Decentralisation of Poverty - Replication Analysis

Jon Minton

5 August 2016

# Caveats

* Work very much in progress
* Please don't steal

# Introduction

This is the introduction section to the report.

## Background

In 2014 Kavanagh, Lee & Pryce produced the AQMeN briefing report [Poverty in Suburbia: Has Glasgow gone the way of American cities?](http://www.research.ed.ac.uk/portal/files/18805757/RB5_poverty_suburbia.pdf). This paper argued that there has been a decentralisation of poverty in Glasgow between 2001 and 2011, and combined a Bayesian spatial statistics approach with the Relative Centralisation Index (RCI) in order to argue firstly that the spatial distribution of poverty has become less decentralised between the two periods, and secondly that the changes in RCE scores are statistically significant. The briefing paper defined 'poverty' in three ways, each relating to specific UK state benefit receipts for persons of working age:

1. The proportion of local populations receiving Income Support (IS)
2. The proportion of local populations receiving Incapacity Benefit (IB)
3. The proportion of local populations receiving Jobseekers' Allowance (JSA)

These analyses were expanded to the four major Scottish cities - Aberdeen, Dundee, Edinburgh and Glasgow - for a paper to appear in *Annals of the American Association of Geographers*. This paper calculated 95% credible intervals (CrIs) for the difference in RCI from 2001 to 2011 for each of the three measures, and each of the four cities, between 2001 and 2011. It found falls in RCI for each benefit type, and for all four cities, which were statistically significant.

## Purpose

The purposes of this briefing paper are two-fold:

1. To reproduce the analysis in Kavanagh, Lee & Pryce 2016 using the proportion of the population 'income deprived' between 2004 and 2016, using Scottish Index of Multiple Deprivation (SIMD) definitions of income deprivation. This will allow the dependence of Kavanagh, Lee & Pryce's conclusions on the measures and time periods used to be assessed.
2. To use descriptive statistics and data visualisation to better understand the contribution of both changes in the numerators (population counts 'in poverty') and changes in denominators (total population counts) to overall falls in RCI.

# Data and methods

## Data

The SIMD is used by the Scottish Government to identify places suffering from deprivation. Different editions of the SIMD have been produced in 2004, 2006, 2012, with a new release scheduled for August 2016, and made available at the small area known as the datazone, equivalent to the Lower Super Output Area (LSOA) in England & Wales. The methodology differs slightly from one edition to the next, but for all releases estimates of the number of individuals within each datazone who are 'income deprived', as well as overall and domain-specific scores and rankings for deprivation. The definition of 'income deprived' varies between household types and age groups, to account for the different needs of different demographic groups and the efficiency gains associated with living in multi-person rather than single person households. SIMD data releases for 2004, 2006, 2012 and 2016 were downloaded from the [SIMD website](http://www.gov.scot/Topics/Statistics/SIMD/DataAnalysis). The total population counts, and the population count of the subpopulation defined as 'income deprived', were extracted from each release and combined in a single dataset. The 'income deprived' and 'total' population counts were the focus of this report as the RCI measure, like many other segregation measures, involves comparing two mutually exclusive groups.

## Methods

Producing the results involved doing the following:

* Defining centroids for each place.
* Identifying the closest centroid for each datazone.
* Reweighting income deprived and total population counts from the SIMD, released at 2001 datazones, onto 2011 datazones.
* etc
* etc

