GLA household income estimates
Methodology of household income data model

July 2015

The GLA household income estimates replace the experimental estimates released in May 2013. The previous estimates covered 2011/12 only. The data for 2011/12 has been revised in these new estimates due to an updated methodology.

Why are small area income estimates needed?

- The most recent Office for National Statistics (ONS) estimates at Middle Super Output Area (MSOA) level for 2007/08 are useful but only available for MSOAs and are now several years out of date.
- It is not yet clear what income estimates ONS will publish in the future, and at what geography.
- CACI/Experian and other companies charge for their data, and there are strict limits on how any organisation can use and publish this data.
- There is a high demand for income estimates for small areas shown by webstats from the London Datastore.

The intention is to create gross annual household income estimates to cover Lower Super Output Areas (LSOA), MSOAs, Wards and Boroughs in London as well as regions in the UK. The previous estimates from 2013 also covered Output Area data but there was little demand for this level of granularity, and additionally because of the level of complexity required to work on data for areas that small, data for output area geography has not been published in these estimates.
The data is average (mean and median) unequivalised income, which means they take no account of household size or composition within each area.

The model has been designed so that the estimates can be updated in the future. Part of the model relies on Census data, which is currently held once every 10 years so to some degree for the model to work in the long term, similar Census outputs will be required in future censuses, unless the model is altered. It is possible that the model could be adapted to use other data sources, should better indicator data become available.

Regional estimates

- London level mean and median income data were calculated from ESRC *Understanding Society* dataset[1], waves 1-4 (2009/10-2012/13) and British Household Panel Study (2001-2008). The data were smoothed using a moving average. The 2007 figures were upscaled to match the 2007/08 ONS household income data (annualised simply by multiplying by 12). All other years were calculated relative to 2007. (Variable "gross household income: month before interview")
- Any differences in household income that may occur during the year, for example Christmas bonuses, will be taken account of because a roughly equal proportion of the

[1] See appendix for an explanation of each of the sources of data used in the model

sample were surveyed in each month of the year. Therefore the effects of seasonality on working patterns and income have been smoothed.
- The regional figures gave averages very close to the Family Resources Survey data. Therefore this survey was deemed suitable to use as a basis for the rest of the model.
- All incomes below London level are modelled.

Borough and Small area estimates

Geographies below regional level have been modelled based on:
- NS-SEC of people (Census),
- Household deprivation in x number of dimensions (Census),
- Median house selling prices (land registry),
- Child Poverty data (HM Revenue and Customs)
- ONS Household Income Estimates from 2001, 2004, and 2007.

A fuller description of these indicators/sources can be found in the appendix.

These indicators have a strong statistical relationship with income. Using a number of indicators reduces the risk of having outliers (that is unexpectedly high or low results) from one of the datasets. Furthermore, each indicator is deemed to be sufficiently different from any of the others to bring useful additional data into the model for example, deprivation and child poverty are poverty indicators and house prices is a measure of wealth.

Weighting the factors
- The standard deviation (SD) for each indicator by year and geography type was calculated. This shows how much variation or dispersion from the average exists i.e. in a particular dataset a measure of how much all the values are either bunched closely together or spread far apart, and their distance from the mean.
- Each of these factors was turned into a Z score based on the standard deviation (SD) of each indicator. A Z score is a way of standardising data on one scale so a comparison can take place between different types of data. The Z scores used here describe how much a point deviates from the London average. For example, if an area returns a Z score of +1.5 it is interpreted as +1.5 standard deviations away from the average. A Z score can be positive or negative, with negative scores indicating a value below the London average.
- A sum of the five indicators was then calculated using the following weightings NS-sec 25%, Household deprivation 20%, Child Poverty 15%, House prices 25%, and ONS Income 15%.

Calculation of the income figures
- The average weighted score from the five indicators from 2007 was placed with the annualised ONS income data (2007/08) at MSOA level. A scatterplot was created and a polynomial trend line added for each of mean and median income. The trend line showed the relationship between the indicator score and household income. The equation of the trend line was then applied to the scores from all years to give an income figure for each area and each year.
- These figures were adjusted for other years using 2007 as the index year (100%), using the London regional figures 2001-2012.

