 2026

Next release:
To be announced

Table of contents

1. [Overview of income estimates](#)
2. [Methodology](#)
3. [Modelling for income and datasets](#)
4. [Developing the models](#)
5. [Quality of the estimates](#)
6. [Comparing data and metadata profiles for financial years ending 2020 and 2023](#)
7. [Guidance on the use of the estimates](#)
8. [Cite this methodology](#)

1 . Overview of income estimates

Importance of income statistics

There is a need for high-quality income statistics at the smallest possible geographical level. Interest in this stems from a variety of sources:

- central government departments
- local authorities
- academics
- commercial organisations
- independent researchers

These data are essential for the identification of deprived and disadvantaged communities, to support work on social exclusion and inequalities, evaluation research, provision of information for practitioners, and the profiling of geographical areas.

Requirement for income data

Questions on income have never been included in the UK census. Alternative methods for obtaining data on income at the small area level were identified and implemented. One of the options identified was the use of small area estimation methodologies to produce small area income estimates.

Use of Middle-layer Super Output Areas

This report is a technical guide to support the financial year ending (FYE) 2023 (April 2022 to March 2023) set of Middle-layer Super Output Area (MSOA)-level income estimates for England and Wales. Super Output Areas (SOAs) are a geographic hierarchy designed to improve the reporting of small area statistics in England and Wales. A range of areas have been developed that are of consistent size and are subject to minimal boundary changes. These areas are built from groups of Output Areas (OAs) used for the census.

The SOA layers form a geographical hierarchy based on aggregations of OAs; these add firstly to form Lower-layer Super Output Areas (LSOA) then to larger areas. MSOAs comprise between 2,000 and 6,000 households and have a usually resident population between 5,000 and 15,000 persons. They are built from groups of LSOAs and are constrained by the local authority boundaries used for 2021 Census outputs. In the 2021 Census, there were 7,264 MSOAs in England and Wales.

Comparability with other sources

These model-based estimates of average household income in MSOAs are not calculated in the same way as the national and regional household income estimates published separately in our accredited official statistics on [Average household income, UK](#) and [Effects of Taxes and Benefits on UK Household Income](#).

The definitions of income and data sources used for these statistics are different. The Small Area Income estimates are produced using the Department for Work and Pensions' (DWP) Family Resources Survey (FRS) together with auxiliary data from other sources. In contrast, other Office for National Statistics (ONS) income publications come from the Household Finance Survey (HFS). It is not possible, therefore, to aggregate the estimates up to match the regional and national estimates.

The method for producing small area estimates combines survey data with auxiliary data that are correlated with the target variable. The approach is to create a model that relates the survey variable of interest (for example, income) to these auxiliary variables (covariates).

The FRS survey sample is too small to provide reliable direct estimates for small areas or domains, but synthetic estimates can be made based upon the model parameters and values for the covariate data, which are available for all the small areas. These estimates and [confidence intervals](#) are published as [accredited official statistics](#).

More information on the different measures of income can be found in our [Income and Earnings Statistics Guide](#).

Data quality and methods

The report contains details of the methods and processes used, and of the assessment of the quality of the models, and the resulting income estimates. Several diagnostic checks are used to assess quality, which show that in general the models are well-specified, and the modelling assumptions are satisfied. Such checks are described in [Section 5: Quality of the estimates](#) and include an assessment of residuals compared with model estimates, estimates of precision, stability, distinguishability, and a Wald-based comparison of the direct survey and modelled MSOA estimates. This provides assurance of the accuracy of the estimates and the confidence intervals.

Also included in this report is a comparison of the model (and tabulated covariate data) used to derive the income estimates for financial year ending 2023 with that used for financial year ending 2020 and guidance on the use of the estimates.

The methodology uses several administrative and survey sources, with the DWP' FRS serving as a main component.

Further technical information can be found in our previously published methodology, [Income estimates for small areas technical report: financial year ending 2016](#).

2 . Methodology

Synthetic estimation produces estimates for domains where survey data are insufficient, by borrowing strength from other data sources. The other data sources (known as auxiliary data or covariates) are available on an area basis and for all areas in the target population. At the level of these small areas, survey sample sources are not generally available, so the covariate data are usually from some administrative system or a previous census.

The small area estimate is based on the area-level relationship between the survey variables and auxiliary variables. This relationship can be fitted by regressing individual survey responses (for example, household income) on area-level values of the covariates (for example, proportion of the Middle-layer Super Output Area (MSOA) population claiming Income Support). The fitted model describes the relationship between the area-level summary (mean) values of the target survey variable and the covariates.

While the model has been constructed only on responses from sampled areas, the relationships identified by the model are assumed to apply nationally. So, as administrative and census covariates are known for all areas, not just those sampled, the fitted model can be used to obtain estimates and [confidence intervals](#) for all areas. This is the basis of the synthetic estimation that we have used in the development of small area estimation. An assessment of the quality is made using several diagnostics.

For more technical details of the methodology, please see our previously published methodology, [Income estimates for small areas technical report: financial year ending 2016](#).

3 . Modelling for income and datasets

Survey data

The survey data were obtained from the [Family Resources Survey: financial year 2022 to 2023](#).

While other surveys contain some income questions, the FRS is the only large survey that collects a comprehensive set of income components across the full target population and was chosen on this basis. The Labour Force Survey measures income for employees only and does not include the self-employed, benefits income, or housing costs, and for this reason was not used.

The FRS allows four survey variables to be modelled, and the average is used as the summary variable. The estimates produced are values of average Middle-layer Super Output Area (MSOA) income for the following four income types:

- total annual household income (unequalised)
- disposable (net) annual household income (unequalised)
- disposable (net) annual household income before housing costs (equalised)
- disposable (net) annual household income after housing costs (equalised)

Note that the definition of "equalisation" is provided later in this section and considers the household size and composition. It acknowledges that, for example, two people do not need double the income of one person to have the same living standards.

