

**Interpolated
Household Income
Estimates by
Percentile
(2018)
Secondary Data
Analysis**

**A Short
Guide for Housing
Affordability Analysis
(October 2020)**

Scottish Government

AIM

This note sets out the methodology used by the Scottish Government (SG) to derive income percentile estimates for 2018 at local authority and Scotland level.

This is secondary analysis which uses the Local Level Household Income Model Estimates (LLHIE) for each of these years as produced by Heriot-Watt University and David Simmonds Consultancy and published by the Scottish Government in October 2020.

Both the percentiles and the LLHIE data on which they are based can be accessed on SG [CHMA website](#). The percentile datasets provide income estimates for every percentile from the 10th to the 90th percentile ranging from lower to higher incomes by Local Authority and for Scotland.

The original LLHIE estimates were specially produced for the purpose of assessing housing affordability across Scotland and, as such, there are limitations attached to what they may reliably be used for. These are explained below. There are several different ways to measure income and alternative income measures may be more suitable for an intended use other than for housing affordability.

DATA USES

The primary purpose of these income estimates is for use by SG CAD to input to its [Housing Need and Demand Assessment \(HNDA\) Tool](#). This uses income to estimate the future amount and tenure of housing needed by comparing income to house prices and rental costs to ascertain levels of housing affordability in an area.

The income estimates are being produced as open data and are freely available to be used for further analysis. However, if you wish to use the income estimates for reasons other than housing affordability, you should be clear about the methodology and limitations associated with the data and you may wish to seek advice from [SG CAD](#).

DESCRIBING THE DATA

The LLHIE estimates on which the percentiles are based relate to gross weekly income i.e. gross earnings from employment, welfare benefits, tax credits, housing benefit, savings and pensions etc.

Estimates are provided for each income percentile from the 10th to 90th percentile. The top and bottom 10 percentiles have been removed in order to remove outliers and because there is more uncertainty around the estimates at the extreme ends of the distribution. As these are estimates only they have been also rounded to the nearest £10.

The data represent unequivalised income. If income is equivalised it means that household income is adjusted to take account of the size of the household in terms of the numbers of adults and children. Unequivalised income does not adjust for size of household, so that the multi-income/multi-adult households are likely to have a higher income.

Unequivalised income was used because the dataset is designed to analyse if a household can afford to buy a property or if they can afford to rent i.e. a joint income household might be in a stronger position to buy than a single income household.

Income percentile are also available on the [CHMA website](#) for 2014, 2015 and 2017, however we urge caution when comparing estimates for different years for small areas. Whilst there may be genuine changes in income, the differences may also reflect refinements to the methodology that were introduced over the course of the first two rounds. Year-on-year changes may also be influenced by 'data noise' in key sources, such as the SHS data, that may be attributed to sampling variance.

The LLHIE estimates are research only and they do not represent National or Official Statistics.

FINDING THE DATA

The income percentiles and LLHIE estimates for 2014, 2015, 2017 and 2018 can be found on the CHMA website at <https://www.gov.scot/publications/centre-for-housing-market-analysis-index/>

THE METHODOLOGY

Interpolation

The method used is called interpolation:

Interpolation is a method of constructing new data points within the range of known data points

What's the issue?

In the raw LLHIE income dataset, the income distribution for each area is comprised of several different income distributions for a number of household types with different income profiles however we do not have enough information about these distributions to retrieve the full information from the raw data. This would be the case if, for example, the information we had were the two arguments that define a log-normal distribution and we knew that the distribution was log-normal, we could retrieve the rest of the information.