### Calculating nearest centres

The following definitions of city centres were used

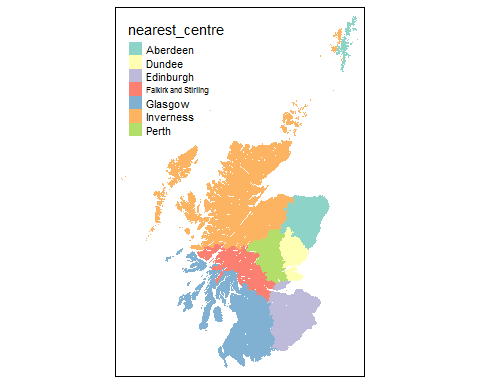
|  |  |  |  |
| --- | --- | --- | --- |
| Place | 2011 Datazone | Postcode | Description |
| Aberdeen | S01006646 | AB10 1AN | Shoe Lane |
| Glasgow | S01010265 | G1 3BU | West End of George Square |
| Edinburgh | S01008677 | EH1 1BQ | 31 Waverley Bridge |
| Dundee | S01007705 | DD1 2AJ | Commercial Street |
| Inverness | S01010620 | IV1 1HY | High Street/Castle Street |
| Perth | S01011939 | PH2 8PA | South Street |
| Falkirk and Stirling | S01013067 | FK8 2LJ | Port Street |

Kavanagh, Lee & Pryce (2016) produced a sensitivity analysis to assess the dependence of their results on the choice of centre, indicating they were robust to this choice. Alternative datazone centres could be used to explore the effect of definition for these analyses too.

The majority of the results presented will be for the four cities Aberdeen, Dundee, Edinburgh and Glasgow, as these are the largest and most centralised of the areas considered. Within this Glasgow is historically the most populous, has the highest population density, and has the highest concentration of deprived areas.

### Apportioning nearest centre to datazones

Below is a map of Scotland in which each datazone has been coloured according to the nearest centre.



As this map indicates places in the Orkney and Shetland Islands, and elsewhere in the remote and rural highlands and islands, are arbitrily assigned to either Aberdeen or Inverness, even though they are not part of any meaningful functional urban geography. However these are easily identifiable through their absolute distance from either centre, and are not part of the main analysis.

### Linking 2001 to 2011 datazones

Income deprived and total population counts were reweighted from 2001 to 2011 datazone boundaries using a lookup file developed by [Paul Norman](http://www.geog.leeds.ac.uk/people/p.norman). This allowed weights to be produced, based on census output area overlap, allowing weighted 2011 population count estimates to be produced from counts enumerated at 2001 datazone level, based on the the level of overlap between 2001 and 2011 datazones. Further details about the lookup file was produced are available from [Paul Norman](p.d.norman@leeds.ac.uk) and the code used to perform the reweighting is reproduced below.

# Results

## RCI calculations

This section will present the results of the RCI calculations, i.e. more formal attempt to replicate the results of Kavanagh, Pryce & Lee (2016) using a different indicator and data source.

### RCI, Aberdeen

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.218 | 0.212 | 0.224 |
| 2006 | 0.227 | 0.221 | 0.233 |
| 2009 | 0.200 | 0.194 | 0.206 |
| 2012 | 0.187 | 0.181 | 0.194 |

### RCI, Dundee

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.304 | 0.298 | 0.309 |
| 2006 | 0.295 | 0.290 | 0.301 |
| 2009 | 0.266 | 0.261 | 0.272 |
| 2012 | 0.256 | 0.250 | 0.261 |

### RCI, Edinburgh

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.013 | 0.009 | 0.017 |
| 2006 | 0.020 | 0.016 | 0.023 |
| 2009 | -0.005 | -0.009 | -0.002 |
| 2012 | -0.013 | -0.017 | -0.009 |

### RCI, Glasgow

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.144 | 0.142 | 0.146 |
| 2006 | 0.140 | 0.137 | 0.142 |
| 2009 | 0.125 | 0.123 | 0.128 |
| 2012 | 0.093 | 0.091 | 0.096 |

## Dissimilarity calculations

This section will present the results of the Dissimilarity index calculations, using the Bayesian modelling approach used in Kavanagh et al 2016.