Total annual household income (unequalised)

This is the sum of the gross income of every member of the household plus any income from benefits, that is, wages and salaries, self-employment, pensions, investments, and social benefits.

Disposable (net) annual household income (unequivalised)

This is the sum of the disposable (net) income of every member of the household, that is, all income (from wages and salaries, self-employment, pensions, investments, benefits) minus Income Tax, National Insurance, Council Tax, maintenance or child payments deducted through pay, and contributions to occupational pensions.

Disposable (net) annual household income before housing costs (equivalised)

This is the same as disposable (net) annual household income unequivalised but is then subject to equivalisation.

Disposable (net) annual household income after housing costs (equivalised)

This uses the same elements as disposable (net) annual household income, but also deducts housing costs, such as rent, water rates, mortgage interest payments, structural insurance premiums, ground rent, and service charges prior to the equivalisation scale.

Equivalisation

Equivalisation is the process of accounting for the fact that households with many members are likely to need a higher income to achieve the same standard of living as households with fewer members.

Equivalisation considers the number of people living in the household and their ages, acknowledging that while a household with two people in it will need more money to sustain the same living standards as one with a single person, the two-person household is unlikely to need double the income. These estimates use the modified Organisation for Economic Co-operation and Development (OECD) equivalisation scale for before housing costs and companion scale for after housing costs.

For more details on these income definitions and the equivalisation scale, see our previously published methodology, [Income estimates for small areas technical report: financial year ending 2016](#).

Income estimates are published in terms of annual income rather than weekly income to aid interpretation. However, the estimates are modelled using weekly income data as per previous outputs. The final weekly estimates are expressed as annual income using a factor of 365.25 divided by seven.

Sample size

The FRS uses a stratified clustered probability sample drawn from the Royal Mail's Postcode Address File (PAF). The survey selects 3,407 UK postcode sectors with a probability of selection that is proportional to size. Each sector is known as a Primary Sampling Unit (PSU). Within each PSU, a sample of addresses is selected.

In the financial year ending 2023, 28 addresses per PSU were selected. More information on the FRS methodology is contained within the [Family Resources Survey \(FRS\): background information and methodology](#).

The FRS aims to interview all adults in a selected household. A household is defined as fully co-operating when it meets this requirement. In addition, to count as fully co-operating, there must be less than 13 "don't know" or "refusal" answers to monetary amount questions in the benefit unit schedule (for example, excluding the assets section of the questionnaire).

In the financial year ending 2023, the achieved sample size (for the UK) was 25,056 households. This reflects the introduction of the planned sample boost. As a result of the sample boost, the precision of income estimates is expected to be higher. The sample is expected to return to normal levels in subsequent years. More information on the FRS methodology is contained within the FRS background information and methodology.

Survey data file

The requirement for this release is to produce MSOA-level estimates of average household income (four types) for England and Wales. As the estimates cover only England and Wales, the survey data file used contained 21,110 households from 2,698 postcode sectors in financial year ending 2023. This contained cases in 4,344 different MSOAs out of a total of 7,264.

The number of cases per MSOA in the achieved FRS sample varies widely particularly because MSOAs cut across the postcode sectors' primary sampling unit. For example, some MSOAs recorded only one response whereas others had as many as 37 (the maximum number of sampled households).

Consistent with the analyses for previous publications, for each different income type, a minority of records (24 of 21,110 for total annual household income) were found with values of income less than or equal to £1. These were removed from the sample dataset.

Additional records with extremely high total income values were removed as they would have had an unduly large influence on the model. These households either had a total weekly household income that equated to over £20,000 per week, or a total weekly household income over £15,000, and were the only household sampled in a MSOA.

For the disposable (net) weekly (unequalised and equalised) income, records were removed where the disposable (net) income was greater than the total income. The disposable (net) equalised weekly income excludes households containing a married adult whose spouse is temporarily absent. This is because disposable (net) weekly income is based on the Households Below Average Income (HBAI) derived income measures, which are produced using the underlying Family Resources Survey (FRS) data. This is a record-level dataset maintained by the Department for Work and Pensions.

Definitions from Family Resources Survey data

Although all the survey data used in the modelling process are obtained from the Family Resources Survey (FRS), three of these income types are defined by a different study that is based on FRS data. Disposable (net) weekly household income, unequalised and equalised both before and after housing costs, is defined and calculated in the [HBAI documentation](#).

Although all four types of income for a particular household will be calculated using the same FRS data, the HBAI methodology makes some changes to the original dataset. The HBAI dataset is a cut-down version of the FRS data since the HBAI excludes households containing a married adult whose spouse is temporarily absent. An adjustment is also made to sample cases at the top of the income distribution to correct for volatility in the highest income captured in the survey.

For more detail on these adjustments and the reasons for them, see the HBAI documentation. Note that because of the differences in the HBAI and FRS methodology, the two sets of data have different grossing factors.

The data used are as close to the reference period of the target income estimates as possible (that is, for financial year ending 2023). This is shown in [Section 6: Comparing data and metadata profiles for financial years ending 2020 and 2023](#). Administrative data are collected primarily for government administrative processes and may change over time.

Covariate datasets

The methodology requires covariates data to be available at a geographic level compatible with MSOAs. A range of data sources, aligned closely with FYE 2023, were used in the modelling process presenting variables that may be related to household income. They are:

- 2021 Census: Wide range of MSOA-level variables for each FRS respondent; examples include the proportion of adults involved in managerial and professional work, and the proportion of households who are defined as deprived in terms of health dimension
- Department for Work and Pensions (DWP): Benefit claimant counts, August 2022 available via Stat-Xplore
- Valuation Office Agency (VOA): Council Tax: stock of properties, 2022, provided as counts
- Office for National Statistics (ONS): House Price Statistics for Small Areas, Quarter 1 2023, median, mean, and lower quartile prices for completed sales
- Department for Energy Security and Net Zero (DESNZ): Middle Super Output Areas gas and electricity consumption data, 2022
- HM Revenue and Customs (HMRC): Pay as You Earn data, tax year ending 2023
- HM Revenue and Customs (HMRC): Child Benefit data, tax year ending 2023
- regional or country identification variable

Department for Work and Pensions data

The Department for Work and Pensions (DWP) data were provided as counts. However, it was more appropriate to include proportions or prevalence rates in the modelling process. MSOA population data from mid-2022 were used as denominators to derive these proportions.

