Experimental Statistics: data being developed - These are new statistics being developed and have been published to involve users and stakeholders in their development and to build in quality and understanding at an early stage.

Local Level Household Income Model Estimates for 2014

A Short Non-Technical Guide for Housing Affordability Analysis (2017)

Scottish Government
Communities Analysis Division

Version 1.2

About this Guide

This guide provides a short, non-technical overview of the Local Level Household Income Model (LLHIM). It describes the income estimates, how they were produced and assessed, what they were designed for and their limitations. A more detailed key findings paper that includes a full summary of the methodology is also available.

The Data Providers

The model was commissioned by the Scottish Government (SG) Communities Analytical Division (CAD) and has been produced by I-SPHERE at Heriot-Watt University (Glen Bramley, Paul Cosgrove and David Watkins) in association with David Simmonds Consultancy.

Access and updates

The intention is that the model will be updated annually for the duration of the research contract period. The data are accessed at statistics.gov.scot which lets you map the data or as an Excel spreadsheet at chma.gov.scot.

Data Uses

The primary purpose of the income statistics 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 at sub-Local Authority level.

Local gross household income data is important for the tool because this allows income levels to be compared to other data on house prices and rental costs to ascertain level of housing affordability across particular geographic areas.

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.

Experimental Statistics: data being developed - These are new statistics being developed and have been published to involve users and stakeholders in their development and to build in quality and understanding at an early stage.

Experimental Statistics

These are Experimental Statistics: Data Being Developed and have been badged this way as they are new statistics being developed and have been published to involve users in their development and to build quality and understanding at an early stage. However, SG CAD considers the estimates to be sufficiently robust for the purpose of assessing future housing need and demand using the HDNA Tool.

Describing the data

These estimates are 2014-based and go down to 2011 data zone level.

The model provides gross weekly income i.e. gross earnings from employment, welfare benefits, tax credits, housing benefit, savings and pensions etc.

Banded income and the proportion of households falling into each band are also provided.

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.

Experimental Statistics: data being developed - These are new statistics being developed and have been published to involve users and stakeholders in their development and to build in quality and understanding at an early stage.

Data Extract

6. An extract of the dataset is shown below. Most variables are self-evident but some are described.

2011 Data Zone code	2011 Data Zone name	Intermediate Zone code	Intermediate Zone name	Local Authority Area code	Local Authority Area name	Total Dwellings 2014 (NRS)	Occupied Dwellings 2014 (NRS)
S01006506	Culter - 01	S02001236	Culter	S12000033	Aberdeen City	479	464
S01006507	Culter - 02	S02001236	Culter	S12000033	Aberdeen City	370	364
S01006508	Culter - 03	S02001236	Culter	S12000033	Aberdeen City	322	309
S01006509	Culter - 04	S02001236	Culter	S12000033	Aberdeen City	276	267

…….continued

SIMD16 Rank	SIMD16 Income Domain Rank	SIMD16 Percentage of people who are income deprived	Mean Gross Weekly Income	Mean Gross Weekly Income (rank)	Median Gross Weekly Income	Median Gross Weekly Income (rank)	Weekly Income under £50 (proportion of households)	Weekly Income under £100 (proportion of households)	Weekly Income under £150 (proportion of households)	Weekly Income under £200 (proportion of households)	Weekly Income under £250 (proportion of households)	Weekly Income under £300 (proportion of households)	Weekly Income under £350 (proportion of households)
5272	4513	0.07	835	5774	707	5625	0.0011	0.008	0.0273	0.0608	0.1037	0.151	0.1993
4838	4334	0.07	786	5346	675	5297	0.0014	0.0102	0.0335	0.0725	0.1206	0.1716	0.222
6321	5266	0.05	845	5874	733	5833	0.0003	0.0039	0.018	0.0449	0.081	0.1225	0.1667
5363	5002	0.05	755	4996	647	4921	0.0015	0.0108	0.0361	0.0778	0.1283	0.1813	0.2333

……..continued

Weekly Income under £400 (proportion of households)	Weekly Income under £500 (proportion of households)	Weekly Income under £600 (proportion of households)	Weekly Income under £700 (proportion of households)	Weekly Income under £800 (proportion of households)	Weekly Income under £900 (proportion of households)	Weekly Income under £1,000 (proportion of households)	Weekly Income under £1,200 (proportion of households)	Weekly Income under £1,400 (proportion of households)	Weekly Income under £1,600 (proportion of households)	Weekly Income under £2,000 (proportion of households)	Approx proportion of households under 60% of median gross income	Approx proportion of households under 60% of median gross income (rank)
0.2468	0.3367	0.4192	0.4949	0.5645	0.6281	0.6855	0.7811	0.8522	0.9022	0.9586	0.0856	6358
0.2702	0.3599	0.4422	0.5192	0.5911	0.6569	0.7159	0.8113	0.8785	0.9233	0.97	0.1096	5503
0.2121	0.3031	0.3912	0.4744	0.5513	0.6213	0.6837	0.7856	0.8588	0.9089	0.9631	0.0634	6785
0.2834	0.3768	0.4627	0.542	0.6146	0.6802	0.7379	0.8296	0.8927	0.9337	0.9751	0.111	5442