### D, Aberdeen

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.382 | 0.377 | 0.388 |
| 2006 | 0.381 | 0.375 | 0.386 |
| 2009 | 0.379 | 0.373 | 0.384 |
| 2012 | 0.361 | 0.356 | 0.367 |

### D, Dundee

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.412 | 0.408 | 0.417 |
| 2006 | 0.401 | 0.396 | 0.406 |
| 2009 | 0.383 | 0.378 | 0.388 |
| 2012 | 0.376 | 0.371 | 0.380 |

### D, Edinburgh

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.385 | 0.382 | 0.388 |
| 2006 | 0.359 | 0.356 | 0.362 |
| 2009 | 0.339 | 0.336 | 0.342 |
| 2012 | 0.320 | 0.316 | 0.323 |

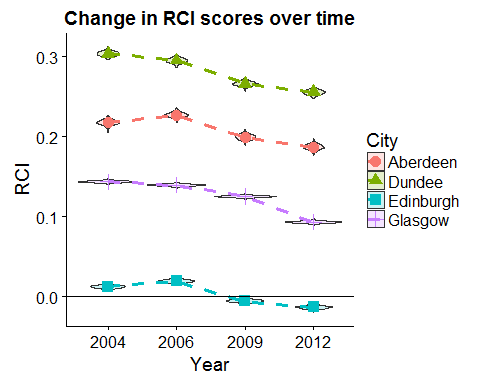
### RCI, Glasgow

|  |  |  |  |
| --- | --- | --- | --- |
| period | median | lower | upper |
| 2004 | 0.367 | 0.366 | 0.369 |
| 2006 | 0.348 | 0.347 | 0.350 |
| 2009 | 0.333 | 0.332 | 0.335 |
| 2012 | 0.316 | 0.314 | 0.317 |

### Change in RCI and D over time

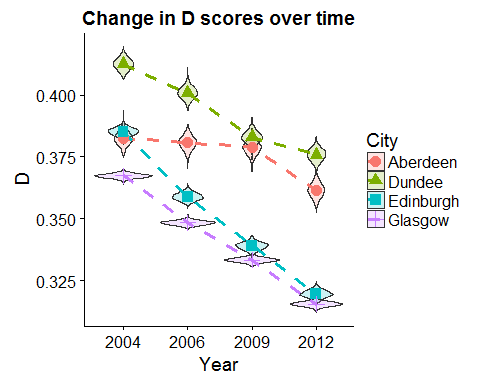
Unlike the census-based variables used in Kavanagh, Lee & Pryce (2016), there are SIMD releases for 2004, 2006, 2009, 2012, and 2016, meaning in RCI and D scores within the period 2001 and 2011 can be observed. The Bayesian spatial modelling approach used by Kavanagh, Lee & Pryce (2016) was applied to each of the SIMD releases, and median RCI and D estimates, along with 95% CrIs, were produced for each release and for each city.

The graph of RCI over time is shown below. The density of RCI estimates for each place and in each year is shown using a violin plot, with longer shapes indicating a greater range of posterior estimates. The median RCI scores are indicated as points, linked for each city with dashed lines.



The overall trend in each place is for RCI to reduce from 2004 to 2012. However within each period the ordering of RCIs between places has not changed, with Dundee showing the most centralisation of poverty, then Aberdeen, and Glasgow, with Edinburgh the most decentralised. In terms of trends over time, Glasgow and Dundee both show steadily declining RCI between SIMD releases, whereas for Aberdeen and Edinbugh RCI values increased from 2004 to 2006, before declining in 2009 and 2012. From 2009 onwards Edinburgh has negative RCI scores, suggesting a peripheralisation rather than centralisation of poverty.