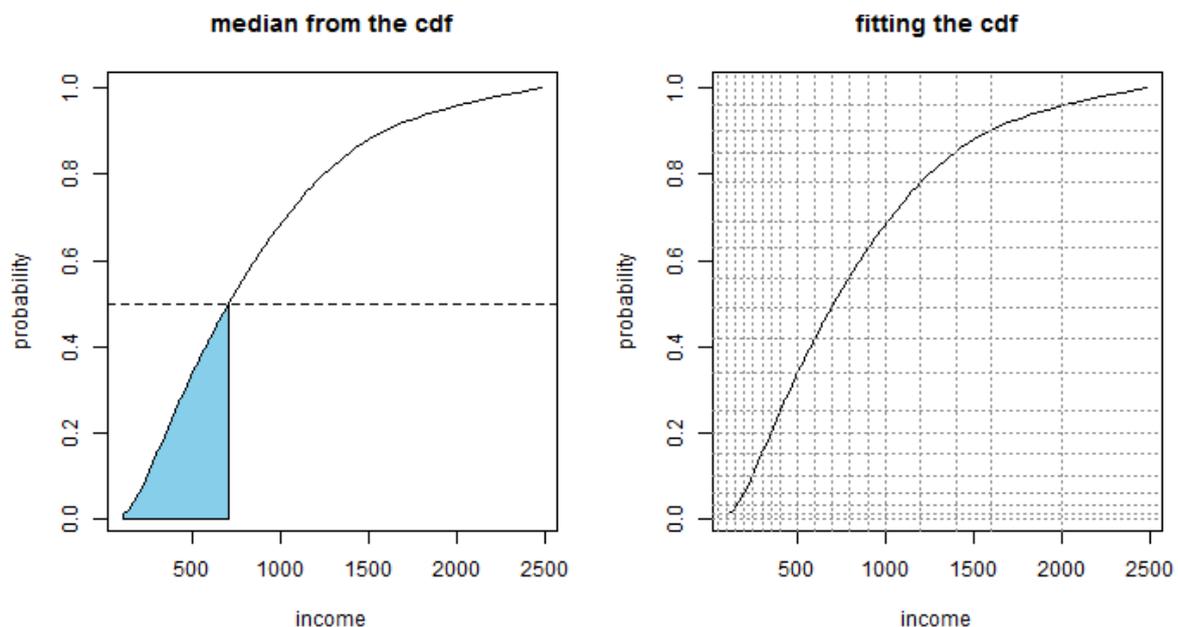
So we don't know enough about the income distribution to estimate the value of income for each percentile across the income distribution. To work this out (as in the box above describes) we need to construct, or interpolate, more data points along the income distribution for each area.

The Cumulative Distribution Function

In order to estimate income percentiles we need to define the Cumulative Distribution Function (CDF) more fully. By interpolating the 18 available points along each CDF, for each Data Zone, we can obtain percentile estimates.

Percentiles give the value below which a given percentage of observations in a group of observations fall. These can be obtained from the inverse of the CDF which is shown below, for this reason the inverse of the CDF is called the quantile function.

By reading from the y axis to the x axis of the CDF below we can obtain values for the appropriate percentile. The example below shows that half of the incomes in the area are below £700. In order to obtain percentiles we need 100 uniformly distributed points along the CDF which we can then use to retrieve the values associated with percentiles 1 to 100.



The CDF can be defined by a well specified distribution (e.g. the log-normal discussed above). However, the sum of log-normal distributions is only approximately log-normal when the tails of the distributions being combined are similar. If these vary substantially then the distribution of the sum is not easy to define. A fitted log-normal was generally found not to be a good approximation of the points on the CDF which are available.

The software used

In order to estimate percentiles, the known points on the CDF were interpolated using a cubic spline in the R statistical software. Known points are shown above where the dotted lines intersect the CDF, these are being joined up by a curve. Points at the extremes are extrapolated by the same function. `splinefun()` and `spline()` belong to R's base functions.

TESTING AND QUALITY ASSURANCE

The estimates have been tested and quality assured in a number of ways including:

- they have been compared to the original LLHIE estimates where possible for example the median value in order to ensure that they do not diverge from the base data on which they are built,
- the top and bottom 10 percentiles of the income distribution have been removed, this is due to the uncertainty around these estimates and because it is where approximations of the distribution perform most poorly (both a log-normal approximation of the income distribution, and reliance on extrapolation from the spline).
- they have been rounded to the nearest £10 to account for error in the underlying data and represent that they are estimates only and not actual incomes,

LIMITATIONS

The following points should be noted regarding the interpolated income estimates:

- the LLHIE base date on which these interpolated income estimates are research only and do not represent Official or National Statistics.
- the figures represent modelled income estimates only and not actual income estimates. Like all modelled estimates they are subject to a degree of error in terms of how closely they match known actual incomes.
- the interpolated income estimates have been produced for the purposes of updating the Scottish Government Housing Need and Demand Assessment (HNDA) Tool.
- the interpolated estimates may also inform work on housing affordability more generally across different tenures and different geographic areas of Scotland, and will help to support local authorities and their partners in the production of Local Housing Strategies and other planning documents.