……..continued

Experimental Statistics: data being developed - These are new statistics being developed and have been published to involve users and stakeholders in their development and to build in quality and understanding at an early stage.

Weekly Income under £50 number of households)	Weekly Income under £100 number of households)	Weekly Income under £150 (number of households)	Weekly Income under £200 (number of households)	Weekly Income under £250 (number of households)	Weekly Income under £300 (number of households)	Weekly Income under £350 (number of households)	Weekly Income under £400 (number of households)	Weekly Income under £500 (number of households)	Weekly Income under £600 (number of households)	Weekly Income under £700 (number of households)	Weekly Income under £800 (number of households)	Weekly Income under £900 (number of households)	Weekly Income under £1,000 (number of households)	Weekly Income nder £1,200 (number of households)	Weekly Income under £1,400 (number of households)	Weekly Income under £1,600 (number of households)	Weekly Income under £2,000 (number of households)
1	4	13	28	48	70	92	115	156	195	230	262	291	318	362	395	419	445
1	4	12	26	44	62	81	98	131	161	189	215	239	261	295	320	336	353
0	1	6	14	25	38	52	66	94	121	147	170	192	211	243	265	281	298
0	3	10	21	34	48	62	76	101	124	145	164	182	197	222	238	249	260

- **'Approx proportion of households under 60% of median gross income'** Before Housing Costs (i.e. with an income of under £330 per week after adjusting for household size) is an estimate of households on low incomes said to be in poverty. This is based on a standard UK definition.

- **'SIMD16 percentage of people who are income deprived'** is another estimate of low income which is defined as people who are income deprived and receiving certain benefits and tax credits.

- **'SIMD16 income domain rank'** ranks the % of people who receive certain benefits due to low income. Rank 1 is the most deprived and rank 6,976 the least deprived.

The Methodology

A non-technical summary of the basic methodology is set out below and this is also shown as a diagram. The project team used a variety of data source to construct the model including:

> the Family Resource Survey (FRS),
> the Scottish Household Survey (SHS) and,
> the Improvement Service/ Scottish Government (Bramley & Watkins 2013) study on Local Income and Poverty in Scotland.
> the Understanding Society Longitudinal Survey (UKHLS)
> the Scottish Census (2011)
> the ONS Standard Occupational Classification (2010)
> Transport Scotland's Transport Model for Scotland (TMfS).

The approach taken is to understand as much as possible about the actual incomes and geo-socio-economic circumstances or determinants of incomes for different types of households using actual survey data and research.

These income determinants are tested and then selected for the model using a statistical technique called regression to see if they are strong or weak predictors of income. How well the model predicts is called it's 'goodness-of-fit' and this can be quantified in terms of a value of R^2 (R-squared).

The tested income determinants are then used to predict the incomes of *all* households types with similar geo-socio-economic circumstances. The 2011 Census, and other local data sources, are used to provide the necessary information about *all* households types.

Net SHS income is converted to gross income before any modelling/estimation takes place. The definition of gross income is in line with the FRS and the UKHLS. Gross income includes welfare benefits, tax credits and housing benefit.

The regression model included consideration of the following potential income determinants;

> age
> household type & size
> occupation
> work participation per household
> prevalence of part-time work
> earnings
> car ownership
> means tested benefits

> SIMD poverty indicators
> unemployment
> qualifications
> ethnicity
> Travel-to-Work Areas
> Housing Market Areas (where people buy and sell houses)
> house prices, size, tenure & council tax band

For the purposes of modelling, households are formed into nine types with certain socio-economic profiles. Then the income variance within each household type and between household types is examined in order to decide that the groups are distinct from each other:

1. Single adult under 50
2. Single adult 50+
3. Singe adult Retirement Age+
4. Lone parent with children
5. 2+ adults under 50
6. 2+ adults 50+
7. 2+ adults + children
8. 2+ adults Retirement Age+
9. All student

The survey data are also used to work out what the distribution of income is for each group i.e. whose income is lower, who's higher and who's in-between. Income distributions for each household type are estimated use a statistical approach known as lognormal distributions. The difference between the actual and estimated distributions is minimised as far as possible to get the best fit.