Valuation Office Agency Council Tax bandings

The Valuation Office Agency (VOA) assigns each residential property in England to one of eight Council Tax bands, depending on its value on 1 April 1991. In Wales, each property is assigned to one of nine Council Tax bands depending on its value on 1 April 2003. The Council Tax data used here were provided as counts for each band for each MSOA. These counts were transformed into proportions.

The Council Tax bands for England and Wales are not consistent, therefore separate covariates are defined for England and Wales.

Regional or country identification variable

England is split into nine International Territorial Level 1 Regions. Binary variables were created for each region and Wales, taking the value one if the MSOA belonged to that region and country, and zero otherwise. The region and country variables included in modelling income were:

- North East
- North West
- Yorkshire and The Humber
- East Midlands
- West Midlands
- East of England
- South East
- South West
- Wales

Note that London was selected as the base case and therefore not specified separately in the modelling procedure.

Data preparation

Before any modelling could proceed, substantial effort had to be channelled into gathering the necessary source data, principally; survey response data and covariates data. The survey dataset comprises the survey response variables of interest, and weekly household income (matched to postcodes and MSOA codes) for the estimation area. The covariate dataset comprises MSOA covariates along with the corresponding MSOA identifiers. These two datasets are matched by reference to the MSOA codes.

While previous small area income estimates releases used the 2011 MSOA boundaries, this release uses the 2021 MSOA boundaries. In 2011, 7,201 MSOA units existed in England and Wales, and this increased to 7,264 in 2021.

Council Tax bandings and benefit claimant counts were published on 2011 MSOA boundaries. To ensure consistency, the 2011 MSOAs were reconstructed to match 2021 MSOAs. This was carried out by weighting according to the numbers of postcodes contained within each MSOA. Such transitions only affected a minority of MSOAs whose borders had moved between the censuses.

The resulting matched dataset, containing the survey variable along with associated covariates and MSOA and Postcode Sector (the latter being the FRS Primary Sampling unit) identifiers, becomes the analysis dataset. The analysis dataset is required for the modelling, and the full covariate dataset (including out-of-sample MSOAs) is required to produce the final estimates once the modelling has been performed.

As with the modelling for previous publications, where missing values existed for any of the covariates, the England and Wales mean of the variable in question was used to impute the missing value.

4 . Developing the models

Linear mixed-effects (multi-level) models were developed for England and Wales to account for the fact that individual households are clustered within specific Middle-layer Super Output Areas (MSOAs). By incorporating area-level random effects, these models recognise that households in the same MSOA often share similar characteristics. The models use "household weekly income" as the response variable, relating it to local area-level covariates to produce estimates for each small area.

The developed models were fitted as multi-level models and can be used to produce estimates of the target variable at the small area level. These models can be used to produce MSOA-level estimates of average weekly household income and calculate [confidence intervals](#) for the estimates.

For all four types of income, the response variable "weekly household income" was not normally distributed but positively skewed (the largest values differ from the mean more than the smaller values do). By using the natural logarithm (ln) of the appropriate type of income as the response variable, this skewness was reduced, and it is assumed for the analysis that the transformed variable follows a normal distribution.

The models were fitted using the statistical software SAS, with postcode sectors at the higher level and households at the lower level. Region and country indicator terms are forced into the model (whether [statistically significant](#) or not) and then the method of stepwise forward selection is used to identify the statistically significant covariates to be included in the models from the set of covariates.

All the appropriate covariates (those expressed as percentages or proportions) were transformed onto the logit scale, and both the transformed and original covariates were considered for inclusion in the models. The covariates were centred by subtracting the corresponding means for England and Wales. Centring the covariates enables easier interpretation of the model parameters, for example, the intercept now represents the weighted average of the response variable (after the ln transformation) over all areas.

Initially, statistically significant (at the 5% level) covariates were selected using a stepwise method for inclusion in the models. Then with these statistically significant covariates, interaction terms were created, tested for significance, and where appropriate, included in the models. Note that covariates were sometimes included in the model even though they did not maintain significance at the 5% level once the interactions terms were included, since they were included in an interaction term, which was statistically significant.

After modelling, adjustments were made to the modelled estimates to ensure they were consistent with the direct survey estimates at regional level for England and country level for Wales (this is known as "benchmarking"). The Family Resources Survey (FRS) data are used to calculate direct estimates of income at these higher geographical levels (estimates at this level are considered robust). The model-based MSOA estimates of income were aggregated to region and country level, and comparisons made between the two sets of estimates. The ratio of direct survey estimate to aggregated model estimate at the region and country level was used to scale all modelled MSOA-level estimates and their confidence intervals.

More detail on this benchmarking approach and aspects of the modelling methodology are given in our previously published [Income estimates for small areas technical report: financial year ending 2016](#).

The significance of each covariate in the models is assessed using the t-ratio, calculated as the parameter estimate divided by its [standard error](#). A larger absolute t-ratio indicates that the covariate has a stronger, more statistically significant relationship with the household income type being modelled.

The subsequent sections describe the models developed for the four income types for England and Wales.

