The impacts of neighbourhood context on attitudes to poverty and inequality

Constructing the neighbourhood database

Working Paper for the ESRC-funded study, "*The impact of neighbourhood context on attitudes to inequality and redistribution*" (RES-000-22-4192)

Nick Bailey
Urban Studies
University of Glasgow

nick.bailey@glasgow.ac.uk

July 2011

Contents

Tables and figures

1. Introduction

This note provides a record of the steps we went through to construct the neighbourhood database which forms the basis for this study.

The overall aim of the project is to examine how neighbourhood context shapes attitudes to poverty, inequality and redistribution. More specifically, it is interested in

- whether ***proximity to poverty/disadvantage or to affluence/advantage*** affect attitudes (RQ1);
- whether ***different kinds of poverty/disadvantage*** (e.g. child poverty, pensioner poverty, or poverty among minority ethnic groups) or ***different kinds of affluence/advantage*** impact more or less on attitudes (RQ2);
- whether the ***scale or patterning*** of poverty/affluence is important (RQ3); and
- whether the impacts of neighbourhood context are ***dependent on individual characteristics*** (RQ4).

The first three aims define the challenges for us in developing the neighbourhood database. In this paper, we consider first how we measure context in terms of the immediate neighbourhood (the LSOA) in which each person lives. We then go on to discuss how we measure the characteristics of the surrounding area, to cover the objectives related to scale and patterning.

2. Immediate neighbourhood

Approaches to measuring context

Broadly, two kinds of measure of neighbourhood characteristics are possible: measures of the *level* of a given characteristic and measures of *variation or mix*.

Levels

With levels, two approaches can be adopted. On the one hand, we can have *direct* or scalar measures of population averages (for continuous measures such as income) or of the proportions of people in different groups in each area (for categorical variables). On the other, we can construct *scaleless indices* which score neighbourhoods as having higher or lower levels on composite indicators of socio-economic status. The direct or scalar approach is the more appealing in many ways because the meaning of the measures is clear and obvious. Models would show the impact on attitudes of a rise of x per cent in the proportion of people in a particular group, for example. It is more demanding in data terms, however, since we do not collect small-area data on many characteristics of interest here: incomes or poverty, for example (Anderson 2007). With single indicators, there is also potentially a higher level of measurement error or 'noise' and the inclusion of large numbers of single indicators can give rise to problems of multicollinearity.

Scaleless indices are easier to implement since we can use a variety of indicators from sources such as the Census to construct measures which capture different aspects of the neighbourhood, including one or more indices capturing levels of socio-economic disadvantage. Various techniques exist to combine indicators but the obvious choice is some form of factor analysis. This also has the advantage of removing problems of multi-collinearity which occur with large numbers of individual indicators. One disadvantage of this approach is that the interpretation of the factor scores is not immediately obvious. We would need to find ways of understanding how any measure of socio-economic disadvantage related to measures or indicators of poverty. One means to do this would be to explore the relationships between the factor scores and direct estimates of the proportions poor or affluent from other sources.

Mix

With continuous measures such as income, measures of mix could include measures of dispersion or variance. With categorical measures where there are three or more groups, one approach to measuring mix or diversity is to combine these proportions into a single continuous variable such as an entropy index. The limitation of this is that the index does not record different forms of mix. An alternative is to use typologies (perhaps derived from cluster analysis) which identify different kinds of mix.

Levels - direct scalar measures

At the small area level, we have limited data on household incomes or wealth, or on the proportions with given levels of disadvantage or affluence. Some data is available from two sources, however, both associated with measuring poverty or deprivation. The first is data from the Indices of Multiple Deprivation 2007 (IMD 2007) which also has the advantage of being closer in time to the BSAS 2004 survey data than the 2001 Census. As well as an aggregate, scaleless measure of area deprivation, the IMD produces estimates of the proportion of people on low income benefits, derived from administrative sources (the Income Deprivation domain score) (Noble et al 2006). From 2007, it has also been producing estimates of the proportion of children and of older persons in households on low-income benefits. Indirectly, it therefore also produces counts of the number of working-age adults in such households. Unfortunately, the latter does not distinguish adults with children from those without.

A second source would be estimates of the numbers in poverty, derived from a combination of survey data on individuals and area data from the Census. The most recent example of such work is provided by Fahmy et al (2011); earlier studies include Anderson (2007). Fahmy et al use the *Poverty and Social Exclusion Survey 1999* (PSE 1999) to identify the set of Census variables that most accurately predicts the odds of being in the 'Breadline Poor' and 'Core Poor' categories. The paper provides full details for these terms, but in brief:

- 'Breadline Poor' households are 'deprivation poor' (lacking two or more items regarded as 'necessities' by the general public) and on low income (low PSE-equivalised household income); and
- 'Core Poor' households are 'deprivation poor', have an equivalised income less than 70% of the median, and report themselves as subjectively poor `sometimes' or `all the time'.