Both these income distributions and the income determinants are then used to predict the incomes of *all* similar household group using the Census and other local level data. The use of both income distributions *and* income determinants, along with a disaggregation of households into different groups is what the project team refer to as a 'blended' methodological approach. The approach allows for estimates to be produced at a datazone level.

At a national level the resulting income estimates are constrained/ controlled back to actual income data on the FRS to ensure that they are consistent with it and to make sure the estimates are kept within real survey data parameters. The local area income estimates, however, will reflect local patterns of income based on actual local data combined with the predictive element of the model.

Outlier (extreme) income estimates are removed to ensure that they do not skew the data.

Testing Results, Quality & Robustness

In terms of how robust the estimates are there are a number of ways to judge this.

'Goodness-of-fit'

A common measure of how good a model is at predicting something is called its 'goodness-of-fit' and this is defined in terms of the value of R-Squared (R^2). This is a measure of the proportion of variance in the data (in this case income) that the model explains. A model with a value close to 0 explains very little whereas a model with an R^2 close to 1 explains almost all of the variance.

Models with very high R^2 values can be over-fitted and should be judged with care i.e. you can explain more of the variation, and get high R^2 values, simply by using feeding in more and more variables. Good models should use the least number of variables possible to explain the variance.

It is also generally easier to achieve high R^2 values where you are using time series data (which is not the case here). Values of R^2 are often much lower when using cross-sectional data as is the case here i.e. lots of different variables from one point in time.

This (cross-sectional) model has the following R^2 values for:

> Single adults households of working age, **R^2 =0.45**
> Multi adult households of working age, **R^2 = 0.45**
> Single adults households of retirement age, **R^2 = 0.15**
> Multi adult households of retirement age, **R^2 = 0.22**

For single adults, therefore, the model (and the factors included in it) can explain almost half (45%) of the variation in income for that group.

One of the reasons why the model is a better predictor of incomes for those of working age than those of retirement age is because the SHS has information about their occupation but it doesn't give the former occupations of retired households.

Constraining

One way to help quality control the modelled income estimates has been to constrain them back to actual incomes recorded on the FRS. This means that the modelled income estimates are kept in line with actual incomes.