Table 1: Key to covariates included in the model for total weekly household income, unequivalised

Covariate Name	Label	Source	t-ratio
northeast	Respondent is in North East	Country/regional indicators	-3.53
northwst	Respondent is in North West	Country/regional indicators	-4.50
york	Respondent is in Yorks / Humber	Country/regional indicators	-4.57
eastmid	Respondent is in East Midlands	Country/regional indicators	-3.34
westmid	Respondent is in West Midlands	Country/regional indicators	-4.89
east	Respondent is in East of England	Country/regional indicators	-2.57
wales	Respondent is in Wales	Country/regional indicators	-4.36
southeast	Respondent is in South East	Country/regional indicators	-3.44
southwst	Respondent is in South West	Country/regional indicators	-2.65
ewPAYEg2md	Standardised PAYE - Males 60 to 64 - Median	Admin	3.50
ewlnPAYEg3md	Standardised Logit of PAYE - Males 65 and over - Median	Admin	2.55
Incaa2	Standardised Logit of Carers Allowance - Age - 25 to 49	Benefits data	2.97
Indlaca3	Standardised Logit of Disability Living Allowance - Care Award (DLA only) - Highest - over 16	Benefits data	-2.34
Indlad2	Standardised Logit of Disability Living Allowance - Duration - 1 to 2 years - over 16	Benefits data	3.08
Inhba5	Standardised Logit of Housing Benefit - Age - 70 and over	Benefits data	2.93
pgroupc1	Proportion of people aged 16 to 74 whose approximated social grade is C1	Census	2.29
phealth	Proportion of people in households reporting good or fairly good health	Census	2.43
phhtype6	Proportion of households that are a couple with dependent child(ren)	Census	2.91

phhtype7	Proportion of households that are a couple with all child(ren) non -dependent	Census	2.01
phrpman	Proportion of HRP's aged 16 to 74 whose NS-SEC is 'managerial and professional'	Census	7.11
ewPAYEg2md_wales	Interaction of Standardised PAYE - Males 60 to 64 - Median with Respondent is in Wales	Admin and Country/regional indicators	-1.93

Source: Office for National Statistics

Notes

1. SocialGradeC1 refersto individuals in supervisory, clerical and junior managerial, administrative, and professional occupations.

With no covariates included in the model, the estimated residual area variance was 0.0481, with an associated standard error of 0.0041. Throughout this report, estimates are presented in the form 'estimate (standard error)', following standard statistical reporting conventions. For example, the notation 0.0481 (0.0041) indicates a residual area variance estimate of 0.0481, with 0.0041 representing the uncertainty around that estimate. When the statistically significant covariates were included in the model, the residual area variance decreased to 0.0055 (0.0027). This represents an 88.56% reduction in unexplained arealevel variance compared with the model without covariates.

The most statistically significant covariate in the model is the census covariate "phrpman" (proportion of Household Reference Persons (HRPs) aged 16 to 74 years whose National Statistics Socio-economic Classification (NS-SEC) is "managerial and professional"), which has a t-value of 7.11. The strong positive effect of this covariate aligns with expectations, as the value of "phrpman" increases for an MSOA, the average household income in that MSOA also increases.

The next most statistically significant covariate is the regional indicator "westmid" with a t-value of negative 4.89. Being negative, this shows that MSOAs in the West Midlands are estimated to have lower average household income than those in the London region.

The relationship of a covariate with the average household income may be different if it is also involved in a model interaction. For example, the interaction variable "ewPAYEg2md_wales" was found to be statistically significant. This suggests that the relationship between "ewPAYEg2md" (median Pay As You Earn (PAYE) earnings for males aged 60 to 64) and the average household income is different for MSOAs in Wales as compared with the rest of England and Wales. As the coefficient for "ewPAYEg2md" is positive while "ewPAYEg2md_wales" is negative, this implies that a unit increase in the median PAYE income for males aged 60 to 64 years has a positive association with the outcome in England, but this effect is noticeably reduced in Wales.

Table 2: Key to covariates included in the model for disposable (net) weekly household income before housing costs, unequivalised

Covariate Name	Label	Source	t-ratio
northeast	Respondent is in North East	Country/regional indicators	-1.98
northwst	Respondent is in North West	Country/regional indicators	-3.75
york	Respondent is in Yorks / Humber	Country/regional indicators	-3.08
eastmid	Respondent is in East Midlands	Country/regional indicators	-1.11
westmid	Respondent is in West Midlands	Country/regional indicators	-3.06
east	Respondent is in East of England	Country/regional indicators	-0.77
wales	Respondent is in Wales	Country/regional indicators	-3.09
southeast	Respondent is in South East	Country/regional indicators	-1.86
southwst	Respondent is in South West	Country/regional indicators	-1.43
Inldad2	Standardised Logit of Disability Living Allowance - Duration - 1 to 2 years - over 16	Benefits data	3.72
Inhba4	Standardised Logit of Housing Benefit - Age - 60 to 69	Benefits data	0.40
ewhbm	Standardised Housing Benefit - Gender - Male - over 16	Benefits data	-4.54
phhdepch	Proportion of households with dependent child(ren)	Census	2.53
phrpman	Proportion of HRPs aged 16 to 74 whose NS-SEC is 'managerial and professional'	Census	2.96
Inphealth	Logit of Proportion of people in households reporting good or fairly good health	Census	3.60
ewPAYEg1md	Standardised PAYE - Males 16 to 59 - Median	Admin	-2.16
ewInPAYEg4mn	Standardised Logit of PAYE - Females 16 to 59 - Mean	Admin	3.17
ewInPAYEg3md	Standardised Logit of PAYE - Males 65 and over - Median	Admin	4.43
ewInPAYEg3md_eastmid	Interaction of Standardised Logit of PAYE - Males 65 and over - Median with Respondent is in East Midlands	Admin and Country/regional indicators	2.60

phhdepch_Inhba4	Interaction of Proportion of households with dependent child(ren) with Standardised Logit of Housing Benefit - Age - 60 to 69	Census and Benefits data	-2.64
Inphealth_Inhba4	Interaction of Logit of Proportion of people in households reporting good or fairly good health with Standardised Logit of Housing Benefit - Age - 60 to 69	Census and Benefits data	2.17

Source: Office for National Statistics

With no covariates included in the model, the estimated residual area variance was 0.0338 (0.0028), compared with 0.0048 (0.0019) when the statistically significant covariates were included in the model; a decrease of 85.87%. Therefore, these covariates together accounted for 85.87% of the total between-area variance.