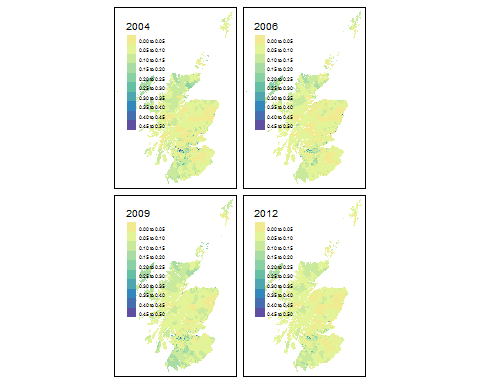
The equivalent graph for the Dissimilarity index is shown below:



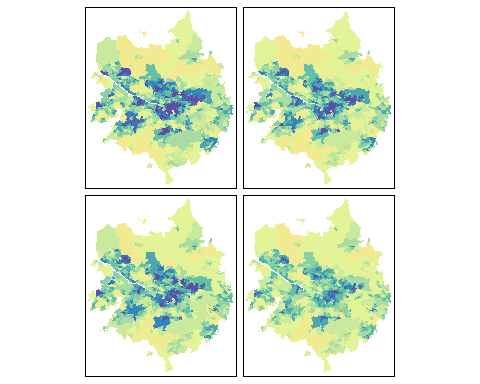
As with RCI, Dissimilarity scores have also fallen over time for all four cities. The greatest contrast in the trends is between Aberdeen and Edinburgh. In 2004 both cities had similar and overlapping D scores, but between 2004 and 2009 D fell rapidly for Edinburgh, while remaining fairly stable in Aberdeen, before falling from 2009 to 2012. Glasgow has always had the lowest D scores of the four cities, and Dundee the highest, but since 2009 the distribution of D estimates for Edinburgh have started to overlap those of Glasgow as the rate of decline in D has been slightly greater in Edinburgh than Glasgow. Similarly, the the greatest fall in D occurred between 2006 and 2009.

## Maps

The following map shows the proportion of the population who are classified as income deprived for each datazone, in the years 2004, 2006, 2009 and 2012. Given that so much of Scotland is rural, few areas appear income deprived using a standard choropleth. However, patterns of high income deprivation are apparent in and near Glasgow.

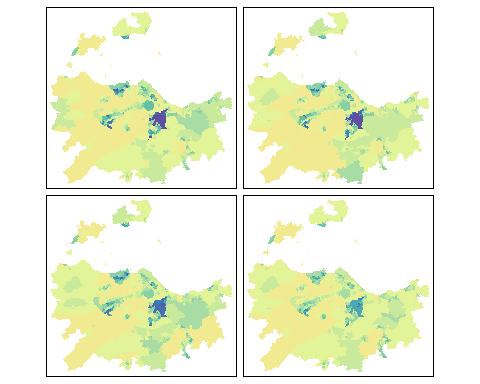


The map below focuses in on income deprivation within Glasgow, and in particular within datazones not more than 15 km from Glasgow city centre.



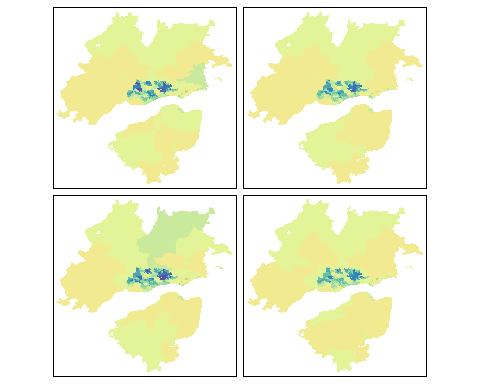
From this map there are suggestions of a decentralisation of poverty, at least as defined as income deprived, in particular between 2009 and 2012. However there are also clear pockets of areas, in the south, east and north west of Glasgow city, with persistently high levels of income deprivation.

An equivalent map for Edinburgh is produced below:



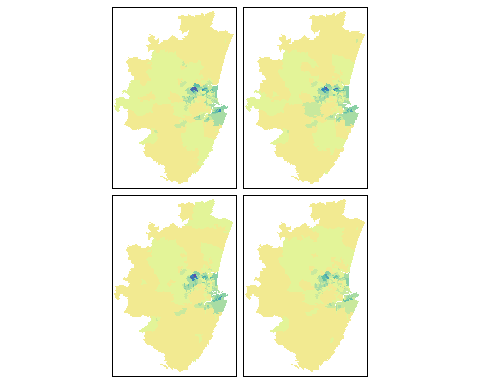
This map also suggests that, as in Glasgow, drivers of increasing levels of decentralisation of poverty in Edinburgh include both rapid falls in poverty in the centre, alongside some more suburban areas with persistently high levels of poverty. In Edinburgh's case, in pockets towards the West, North West, and South/South East of the city centre.