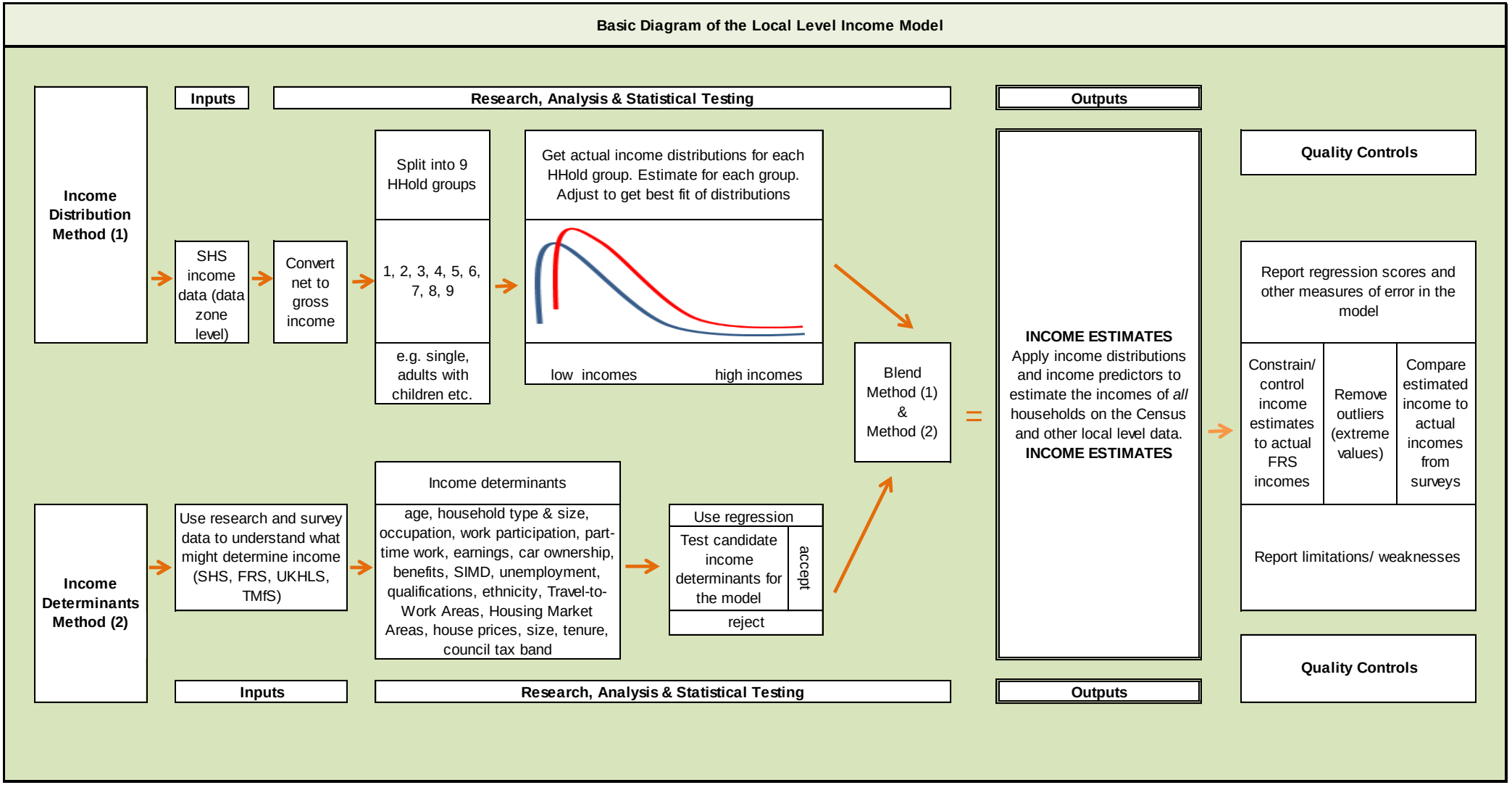
Outliers & Cells with insufficient data

The results were assessed for outliners (extreme values) and four data zones were removed from the final results due to high income level that appeared anomalously high. The data were also assessed for blank cells where there were insufficient data to drive the model. Two such datazone was detected and removed.

Seeking Expert Views

The project team engaged with the research team responsible for the UKHLS (London School of Economics and ISER-Essex University) to determine that the income data on this survey was adequate. They reported no reason to doubt the quality of the UKHLS income data.

Experimental Statistics: data being developed - These are new statistics being developed and have been published to involve users and stakeholders in their development and to build in quality and understanding at an early stage.



Limitations

What the income estimates are not

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.

Unequivalised income

The household income figures presented here are unequivalised. Unequivalised household income is commonly used when considering housing affordability. However for some purposes, including analysing poverty rates, it is usual to use equivalised household income instead.

Equivalised household income takes account of different household sizes, reflecting the fact that, for example, a small household will need less income than a large household to achieve the same standard of living.

This should be borne in mind when considering these figures. For example, one area may appear to have a lower income than another partially or wholly because it has lots of smaller households with correspondingly lower incomes than the other area.

Consider other income sources

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.

Experimental Statistics: data being developed - These are new statistics being developed and have been published to involve users and stakeholders in their development and to build in quality and understanding at an early stage.

Experimental statistics

These are Experimental Statistic: Data Being Developed and have been published to involve uses and stakeholder in their development and to build in quality and understanding at an early stage.

SHS missing income data

Around a third of respondents did not answer the income question. The missing values were imputed based on similar households who did answer the income question. This is another way in which error may be introduced into the modelled estimates.

Other SHS issues

The SHS income questions are very basic compared to the FRS questions for example and several elements of income are not recorded and must be imputed. In addition responses are recalled from memory only and are not based on documentary evidence. This increases the likelihood of error around the estimates.

Student households

The incomes of students are very variable and this makes them difficult to predict using regression. Of the different household groups that were used in the modelling process, the project team express some concern about the accuracy of estimates for this group.

Example Results

Some example results are set out below.

Local Authority	Mean Gross Weekly Income (£)
Aberdeen City	722
Aberdeenshire	818
Angus	646
Argyll and Bute	635
City of Edinburgh	701
Clackmannanshire	654
Dumfries and Galloway	611
Dundee City	566
East Ayrshire	640
East Dunbartonshire	795
East Lothian	730
East Renfrewshire	809
Falkirk	679
Fife	648
Glasgow City	580
Highland	691
Inverclyde	616
Midlothian	718
Moray	679
Na h-Eileanan an Iar	661
North Ayrshire	613
North Lanarkshire	659
Orkney Islands	672
Perth and Kinross	697
Renfrewshire	640
Scottish Borders	633
Shetland Islands	737
South Ayrshire	646
South Lanarkshire	682
Stirling	741
West Dunbartonshire	597
West Lothian	725
Scotland	**668**

Your Feedback

These data are Experimental Statistics: Data Being Developed and should be viewed as in development. As part of this we would find your feedback useful. For example:

> is there any additional analysis you would find useful?
> do you have any questions about the dataset?
> what might you use the data for?

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