The former covered 28 per cent of the British population, while the latter covered 12 per cent (1999 figures). We can use the weights Fahmy et al provide to calculate both poverty rates for LSOAs (Table 1). Fahmy et al did not try to estimate the numbers in different poverty sub-groups since the PSE survey has a relatively small sample, making estimates for sub-groups increasingly unreliable.

Table 1: Variables and weights for 'Breadline Poor' and 'Core Poor'

Variable	**Census table/cells**	**Weight**	
		Breadline poor	**Core poor**
Unemployed (hhlds)	CS0130005	0.211	0.074
Lone parent (hhlds)	KS0200011	0.271	0.101
LLTI (people)	CS0160002	0.161	0.067
No car hhlds (people)	CS0220010	0.164	0.027
Social renting (hhlds)	KS0180005+KS0180006	0.286	0.098
Private renting (hhlds)	KS0180007	0.130	0.071
Overoccupancy hhlds (people)	UV0830004+UV0830005	0.435	0.038
NS-Sec 6-8 (HRP, 16-74)	CS0460031+CS0460036+CS0460041	0.072	0.165
No CH/shared amenities (hhlds)	CS0550011+CS0550086	0.109	0.042
Denominator (all hhlds)	UV0630001		

Source: Fahmy et al (2011).

Other indicators of advantage or disadvantage are available from Census 2001 small area statistics. These provide basic measures of the socio-economic characteristics of the population, including occupational status (NS-Sec), employment status, and housing tenure.

In addition, the Census also permits us to identify the presence of different social groups, e.g. by age, household type, or ethnicity. We cannot identify the presence of deprivation or affluence within those groups directly, but we can identify where neighbourhoods have both concentrations of deprivation or affluence and concentrations of particular age groups, household types or ethnic groups. We can test whether the interaction between these measures has any significant impact on attitudes. We do need to be aware of a possible 'ecological fallacy': neighbourhoods with concentrations of poverty and of elderly households may not have poor elderly households. Nevertheless, this is a reasonably convincing approach to take.

In summary, we have a number of direct measures or estimates of advantage or disadvantage rates for LSOAs:

- the proportion of all people, children, working age adults and pensioners in households on low income benefits (IMD 2007);
- estimates of 'breadline poor' and 'core poor' (derived from Census data);
- (also from Census data) the proportions in the following socio-economic groups:
    - NS-Sec (two groups – professional/managerial and routine manual/never worked);
    - employment status (three groups – unemployed, inactive and students); and
    - housing tenure (two groups – social renters and private renters).

In addition, we would add further Census variables to measure the presence of particular socio-demographic groups. These variables are:

- percent of different household types (with children; lone parent; working age adults only; and pensioners only);
- percent of population from different ethnic groups (Asian; Black; Other);
- percent of population 16-74 who are full-time students.

Levels - scaleless measures

Approach to factor analysis

Exploratory factor analysis is a rather subjective technique in which a great many choices have to be made, all of which can influence results. Guidance on the technique stresses the importance of making these choices explicit and of examining the extent to which they shape the final analysis.

Number of cases

There are various cautions against conducing factor analysis with too few cases, but these tend to be directed to studies with a few hundred cases. We have over 32,000 LSOAs, so far exceed any recommended minimum thresholds. In our analyses, the Kaiser-Meyer-Olkin test of sample adequacy is always close to 0.9, compared with a minimum recommended value of around 0.6 (and a maximum value of 1.0).

Selection of variables

Variables should be selected with a clear intent, linked to theory and building on previous studies. For us, the selection is made quite straightforward due to the fact that the theories or hypotheses we wish to test fit well with what is known about urban structures. Timms (1971) summarised a wide range of studies on residential differentiation, arguing that four dimensions emerge almost universally: socio-economic status; demographics or family type; ethnic composition; and mobility. All of these would appear to be useful for our study. The first relates to the core interest in levels of poverty or affluence. The second and third help to identify different forms of poverty or affluence which we suspect may have different influences on attitudes. The last is not explicit in our previous theorising but might still be worth capturing: a highly mobile population might have different impacts on attitudes.

Thirty years later, Johnston et al (2004) use 18 variables from the 2001 Census, and identify five consistent factors which replicate this pattern almost exactly; the one addition is a factor that picks up ‘rurality’ (through a combination of individuals working in agriculture and poor housing quality); many of the studies that Timms reviews were of single cities so a rural dimension might not be present. The last of these would also be a useful aspect to capture, since it relates to density and hence to physical proximity.