The equivalent map for Dundee is shown below:



Within Dundee the patterns of spatial persistence of poverty appear most consistent over time, with a clearer demarcation between urban and non-urban areas. The functional urban geography of Dundee appears more spatially concentrated than for Glasgow and Edinburgh.

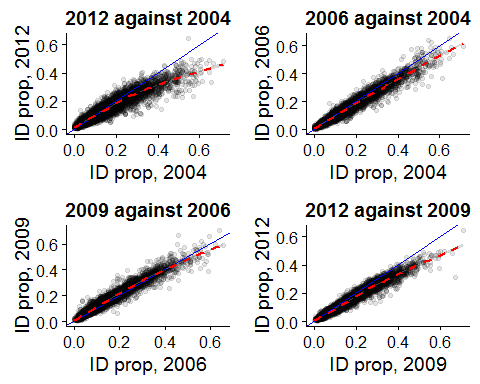
The equivalent map for Aberdeen, historically the most affluent and rural of the four cities, is shown here:



As with Glasgow and Edinburgh, it appears both that the greatest changes have occurred between 2009 and 2012, and also that any fall in RCI scores are likely due to persistence of poverty in some local areas, with a high income deprivation area clearly visible in all years to the north west of the city centre.

## Plotting temporal dependence

The following plot shows scatterplots of the ID proportions in each Scottish datazone in 2004, 2009 and 2012. Within these plots the blue line indicates parity between values on the x and y axis, and the red dashed line shows a nonlinear smoother function between the two ID proportions being plotted, and hence the level of deviation between the actual relationship over the two years and parity.



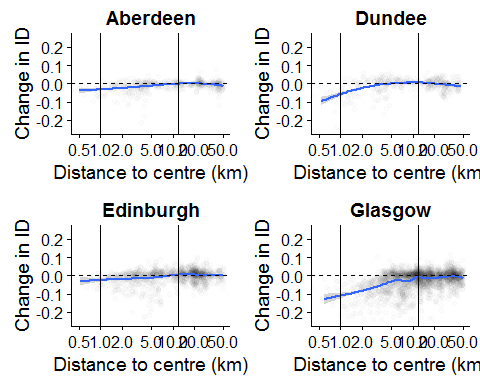
There were indications in the maps above that the greatest changes occurred between the 2009 and 2012 SIMD releases, suggesting that the temporal association between 2004 and 2009 ID proportions would likely be strong compared with the temporal association between 2009 and 2012.

The first of the plots, 2004 compared with 2012, shows strong spatial dependence in areas with low to moderate ID proportions, but much less temporal dependence in areas with high ID proportions in 2004. The second subfigure plots 2004 against 2006 values, and shows a near linear temporal dependence for the complete range of 2004 values, with a global trend towards lower ID proportions throughout Scotland. The temporal relationship between 2006 and 2009 starts to become nonlinear at the upper range of 2006 values, and this nonlinearity at the upper range becomes much stronger between 2009 and 2012.

These plots indicate a complex change in the temporal dependence of ID proportions, likely occurring over the period 2008 to 2012. SIMD 2016 values will be invaluable in indicating whether these changes have continued, altered or settled between 2012 and 2016.

## Plots of the distance/deprivation relationship

In the figure below, the vertical axis shows the proportion change in income deprivation from 2004 to 2012, and the horizontal axis shows the distance of datazones from the city centres, plotted on a log scale, for each of the four main Scottish cities. Individual datazones are indicated with faint circles, and a nonlinear smoother has been added showing how this the change in income deprivation varies as a function of distance. A dashed horizontal line has been added indicating no change in deprivation between the two period, and vertical lines added at 1km and 12km, as below 1km there are few observations, and above 12km the effect of a functional urban geography appears small.