The most statistically significant covariate in the model is the benefits covariate "ewhbm" (the male population aged over 16 years in receipt of Housing Benefit), which has a t-value of negative 4.54. This covariate has a negative coefficient; as the proportion of males receiving Housing Benefit increases, the average household income for that MSOA decreases.

The standardised logit of the median PAYE earnings for males aged 65 years and over, "ewlnPAYEg3md", is the next most statistically significant covariate in the model, with a positive coefficient and a t-value of 4.43. This shows that, as the MSOA median PAYE income across males aged 65 years and over increases, so does the average household income.

Table 3: Key to covariates included in the model for disposable (net) weekly household income before housing costs, equivalised

Covariate Name	Label	Source	t-ratio
northeast	Respondent is in North East	Country/regional indicators	-1.62
northwst	Respondent is in North West	Country/regional indicators	-3.96
york	Respondent is in Yorks / Humber	Country/regional indicators	-3.07
eastmid	Respondent is in East Midlands	Country/regional indicators	-0.95
westmid	Respondent is in West Midlands	Country/regional indicators	-2.82
east	Respondent is in East of England	Country/regional indicators	-1.18
wales	Respondent is in Wales	Country/regional indicators	-3.07
southeast	Respondent is in South East	Country/regional indicators	-1.22
southwst	Respondent is in South West	Country/regional indicators	-1.69
Incaa2	Standardised Logit of Carers Allowance - Age - 25 to 49	Benefits data	3.22
Inspf	Standardised Logit of State Pension - Gender - Female - over 16	Benefits data	0.75
ewhbtot	Standardised Housing Benefit - Total - over 16	Benefits data	-3.92
pcommun	Proportion of people living in communal establishments	Census	2.89
pmanprof	Proportion of people aged 16 to 74 whose NS-SEC is 'managerial and professional'	Census	7.12
Inpecactiv	Logit of Proportion of people aged 16 to 74 who are economically active	Census	3.92
Inphrpmale	Logit of Proportion of household reference persons who are male	Census	-3.80
ewPAYEg3md	Standardised PAYE - Males 65 and over - Median	Admin	4.65
ewPAYEg5tp	Standardised PAYE - Females 60 to 64 - 10th P'ile	Admin	-2.43
ewPAYEg6tp	Standardised PAYE - Females 65 and over - 10th P'ile	Admin	2.55

ewlnPAYEg0tp	Standardised Logit of PAYE - All - 10th P'ile	Admin	-1.99
ewlnPAYEg2mn	Standardised Logit of PAYE - Males 60 to 64 - Mean	Admin	2.44
Inpecactiv_Inspf	Interaction of Standardised Logit of Proportion of people aged 16 to 74 who are economically active with Standardised Logit of State Pension - Gender - Female - over 16	Census and Benefits data	-2.78
ewPAYEg3md_eastmid	Interaction of Standardised PAYE - Males 65 and over - Median with Respondent is in East Midlands	Admin and Country/regional indicators	2.48
ewlnPAYEg0tp_pcommun	Interaction of Standardised Logit of PAYE - All - 10th P'ile with Proportion of people living in communal establishments	Admin and Census	-2.29
Incaa2_westmid	Interaction of Standardised Logit of Carers Allowance - Age - 25 to 49 with Respondent is in West Midlands	Benefits data and Country /regional indicators	-2.99
ewhbtot_westmid	Interaction of Standardised Housing Benefit - Total - over 16 with Respondent is in West Midlands	Benefits data and Country /regional indicators	2.11

Source: Office for National Statistics

With no covariates included in the model, the estimated residual area variance was 0.0253 (0.0021), compared with 0.0013 (0.0013) when the statistically significant covariates were included in the model; a decrease of 95.05%. Therefore, these covariates together accounted for 95.05% of the total between-area variance.

The most statistically significant covariate in the model is the census covariate, "pmanprof". This refers to the proportion of people aged 16 to 74 years whose NS-SEC classification is "managerial and professional", which has a t-value of 7.12. Therefore, as the proportion of people aged 16 to 74 years whose NS-SEC is "managerial and professional" increases, so does income (before housing costs).

The next most statistically significant covariate is "ewPAYEg3md" (standardised PAYE, males aged 65 years and over, median) in the model with a positive coefficient and a t-value of 4.65. This shows that, as the median MSOA of PAYE income for males aged 65 years and over increases, the average household income increases.

Table 4: Key to covariates included in the model for disposable (net) weekly household income after housing costs, equalised

Covariate Name	Label	Source	t-ratio
northeast	Respondent is in North East	Country/regional indicators	2.52
northwst	Respondent is in North West	Country/regional indicators	0.88
york	Respondent is in Yorks / Humber	Country/regional indicators	1.15
eastmid	Respondent is in East Midlands	Country/regional indicators	1.36
westmid	Respondent is in West Midlands	Country/regional indicators	-0.17
east	Respondent is in East of England	Country/regional indicators	1.15
wales	Respondent is in Wales	Country/regional indicators	0.87
southeast	Respondent is in South East	Country/regional indicators	1.13
southwst	Respondent is in South West	Country/regional indicators	0.33
ewpmed	Standardised Median House prices	House Prices	9.21
ewPAYEg4lq	Standardised PAYE - Females 16 to 59 - Lower Quartile	Admin	7.54
ewhbf	Standardised Housing Benefit - Gender - Female - over 16	Benefits data	-4.19
phhdepr_hous	Proportion of households classed as deprived (housing)	Census	-1.14
Inphhshare	Logit of Proportion of household residents living in a shared dwelling	Census	2.84
Inphrpmale	Logit of Proportion of household reference persons who are male	Census	0.24
Inpcommun	Logit of Proportion of people living in communal establishments	Census	2.32
Incaa1	Standardised Logit of Carers Allowance - Age - under 25	Benefits data	-2.07
Inphrpmale_phhdepr_hous	Interaction of Logit of Proportion of household reference persons who are male with Proportion of households classed as deprived (housing)	Census	-3.30

Inphhshare_ewhbf	Interaction of Logit of Proportion of household residents living in a shared dwelling with Standardised Housing Benefit - Gender - Female - over 16	Census and Benefits data	-2.35
phhdepr_hous_ewpmed	Interaction of Proportion of households classed as deprived (housing) with Standardised median House prices	Census and House prices	-2.34

Source: Office for National Statistics

With no covariates included in the model, the estimated residual area variance was 0.0290 (0.0026), compared with 0.0042 (0.0017) when the statistically significant covariates were included in the model; a decrease of 85.53%. Therefore, these covariates together accounted for 85.53% of the total between-area variance.