- it is important to note that gross household income estimates are only one measure of income, and should not be considered on their own without consideration of other local level information. Users are strongly encouraged to use other detailed statistics such as the Scottish Index of Multiple Deprivation or the Scottish Census to develop a basket of evidence and statistics to build up a comprehensive picture of people and households in local areas.
- it is also important to note that the gross household income estimates are not intended to be a measure of person-level income, they do not reflect household income adjusted by household size, they do not reflect income levels after tax or after housing costs, they do not provide information on wealth or assets, and they are not intended as a measure of income based deprivation. Not all people in areas of low average gross household incomes will necessarily be deprived or in poverty, and not all households in areas of high average gross household incomes will necessarily contain people with high levels of personal disposable income or wealth.

SUMMARY FINDINGS

Table 1 (Annex A) shows summary results at Local Authority Level. The estimates have been rounded.

In 2018, weekly gross household income estimates (£), at the lower quartile, range from £270 in Inverclyde council to £420 in East Renfrewshire council. At Scotland level they are £330 per week.

In 2018, weekly gross household income estimates (£), median, range from £450 in Inverclyde council to £740 in East Renfrewshire council. At Scotland level they are £550 per week.

In 2018, weekly gross household income estimates (£), at the upper quartile, range from £770 in Inverclyde council to £1,190 in East Renfrewshire. At Scotland level they are £920 per week.

YOUR FEEDBACK

Your feedback on this secondary income analysis is welcome, for example:

- > is there any additional analysis you would find useful?
- > do you have any questions about the dataset?

Please don't hesitate to get in touch using our mailbox chma@gov.scot.

Scottish Government
October 2020

Table 1 - Interpolated Household Income Estimates, 2018, by Income Quartile and Local Authority

Local Authority	Gross Weekly Household Income ^{1,2}		
	Lower Quartile	Median	Upper Quartile
	25th Percentile	50th Percentile	75th Percentile
Aberdeen City	£330	£560	£920
Aberdeenshire	£410	£700	£1,100
Angus	£330	£540	£890
Argyll and Bute	£320	£530	£880
City of Edinburgh	£360	£600	£970
Clackmannanshire	£300	£490	£820
Dumfries and Galloway	£310	£510	£830
Dundee City	£290	£470	£770
East Ayrshire	£320	£520	£860
East Dunbartonshire	£420	£720	£1,160
East Lothian	£370	£620	£990
East Renfrewshire	£420	£740	£1,190
Eilean Siar	£310	£490	£810
Falkirk	£320	£540	£890
Fife	£320	£530	£880
Glasgow City	£300	£480	£810
Highland	£360	£590	£940
Inverclyde	£270	£450	£770
Midlothian	£370	£630	£990
Moray	£350	£580	£940
North Ayrshire	£300	£490	£810
North Lanarkshire	£330	£550	£910
Orkney Islands	£370	£610	£960
Perth and Kinross	£380	£640	£1,010
Renfrewshire	£320	£540	£900
Scottish Borders	£320	£530	£870
Shetland Islands	£320	£530	£860
South Ayrshire	£330	£550	£900
South Lanarkshire	£360	£600	£980
Stirling	£360	£610	£990
West Dunbartonshire	£290	£480	£790
West Lothian	£340	£580	£940
SCOTLAND	£330	£550	£920

Source: Scottish Government (2020).

Notes:

1. This analysis is based on the Scottish Government Local Level Household Income Estimates (LLHIE) 2018.

2. Gross weekly household income including gross earnings from employment, welfare benefits, tax credits, housing benefits. Unequalised income which does not adjust for size of household where there may be multiple incomes.