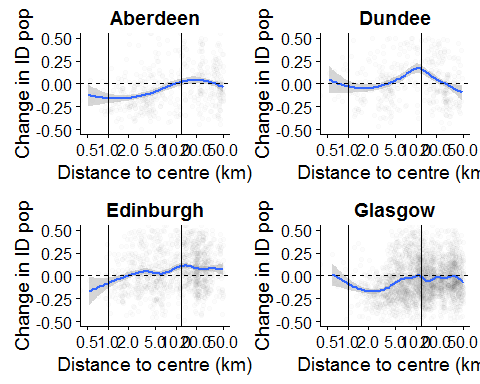


These plots suggest a qualitatively similar relationship between change in deprivation and distance, up to around 12km from city centre, for Aberdeen, Edinburgh and Glasgow, with Glasgow showing the strongest gradient of change between these two periods. Dundee shows a simpilar kind of relationship, but concentrated within the first 5km, rather than 12km, of distance from city centre. This is consistent with the maps of income deprivation in Dundee, showing much greater levels of spatial concentration than for the other cities.

### Numerator and denominator changes

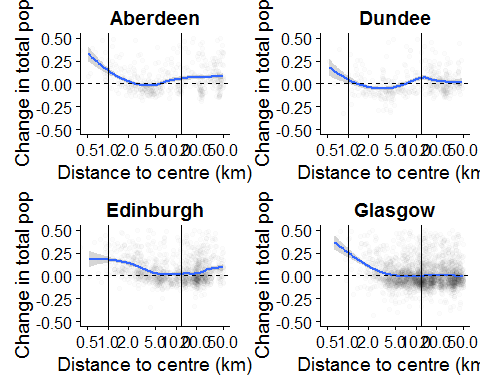
The falls in income deprivation nearer city centres are a function both of changing numerators - i.e. changes to income deprived population sizes - and changing denominators - i.e. changes to total population sizes. To gain a better understanding of the relative influence of these two changes to the relationship shown in the previous figure, the next two figures show the proportionate change from 2004 to 2012 in income deprived populations, and total populations.

The next figure shows how the income deprived population has changed between the two periods.



This plot indicates proportionate falls in the ID population shares at distances closes to the city centre, up to around 20% in Aberdeen and Glasgow, and either proportionate rises (Dundee and Edinburgh) or little change in proportions (Aberdeen and Glasgow) at greater distance from the city centre. This seems to suggest, for Aberdeen and Glasgow in particular, that 'gentrification of the centre' could be as much or more of an explanation for falls in RCI as 'impoverishment of the periphery', though that there are somewhat different socioeconomic mechanisms in place for the four cities.

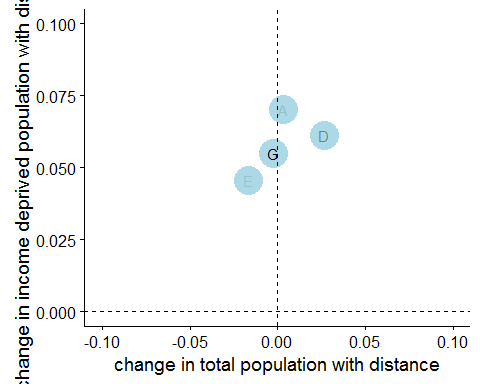
The following plot shows the change in the total population size as a function of distance from city centre. (The smoother line at distances less than 1km should likely be ignored for most cities as it involves extrapolation based on very few observations).



This plot suggests that there have been substantial increases in population sizes in Aberdeen, Edimburgh and Glasgow over this period, in datazones closer to the city centre, with the greatest changes occurring near the centre of Glasgow. In Aberdeen, and to a lesser extent for Dundee, there have also been increases in population shares living at distances of 12km and more from the city, suggesting greatest consistency in population sizes at intermediate distances.