The most statistically significant covariate in the model is the House Prices covariate "ewpmed", which has a t-value of 9.21. The large positive coefficient for this covariate is expected when modelling income across MSOAs, as areas with higher median house prices typically correspond to higher household incomes. The next most statistically significant covariate is "ewPAYEg4lq", and is the proportion of females aged 16 to 59 years in the lower quartile of PAYE earnings. Its large positive coefficient of 7.54 suggests that, as the standardised MSOA lower quartile PAYE earnings recorded for females aged 16 to 59 years increase, so does the average household income.

Observations

Although some of the covariates may be different between the four equations, the models are generally explaining the same MSOA characteristics, and all four are similar.

1. Across the four models, covariates relating to PAYE amounts are one of the strongest determinants of household income at the small area level. PAYE, as an administrative data source, is particularly useful as it is available for a high proportion of individuals across the country.
2. Other statistically significant positive factors influencing income at MSOA level include the proportion of HRPs aged 16 to 74 years whose NS-SEC is "managerial and professional". This also includes house prices when looking at equivalised disposable (net) weekly household income, after housing costs. Negative determinants include benefits such as Housing Benefit, Care Allowance and Disability Living Allowance.
3. Some regional or country indicators in each model are not statistically significant but are included. This is because benchmarking is carried out on the raw income estimates to benchmark regional and Wales-level average income estimates to those directly derived from the FRS.
4. The final types of covariates included in the models are interaction effects. Approximately half of the interaction terms involve regional or country indicators. This shows that some covariates have different effects in different regions.

Some of the results described may be unexpected. However, it should be remembered that the relationships observed should not be taken in isolation, but alongside the other relationships described by the other covariates present in the model.

5 . Quality of the estimates

Once a model has been selected, an assessment of the quality is made using several diagnostics as described in this section to assess the appropriateness of the models developed. The diagnostic checks employed here are those developed by the Office for National Statistics (ONS) for small area estimation and are published in our [Evaluation of small area estimation methods - an application to unemployment estimates from the UK Labour Force Survey \(LFS\)](#) article, as well as some additional diagnostic checks. The analysis shows that, in general, the models are well specified, and the assumptions are satisfied. This provides confidence in the accuracy of the estimates and the [confidence intervals](#) produced from the models.

The results of the diagnostics for all four income types are summarised in Table 5, with further information about each given after the table. Further information about the diagnostic tests and why they are performed can be found in our previously published [Income estimates for small areas technical report: financial year ending 2016](#) methodology.

Table 5: Diagnostic results for all four income types estimated, England and Wales

Diagnostic Measure		Gross total weekly household income (unequalised)	Disposable (net) weekly household income before housing costs (unequalised)	Disposable (net) weekly household income before housing costs (equalised)	Disposable (net) weekly household income after housing costs (equalised)
Household Level: Residual vs Model Estimates	Constant (SE)	-0.754 (0.188)	-0.880 (0.178)	-0.303 (0.170)	-0.893 (0.183)
	Slope (SE)	0.109 (0.027)	0.134 (0.027)	0.046 (0.026)	0.137 (0.028)
Area Level: Residual vs Model Estimates	Constant (SE)	-0.057 (0.011)	-0.079 (0.013)	-0.012 (0.005)	-0.081 (0.013)
	Slope (SE)	0.008 (0.002)	0.012 (0.002)	0.002 (0.001)	0.012 (0.002)
Household Level: Model vs Sample Estimates	Constant (SE)	-20.63 (36.77)	-59.68 (28.96)	-45.27 (25.85)	-9.13 (26.75)
	Slope (SE)	0.988 (0.034)	1.051 (0.036)	1.064 (0.035)	1.021 (0.039)
Area Level: Model vs Sample Estimates	Constant (SE)	-161.34 (117.69)	158.59 (111.37)	-140.80 (79.76)	83.39 (87.80)
	Slope (SE)	1.240 (0.203)	0.510 (0.269)	1.309 (0.197)	0.752 (0.246)
	Quadratic term (SE)	-0.0001 (0.0001)	0.0003 (0.0002)	-0.0002 (0.0001)	0.0002 (0.0002)
Coverage	%	99.98	100	100	99.98
Wald	p-value	1	1	1	1
Stability Analysis	RRMSE	0.049	0.039	0.038	0.037
Distinguishability	%	28.47	26.05	32.92	23.17

Source: Office for National Statistics

Notes

1. SE refers to the standard error.
2. RRMSE refers to the relative root mean square error.

The following paragraphs describe some of the diagnostic tests performed on the data. If you require more detail, please see our previously published [Income estimates for small areas technical report: financial year ending 2016](#) methodology.

Residual compared with model estimates diagnostic plot

A plot of model estimates against model residuals, both at the household and area level, is a method of checking that the model assumptions are satisfied, and the model accurately describes the population. We are testing model misspecification and non-constant variance of the residuals (heteroscedasticity), and if any pattern remains in the residuals, this suggests model misspecification. For example, a covariate influential to income may have been left out of the model. The assumption of heteroscedasticity was approximately met for both the household level residuals and the area-level residuals.

We require constant variance in the area-level residuals, as this will have an effect on the calculation of the confidence intervals. Model estimates are calculated at the household level (on the natural log (ln) scale) and plotted against the household-level residuals. The standard errors can be used to determine whether the constant and linear terms are significantly different from zero, statistically.