We therefore began with a set of variables designed to cover these five main dimensions (Table 2). Data was extracted from the Census 2001, the Indices of Multiple Deprivation 2007 and the General Land Use Database 2005. Variables should not be too highly correlated (greater than 0.9 according to Field 2005) nor should they be linear combinations of each other. Both of these issues can affect the determinant, bringing it close to zero and making factor solutions impossible. Initial examination of correlations eliminated some variables (e.g. employment, health and education domain scores from the IMD) as these were very highly correlated with others which were retained. The initial set is listed by the domain they fit most obviously, although we expect loadings for factors to cross some of these boundaries.

Table 2: Initial variable set

Dimension	**Variables**	**Source**
Socio-economic status		
Occupational group (NSSec)	Professional/managerial (NSSEC1) Routine occupations/never worked (NSSEC5)	Census 2001
Employment status	Unemployed Inactive Students	Census 2001
Educational attainment	Level 0/1 qualifications	Census 2001
Housing tenure	Social rent Private rent	Census 2001
Health	LLTI	Census 2001
Overoccupancy of housing	Overoccupancy (occupancy rating -1 or lower)	Census 2001
Deprivation domain scores	Income deprivation Crime deprivation Housing deprivation Environment deprivation	IMD 2007
Demographics		
Age	Age 18-24; 25-29; 30-39; 40-49	Census 2001
Household type	Single pensioner Couple pensioner Couple + dependent children Lone parent + dependent children Other + dependent children All adult	Census 2001
Gender	Female	Census 2001
Ethnicity		
Ethnic composition	Asian Black/Mixed/Other	Census 2001
Residential mobility		
Gross turnover	Gross turnover (in-migrants + out-migrants + within-area migrants)	Census 2001
Rural/urban		
Population density	Density (persons per hectare)	Census 2001
House type	Semi-detached Terraced Flats	Census 2001
Land use	Area covered by domestic properties Area covered by greenspace	General Land Use Database 2005

Checking or screening of variables
Although factor analysis does not make specific assumptions about the distribution of variables, the presence of variables with high levels of skew and/or kurtosis can lead to 'artefactual factors' or to factors that load heavily on a single variable (Bandalos and Finney 2010). Bandalos and Finney (2010) recommend keeping absolute skewness below 2.0 and kurtosis below 2.0, although they note that others are content with kurtosis below 7.0. We have a number of variables with high kurtosis and some with high absolute levels of skew. We apply a range of transformation (natural logs, square root, square and cubic) to bring all values of skewness and kurtosis within or very close to the broader limits (Table 3).

Table 3: Descriptives of transformed variables

	Mean	Std. Deviation	Skewness	Kurtosis
% routine/never worked	17.45	9.80	1.00	.68
% unemployed	3.39	2.06	1.38	1.97
% inactive	33.20	8.25	.61	1.03
% students (log)	1.78	.49	1.62	4.41
% no quals/level 1	49.58	14.56	-.30	-.34
% Social Rent	17.98	18.80	1.32	.99
% Private Rent (sqrt)	2.77	1.24	1.22	1.73
% LLTI	17.97	5.50	.59	.56
% overoccupying (sqrt)	6.01	2.95	1.17	1.53
IMD Income Depvn	15.62	12.18	1.33	1.49
IMD Crime Depvn	.00	.83	.01	-.19
IMD Housing Depvn	21.69	11.05	.63	.04
IMD Envt Depvn	21.69	16.87	1.09	.63
% age 18-24 (log)	2.02	.39	1.49	5.19
% age 25-29	6.66	3.17	1.75	5.12
% age 30-39	15.62	3.65	.74	1.41
% age 40-49	13.39	2.29	-.06	1.13
% single pensioner	6.42	3.14	1.43	4.38
% couple pensioner	8.28	4.22	1.08	2.95
% cple+kids	36.81	8.28	-.21	1.08
% lone parent	8.02	5.34	1.37	1.84
% other+kids (sqrt)	2.07	.70	1.39	3.28
% all adult (log)	2.50	.31	-.42	4.21
% female (cube)	136045	16075	-.14	7.36
% Asian (sqrt)	1.45	1.55	2.19	5.59
% Bl/Ch/Mx/Ot	1.71	1.25	1.84	3.63
Gross turnover % (log)	1.30	.17	.85	1.60
Popln density (log)	1.30	.66	-1.22	.89
% semi	34.44	20.66	.55	-.32
% terraced	26.69	21.69	.88	-.07
% flat	13.51	19.03	2.33	5.23
% Area Domestic Bldngs	8.86	6.34	.76	.85
% Area Greenspace	43.03	30.39	.44	-1.15

Extraction method
There are broadly two related approaches covered by the term factor analysis: exploratory factors analysis (EFA) and principal components analysis (PCA). There is some debate about which is preferable and under which conditions although it has also been suggested that, most of the time, the choice makes little difference (Costello and Osbourne 2005). The latter is seen as more appropriate where the aim is simply data reduction, collapsing a large number of variables into a smaller number of groups. The former is seen as more appropriate where the aim is to identify latent variables which cannot be identified directly. PCA is the default in SPSS but this may be for historic reasons since it is computationally less intensive.