The following plot attempts to summarise the previous series of results, in order the represent how both total population sizes and income deprived population sizes have changed with distance, and the overall strength of the relationship between income deprived population share and distance from city centre.

For each of the four cities, separate linear regressions were produced to extract the gradient of changes in total population size, income deprived population size, and changes in income deprived population shares with distance, over distances between 1km and 12km. The horizontal axis shows the gradient of change in total population with distance, the vertical axis shows graident of change in income deprived population with distance. Each city is represented by a separate point, labelled G for Glasgow, A for Aberdeen, D for Dundee and E for Edinburgh. The darkness of these symbols indicates the magnitude of the overall change in the income deprived population shares as a function of distance, with darker shades indicating a larger relationship and lighter shades a smaller relationship.



# Discussion

## Substantive Findings

The apparent fall in income deprivation proportions in more central areas of cities, in particular Glasgow, and in particular the fact the fall appeared greatest between 2008 and 2012, suggests there may be issues with the way in which income deprivation has been defined, and how effectively or otherwise this connects to broader and alternative definitions of poverty other this period. Although the definition of income deprived is not completely consistent from one SIMD release to the next, they tend to be based on claims for particular types of state benefit. During the 2008 to 2012 period the level of conditionality associated with many benefits increased for people of working age, while at the same time the longest sustained fall and plateau in real wages in over half a century occurred. Unemployment rates did not rise as they had in previous recessions, but levels of self-employment, short-term contracts, and part time work all increased. The self-employed, in particular, are not eligible for many of the benefits that people earning low wages as employees receive, and so the fall the income deprivation may reflect this compositional change in the workforce. Similarly increasing casualisation of employment, with variable hours and pay from one week to the next, may put off many people earning low wages on average from applying for 'top up' benefits such as tax credits. Again, indications that there have been falls in Scotland in income deprivation between 2008 to 2012 suggest statistical artefacts.

## Limitations and recommendations for further research

Alternative methods of reweighting SIMD values released at 2001 datazone levels could be explored, including those which make use of more sophisticated GIS-based methods to apportion values to 2011 datazones, and of approaches which use SIMD component values available at a higher level of spatial disaggregation (e.g. output areas) to recalculate and reapportion SIMD scores to 2011 datazone geographies.

The robustness of the results to the choice of city centre could be explored in a sensitivity analysis.

# Appendices

## Appendix A: Code for reweighting SIMD available for 2001 datazones onto 2011 datazones

Within the script below, the object lkup is the lookup file produced by Paul Norman, and simd\_combined contains data from the 2004, 2006, 2009, 2012 and 2016 releases of the SIMD.

lkup <- read\_csv("data/paul\_norman\_file/paul\_norman\_dz2011\_table.csv")  
simd\_combined <- read\_csv("data/simd/simd\_combined.csv")  
  
lkup

## Source: local data frame [59,492 x 7]  
##   
## oa\_2001 oa\_nrs dz\_2011 oa\_2011 dz\_2001 la\_2011  
## (chr) (chr) (chr) (chr) (chr) (chr)  
## 1 S00000002 60QA000002 S01006732 S00088957 S01000215 S12000033  
## 2 S00000002 60QA000002 S01006732 S00088958 S01000215 S12000033  
## 3 S00000002 60QA000002 S01006732 S00088956 S01000215 S12000033  
## 4 S00000002 60QA000002 S01006736 S00088957 S01000215 S12000033  
## 5 S00000002 60QA000002 S01006736 S00088958 S01000215 S12000033  
## 6 S00000002 60QA000002 S01006736 S00088956 S01000215 S12000033  
## 7 S00000002 60QA000002 S01006737 S00088957 S01000215 S12000033  
## 8 S00000002 60QA000002 S01006737 S00088958 S01000215 S12000033  
## 9 S00000002 60QA000002 S01006737 S00088956 S01000215 S12000033  
## 10 S00000003 60QA000003 S01006710 S00088959 S01000123 S12000033  
## .. ... ... ... ... ... ...  
## Variables not shown: la\_name (chr)