Model compared with sample estimates diagnostic plot

A plot of direct survey estimates (y-axis) against model-based estimates (x-axis) for Middle-layer Super Output Areas (MSOAs), for which there is a sample, is one method of assessing whether the relationship between the target variable and the covariates has been specified properly. For good model-based estimates, the direct estimates will be randomly distributed around the estimates, and the regression line between the two will be very close to the line "y equals x".

If the relationship between the target variable and the covariates has been mis-specified or mis-estimated, then the relationship between the direct and model-based estimates would be expected to be curved or possibly scattered around a different straight line than the "y equals x" line.

An important assumption when using this diagnostic is that the direct estimates are unbiased. The technique for calculating direct survey estimates at an MSOA level is described in our previously published [Income estimates for small areas technical report: financial year ending 2016](#) methodology, along with further detail about this diagnostic test.

The results show that in the quadratic fit, the quadratic term is not [statistically significant](#), and neither is the intercept. In the linear fit, three of the four intercept terms are not significantly different from zero, statistically. The intercept for disposable (net) weekly household income (unequivalised) shows a small but statistically significant deviation. However, in all four models, the estimated slopes are not significantly different from one, statistically. Therefore, the fit is very close to the "y equals x" line. This shows that at least in sampled areas, the modelled estimates do show either none or only very small, occasional signs of bias; ones that are in line with previously published results.

Coverage diagnostic

The purpose of this diagnostic is to examine the validity of the confidence intervals for the model-based estimates. For the MSOAs in the sample, there will be direct survey estimates with associated 95% confidence intervals. This diagnostic measures the overlap between the direct confidence intervals and the corresponding model-based estimate confidence intervals. For example, it measures the percentage of MSOAs for which the model and direct confidence intervals overlap.

However, the overlap between two independent 95% confidence intervals for the same quantity is higher than 95%. Therefore, it is necessary to modify the nominal coverage levels (that is, to narrow the width) of the confidence intervals being compared, to ensure a 95% overlap. Further details of the modification and this test are available in our [Income estimates for small areas technical report: financial year ending 2016](#) methodology. Any statistically significant deviation from a 95% overlap indicates that the model-based confidence intervals are generally too wide or too narrow.

The coverage diagnostic indicates that all four models achieve coverage rates above 95%. This suggests that the modelbased confidence intervals may be somewhat conservative, meaning that the true mean income would fall within the interval for more than 95% of MSOAs. However, the high sample coverage may also affect the variance of the direct survey estimates, which can influence the width of the resulting confidence intervals.

Wald statistic

This diagnostic test assesses the assumptions underlying the model by using a Wald goodness-of-fit statistic. This is used to test whether there is a statistically significant difference between the expected values of the direct estimates and the model-based estimates. Typically, small area-level model-based and direct survey estimates should be approximately correlated and there should be a non-statistically significant p-value associated with the Wald statistic. For all four models, the Wald goodness-of-fit statistic shows no statistically significant difference between the expected values of the direct estimates, and the model-based estimates.

Stability analysis

This diagnostic test analyses the stability of the model's predictive power. The data are split into two datasets similar in size and MSOA representation. The model is fitted to one-half of the data to obtain regression coefficients.

The other half of the data are similarly used in the model to obtain the regression coefficients. These two sets of regression coefficients are then used to obtain two sets of comparable model-based estimates for all MSOAs. This process is repeated 10 times and for each repetition, the difference between the two sets of estimates is measured to evaluate the stability of the model.

A relative root mean square error (RRMSE) is also used as a measure of how close the two sets of model-based estimates are. A small RRMSE indicates that the differences between the two sets of estimates are not statistically significant. The RRMSE stability measures for the four models are all low, ranging from 0.037 to 0.049, therefore indicating a high degree of stability; similar to the models for financial year ending (FYE) 2020 (which ranged from 0.035 to 0.051).

Distinguishability

This diagnostic assesses how well the model differentiates between areas with the lowest and highest estimates. It calculates the percentage of MSOAs in the lowest range, of which the confidence intervals overlap with those in the highest range. Preferably, fewer than 20% of confidence intervals at the lower end should overlap with 20% at the upper end. If the overlap exceeds this threshold, it suggests that confidence intervals are too wide, reducing precision.

In the distinguishability diagnostic assessment for 2023, all four models exceeded the 20% threshold, which also occurred in FYE 2018 and 2020. This is not a reason to reject the models but indicates a need for monitoring. Greater overlap reduces confidence in comparisons, so users should interpret results with caution.

6 . Comparing data and metadata profiles for financial years ending 2020 and 2023

Middle-layer Super Output Area (MSOA)-level model-based estimates of average annual household income have been produced for financial year ending (FYE) 2023 in England and Wales, fulfilling users' requirements for income information at MSOA level. This section outlines the metadata and structural differences between the two years. However, because of these differences, model results from the two years are not comparable.

The only variables that were part of all four models in FYE 2020 and 2023 were region and country, and these were included by design.

Diagnostics

Some plots of household-level, and area-level residuals for all models showed a slight pattern in the data after modelling for FYE 2020 and 2023. However, where there were patterns with the residual plots, the plots of the modelled estimates against the direct estimates showed little or no pattern.

For both years, the coverage diagnostic shows coverage greater than 95% for all four models, indicating that the [confidence intervals](#) of the model-based estimates are possibly conservative. However, this may be caused by overestimating the variances for the direct estimates. For both time periods and all models, the Wald goodness-of-fit statistic shows no statistically significant difference between the expected value of the direct and model-based estimates. Also, the stability analyses for both time periods indicate that the different sets of data produce similar sets of estimates for all four of the models.

The diagnostics for FYE 2020 and 2023 models produce moderately consistent results. This indicates that, in general, the models for England and Wales are well-specified and the assumptions are satisfied. The percentage of variability explained in three out of four of the income models for 2023 exceeded those for 2020.