Final adjustments

- At borough level the difference to the London average was calculated using a mix of Annual Survey of Hours and Earnings (ASHE from ONS) and Survey of Personal Income data (HMRC) with a 25/75% weighting respectively because the SPI is a better source of income as it also includes pension, investments and self-employment income. These are two different sources to help increase reliability.
- Each of these sources provided the percentage difference to the London average for each borough. The mean and median values had separate calculations.
- The results from the modelled results at borough level were analysed to find the difference from the London average and compared with ASHE/SPI borough figures.
- A figure half way between the modelled figure and the ASHE/SPI figure was used to adjust all areas based on the borough they belonged to (LSOA, MSOA, wards and the borough itself). This adjustment was required because even if two areas have the same weighted indicator score, the average costs of living can vary across the boroughs so the adjustment is needed to compensate for that difference.

Appendix

Understanding Society is a longitudinal survey and study that captures information every year about the social and economic circumstances and attitudes of people living in 40,000 UK households. It began in 2009 and replaced the **British Household Panel Survey** (BHPS). Many of the respondents in the BHPS are still being surveyed in this study. Funding comes primarily from the Economic and Social Research Council (ESRC). The survey is designed and managed by a team of longitudinal survey experts at the Institute for Social and Economic Research (ISER), at the University of Essex. Data in the model used household income data broken down by region.

The data from the England and Wales **Census** used in this model took place on 29 April 2001 and 27 March 2011.
NS-SEC of People aged 16-74 data was taken from Table KS14 (2001) and QS607 (2011). The National Statistics Socio-economic Classification (NS-SEC) is the primary social classification in the United Kingdom. The appropriate 'mix' in NS-SEC variables was calculated using income figures from the Understanding Society dataset wave 3. The aim was to produce a single figure for each of these indicators for each area. The average level of income for each category within the indicator was compared with the average income.

Households by deprivation dimensions data was taken from Table UV67 (2001) and QS119 (2011).
The dimensions of deprivation used to classify households are indicators based on the four selected household characteristics. A household is deprived in a dimension if they meet one or more of the following conditions:

- Employment: any member of a household not a full-time student is either unemployed or long-term sick,
- Education: no person in the household has at least level 2 education, and no person aged 16-18 is a full-time student,
- Health and disability: any person in the household has general health 'bad or very bad' or has a long-term health problem, and
- Housing: Household's accommodation is ether overcrowded, with an occupancy rating -1 or less, or is in a shared dwelling, or has no central heating.

A household is classified as being deprived in none, or one to four of these dimensions in any combination.

A straight line trend was used between 2001 and 2011 to interpolate data for the intervening years.

House prices data comes from Land Registry price paid dataset, which records the selling price of every residential property sold in England and Wales. The dataset is one of the most reliable sources of house price information. In order to increase reliability three-year averages (mean and median) were calculated. The same method was used for all levels of geography for consistency.

HMRC Child Poverty This Children in Low-Income Families Local Measure shows the proportion of children living in families in receipt of out-of-work (means-tested) benefits or in receipt of tax credits where their reported income is less than 60 per cent of UK median income. This measure provides a broad proxy for relative low income child poverty as set out in the Child Poverty Act 2010, and enables analysis at a local level. Statistics are published at various levels of geography providing an annual snapshot as at 31 August from 2006 onwards.

ONS Model-Based Income Estimates. This dataset provides model based Income Estimates for Households for wards 2001/12, and MSOAs 2004/05 and 2007/08. Only the average weekly household total income (unequivalised) was used in the GLA model. Missing years were interpolated and extrapolated. LSOAs were given the same figure as the MSOA that they were contained in.

Annual Survey of Hours and Earnings (ASHE) data is annual gross pay based on place of residence. ASHE is based on a one per cent sample of employee jobs taken from HM Revenue & Customs (HMRC) PAYE records. Information on earnings and hours is obtained from employers and treated confidentially. ASHE does not cover the self-employed nor does it cover employees not paid during the reference period. The earnings information presented relates to gross pay before tax, National Insurance or other deductions.

The Survey of Personal Incomes by HMRC includes mean and median Income (Personal incomes by tax year) since 1999. These are estimates based on a survey with an annual sample of HMRC records for individuals who could be liable to UK Income Tax. The data includes income related to employment, self-employment, pension income and investment income.

In all cases where geographical boundaries of the source data was not the same as the output boundaries, the data was modelled based on the percentage of one area contained within the other.

For more information please contact Gareth Piggott, GLA Intelligence
Greater London Authority, City Hall, The Queen's Walk, More London, London SE1 2AA
Tel: 020 7983 4327 e-mail: Gareth.piggott@london.gov.uk