

# **Interpolated, Household Income Estimates, by Percentile, 2014**

## **A Short Guide for Housing Affordability Analysis (2019)**

**Scottish Government  
Communities Analysis Division**

## **AIM**

This note sets out the methodology used by the Scottish Government (SG) to derive income percentile estimates, for 2014, at local authority and Scotland level.

This is secondary analysis which uses the Local Level Household Income Model Estimates (LLHIE) produced by Heriot-Watt University and David Simmonds Consultancy and published by the Scottish Government in May 2017.

The estimates can be accessed on SG [CHMA website](#). Summary results are set out in table 1 below. However the full dataset provides income estimates for every percentile from the 10<sup>th</sup> to the 90<sup>th</sup> 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. This is 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.

Small area income estimates were previously produced by Bramley and Watkins in (2013) as part of an Improvement Service research project on income and poverty. The tool has used these income estimates and projected them forward. Now that the 2014-based income estimates are available these will be input to the tool to produce future housing estimates.

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

These estimates are for 2014 and they are at local authority and Scotland level.

The modelled income estimates on which they 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 10<sup>th</sup> to 90<sup>th</sup> 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.

## FINDING THE DATA

The full dataset 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.

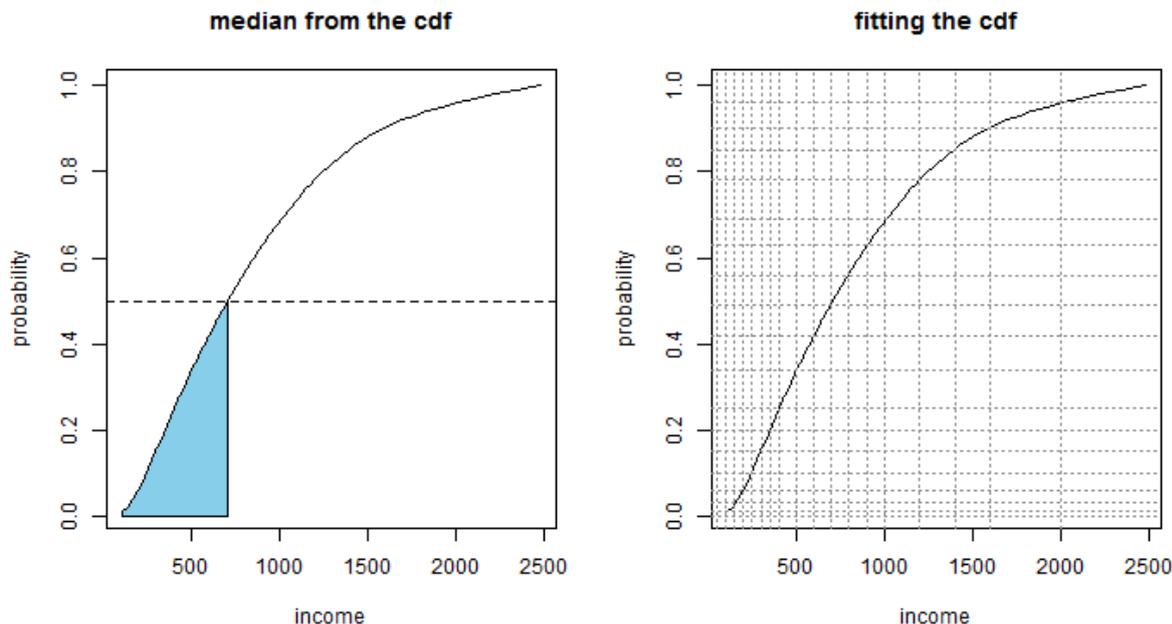
So we don't know enough about the income distribution to estimate the value of income for each percentile across the income distribution. To work this out (as in the box above describes) we need to construct, or interpolate, more data points along the income distribution for each area.

### ***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 do not diverge from the base data on which they are based,
- 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,
- they have been compared to the income estimates used in the last version of the HNDA Tool (2014), that were produced by Bramley and Watkins in 2013, in order to check that they are not diverging more than might be reasonably expected given that a new methodology and new input data have been used for the current estimates.

## **LIMITATIONS**

### ***Experimental Statistics***

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

- the base date on which these interpolated income estimates are based - the [Local Level Household Model Income Estimates](#) – are Experimental Statistics: data being developed which are new statistics being developed and which were published to involve users and stakeholders in their development and to build in quality and understanding at an early stage.
- the interpolated income estimates for 2014 have been produced for the purposes of updating the Scottish Government Housing Need and Demand Assessment (HNDA) Tool.
- the interpolated estimates will 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 RESULTS**

Table 1 below sets out summary results by local authority and for Scotland. Key findings include:

- Median incomes (at the 50<sup>th</sup> percentile) varied from £700 per week in Aberdeenshire to £460 per week in Dundee City. This compares to a Scottish median of £550 per week.
- Lower quartile incomes (at the 25<sup>th</sup> percentile) varied from £390 per week in Aberdeenshire to £280 per week in Dundee City. This compares to a Scottish lower quartile of £330 per week.

Table 1 - Interpolated Household Income Estimates 2014, by Income Decile and Local Authority

2014	Gross WEEKLY Household Income (£) <sup>1,2</sup>		
	Lower Quartile	Median	Upper Quartile
Local Authority	25 <sup>th</sup> Percentile	50 <sup>th</sup> Percentile	75 <sup>th</sup> Percentile
Aberdeen City	£350	£590	£950
Aberdeenshire	£390	£700	£1,100
Angus	£320	£540	£860
Argyll and Bute	£320	£530	£840
City of Edinburgh	£340	£580	£930
Clackmannanshire	£330	£540	£860
Dumfries and Galloway	£320	£510	£810
Dundee City	£280	£460	£740
East Ayrshire	£320	£530	£840
East Dunbartonshire	£370	£650	£1,080
East Lothian	£350	£610	£980
East Renfrewshire	£360	£660	£1,100
Falkirk	£330	£560	£900
Fife	£330	£540	£860
Glasgow City	£300	£470	£760
Highland	£350	£590	£920
Inverclyde	£300	£500	£810
Midlothian	£360	£610	£960
Moray	£350	£580	£900
Na h-Eileanan an Iar	£340	£560	£880
North Ayrshire	£320	£510	£800
North Lanarkshire	£330	£550	£870
Orkney Islands	£330	£570	£910
Perth and Kinross	£350	£580	£930
Renfrewshire	£310	£510	£850
Scottish Borders	£320	£530	£840
Shetland Islands	£370	£640	£990
South Ayrshire	£320	£530	£850
South Lanarkshire	£330	£560	£910
Stirling	£350	£610	£990
West Dunbartonshire	£300	£490	£790
West Lothian	£360	£610	£970
<b>SCOTLAND</b>	<b>£330</b>	<b>£800</b>	<b>£890</b>

Source: Scottish Government, Communities Analysis Division (2019).

Notes:

1. This analysis is based on the Scottish Government Local Level Household Income Estimates (LLHIE) 2014.
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.

## 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](mailto:chma@gov.scot).

**Scottish Government**  
**June 2019**