The only income model which had a smaller percentage was that of equivalised disposable (net) weekly household income after housing costs. For equivalised disposable (net) weekly household income after housing costs, this did not reach the high level seen in FYE 2018 and 2020 but was still strong at 85.53%. This shows confidence in the accuracy of the estimates and their confidence intervals produced from the models.

Covariates

The following covariates were used for modelling FYE 2020:

- Census data, 2021
- HM Revenue and Customs (HMRC): Pay as You Earn (PAYE) data, March 2019
- Department for Work and Pensions (DWP): benefit data, August 2019
- Region and country indicators
- Office for National Statistics (ONS): House Price Statistics for Small Areas, year ending March 2020
- Valuation Office Agency (VOA): Council Tax data, March 2019
- Department of Energy and Climate Change (DECC): Energy Consumption data 2019

The following covariate data were used in the model-based estimates of income for FYE 2023:

- Census data, 2021
- HMRC: PAYE data, tax year ending 2023
- HMRC: Child Benefit Data, tax year ending 2023
- DWP: benefit data, August 2022
- Region and country indicators
- ONS: House Price Statistics for Small Areas year ending March 2023
- VOA: Council Tax data, March 2022
- Department for Energy Security and Net Zero (DESNZ): Energy Consumption data 2022

These lists of data sources show that different covariate datasets were available and used at the time of modelling FYE 2020 and 2023 model-based estimates of average income.

Different covariates have been selected in the models for FYE 2020 and 2023. This is both a consequence of the covariate selection process, as well as the availability of different covariate datasets for the two time periods. The covariate selection procedure ensures that only covariates strongly related to income are selected for each model. However, because of the selection of different covariates, this could result in sharp changes in the estimates for particular areas. A difference in the estimates for an MSOA between FYE 2020 and 2023 could partly reflect differences in the covariates selected in the models, rather than the true change in the mean household income for that area.

Geography of estimation

In FYE 2020, small area income estimates used the 2011 MSOA boundaries. The FYE 2023 uses the 2021 MSOA boundaries. In 2011, 7,201 MSOA units existed in England and Wales, and this increased to 7,264 in 2021.

As mentioned in [Section 3: Modelling for income and datasets](#), the 2021 Census data were extracted for each of the published 7,264 MSOAs.

Caution should be applied when interpreting trends, as the methodology is optimised for a given year rather than for estimating change over time. Non-overlapping confidence intervals suggest possible change over time, but differences between point estimates should not be interpreted as precise measures of change. Each estimate reflects the best available data for its year, and is therefore not an optimised measure of temporal change. Different covariates in earlier models may also create apparent changes where none exist.

7 . Guidance on the use of the estimates

The results of the diagnostic checks presented previously show that the models are well specified, and the modelling assumptions generally hold. However, users should be aware of possible limitations of these model-based estimates. The quality of the estimates is strongly dependent upon the quality and relevance of the input data sources (covariates) used and the fit of the model achieved. In most cases, the estimates are produced using the most up-to-date covariate data sources to match the financial year ending (FYE) 2023 survey data. As such, the estimates should be fully consistent with the current profile of the area.

As with any ranking based on estimates, care should be taken when interpreting Middle-layer Super Output Area (MSOA) income rankings. Users should consider the variability of the estimates when using these figures. For example, the [confidence interval](#) around the highest-ranked MSOA suggests that the estimate lies among the group of MSOAs with the highest income levels rather than being the MSOA with the highest average MSOA income. Estimates for two particular MSOAs can be described as significantly different, statistically, if the confidence intervals for the estimates do not overlap.

Although these model-based estimates can be used to rank MSOAs by income, they cannot be used to make any conclusions on the distribution of income over the MSOAs. The estimation procedure will tend to shrink estimates towards the average level of income for the whole population so estimates at each end of the scales tend to be over- or under-estimated.

Estimates can be used to make inferences, such as the average household income for MSOA "A" is greater than the value for MSOA "B" (if the appropriate confidence intervals do not overlap).

The model-based methodology produces MSOA-level estimates of average income. The model does not support the disaggregation of incomes below MSOA level, nor the aggregation to any level (including to local authorities) apart from International Territorial Level 1 (ITL 1) region or nation.

Models have been developed for four different types of income. In some cases, slight inconsistencies (when examining point estimates) may occur between the income types for particular MSOAs. For example, an MSOA may have a larger modelled estimate for disposable (net) weekly household income (unequalised) when compared with total household income (unequalised). Although there may be some inconsistencies, the models selected are the best possible to model the general patterns of income over all MSOAs. This reinforces the need to look at the confidence intervals for the income estimates, not just the point estimate, since the confidence intervals summarise the variability in the estimates caused by the modelling process.

The model-based method has been developed to ensure that the model-based estimates for MSOAs are constrained to direct survey estimates from the Family Resources Survey (FRS) at the region level for England, and the country level for Wales. However, the model-based estimates will not be consistent with FRS estimates of average household income for other geographical levels.

These estimates have been produced on 2021 MSOA boundaries. Users must be aware of this when using the estimates in any application or drawing conclusions from the data. The estimates are also based on FYE 2023 survey data, and so are only valid for this period.

The different models described previously have been independently chosen to give the best point-in-time estimates of household income for the appropriate time period and geography. In particular, the synthetic estimation methodology, by borrowing strength nationally, tends to draw estimates at the low and high ends of the distribution towards the national mean. This is an acceptable drawback for point-in-time estimation as it is more than compensated by the advantages of borrowing strength nationally in increasing estimate precision.

However, it is problematic when the focus is on measuring local area change over time. This is because the small area estimate of change is drawn towards the national mean of change and no longer distinguishes local variability, which in many cases is what is of particular interest. For this reason, the synthetic estimation applied here is not optimised to give the best estimate of local change.

8 . Cite this methodology

Office for National Statistics (ONS), released 11 March 2026, ONS website, methodology, [Income estimates for small areas in England and Wales, technical report: financial year ending 2023](#)