simd\_combined

## Source: local data frame [26,020 x 7]  
##   
## datazone year pop\_total pop\_workingage pop\_incomedeprived simd\_score  
## (chr) (int) (int) (int) (int) (dbl)  
## 1 S01000001 2004 864 575 62 6.706758  
## 2 S01000002 2004 687 505 14 4.187703  
## 3 S01000003 2004 838 568 24 4.829398  
## 4 S01000004 2004 641 460 13 5.345036  
## 5 S01000005 2004 663 470 21 4.636054  
## 6 S01000006 2004 626 346 83 11.475745  
## 7 S01000007 2004 857 464 38 4.003078  
## 8 S01000008 2004 584 358 65 11.374401  
## 9 S01000009 2004 710 453 40 9.133005  
## 10 S01000010 2004 568 410 15 6.649007  
## .. ... ... ... ... ... ...  
## Variables not shown: simd\_rank (int)

lkup %>%  
 select(dz\_2001, dz\_2011) %>%  
 arrange(dz\_2001, dz\_2011) %>%  
 group\_by(dz\_2001, dz\_2011) %>%  
 tally %>% # produce n, giving number of OAs which contain particular groupings of dz\_2001 and dz\_2011  
 arrange(dz\_2011, dz\_2001) %>%  
 select(dz\_2011, dz\_2001, n) %>%  
 group\_by(dz\_2011) %>%  
 arrange(dz\_2011, dz\_2001) %>%  
 mutate(weight = n / sum(n)) -> weighting  
weighting

## Source: local data frame [15,540 x 4]  
## Groups: dz\_2011 [6976]  
##   
## dz\_2011 dz\_2001 n weight  
## (chr) (chr) (int) (dbl)  
## 1 S01006506 S01000001 9 1.0000000  
## 2 S01006507 S01000005 1 0.1250000  
## 3 S01006507 S01000013 6 0.7500000  
## 4 S01006507 S01000024 1 0.1250000  
## 5 S01006508 S01000001 3 0.3333333  
## 6 S01006508 S01000005 6 0.6666667  
## 7 S01006509 S01000008 6 0.8571429  
## 8 S01006509 S01000013 1 0.1428571  
## 9 S01006510 S01000006 6 0.8571429  
## 10 S01006510 S01000008 1 0.1428571  
## .. ... ... ... ...

weighting %>% # Weighting by dz\_2011  
 left\_join(simd\_combined, by = c("dz\_2001" = "datazone")) %>%  
 select(-simd\_rank) %>% # Not meaningful to reweight rank  
 group\_by(dz\_2011, year) %>%  
 summarise\_each( ~ sum(. \* weight), 6:9) -> simd\_2011\_reweighted  
  
simd\_2011\_reweighted

## Source: local data frame [27,904 x 6]  
## Groups: dz\_2011 [?]  
##   
## dz\_2011 year pop\_total pop\_workingage pop\_incomedeprived simd\_score  
## (chr) (int) (dbl) (dbl) (dbl) (dbl)  
## 1 S01006506 2004 864.0000 575.000 62.00000 6.706758  
## 2 S01006506 2006 842.0000 566.000 69.00000 9.061800  
## 3 S01006506 2009 836.0000 554.000 60.00000 7.916251  
## 4 S01006506 2012 914.0000 612.000 75.00000 10.043655  
## 5 S01006507 2004 820.3750 507.000 43.87500 8.488053  
## 6 S01006507 2006 756.8750 471.500 39.62500 8.937893  
## 7 S01006507 2009 740.5000 450.750 65.00000 11.280132  
## 8 S01006507 2012 736.7500 466.375 50.62500 9.066177  
## 9 S01006508 2004 730.0000 505.000 34.66667 5.326289  
## 10 S01006508 2006 699.3333 494.000 37.66667 6.905884  
## .. ... ... ... ... ... ...