Overall, factor analysis would appear the more suitable extraction method for us. We start from a theory about the (latent) variables which drive urban or neighbourhood structures, and we would therefore expect some, but not all, of the variance in the indicators to be driven by these factors. EFA works only on this shared or common variance, whereas PCA works with all the variance.

There are several factor extraction methods and no clear means to choose between them. Maximum likelihood has some advantages but is most suitable where variables are near-normally distributed (which ours are not). Principal axis factoring is recommended by Costello and Osbourne (2005) for cases where data are non-normal. Since we have chosen more relaxed criteria for skewness and kurtosis, we use principal axis factoring.

Number of factors retained
The standard choice is based on the Kaiser criterion (any factor with an eigenvalue greater than 1.0) although there is also a consensus that this is the least desirable criteria. Of the easily implemented alternative strategies, the scree test is recommended but it is also a question of examining factor loadings, looking for a solution with minimal cross-loadings. We explore a range of solutions with factor numbers around the level indicated by both criteria.

Rotation
The factors can be rotated to produce more easily interpreted factor loadings. It also possible to extract solutions where factors are oblique rather than orthogonal (i.e. where there is some correlation between factors). Oblique solutions are seen by many as providing solutions which reflect the ‘real world’ better (Costello and Osbourne 2005). That certainly seems appropriate here. For example, it would be difficult to imagine that the distribution of minority ethnic groups in the UK would be wholly unrelated to the distribution of socio-economic disadvantage, given what we know about discrimination in labour and housing markets. In SPSS, oblimin rotation with the default value of delta is recommended, and that is what we use here.

Results of factor analysis
A wide variety of analyses were performed, using different combinations of variables and numbers of factors extracted. The scree test tends to suggest five or six factors would be optimal, while the Kaiser criterion tends to suggest five to seven. A fairly consistent picture emerges in line with previous research, although precise details vary. Two solutions are presented below for comparison – a five and a six factor solution (Tables 4 to 6). (One variable is dropped for the five factor solution as the analysis failed to converge initially.)

As previous studies have found, one consistent factor identifies socio-economic disadvantage, loading as expected on a range of variables measuring socio-economic status (in particular, low occupational status, unemployment, inactivity, low educational attainment, social rented housing, poor health and lone parent households). This is almost always the first factor extracted, and accounts for the largest share of the variance. The IMD Income Deprivation variable loads heavily on this factor when included but it is also quite highly correlated with several other variables. Since this factor has a number of other variables with heavy loadings, the Income Deprivation score is not critical. It makes sense to remove this variable and to use it as a means of understanding or interpreting different levels on this factor. It is therefore omitted from both analyses presented here.

Another consistent factor captures the urban-rural dimension, with positive values indicating more rural, less dense areas. This tends to be the fourth or fifth factor to emerge so it accounts for a relative small proportion of the variance. It loads on the obvious variables of population density and various measures of land-use (the proportion of land given to domestic buildings or to greenspace). It also has a modest loading on housing deprivation (a feature of more rural areas) if that variable is included, as well as negative loadings on minority ethnic populations and young adults.

Other variables capturing aspects of demography, ethnicity and mobility appear in slightly variable combinations although the underlying picture is consistent. Factors identify dimensions covering: older people; minority ethnic groups; and mobile young adults. With more factors, a factor covering settled younger adults emerges as well. Students can appear in slightly different factors. In the six factor solution, for example, they are seen as co-occurring with mobile young people (18-29) and with minority ethnic groups, and they have a negative loading on the factor identifying settled young adults. In the five factor solution, the co-occurrence with minority ethnic groups appears to be the strongest relationship. Neither solution separates minority ethnic households entirely from students.

Inevitably, the choice of a final factor solution is the result of a great many judgements, many quite subjective. Our preferred one is the five factor solution shown in Table 4 since it appears to provide a clearer separation of the groups. In this solution, the factor correlation matrix shows some modest correlations: negative relationships between rurality and both socio-economic disadvantage and mobile young people (18-29); note that, in the latter case, we invert the values. There are also slightly weaker correlations between the factor identifying minority ethnic groups and students and both rurality (negative) and mobile young people 18-29 (positive).

Table 4: Summary of two factor analyses

Six factor solution			Five factor solution	
No.	***Label***	*(Correlations)*	***No.***	***Label***
1	Socio-economic disadvantage	(.999)	1	Socio-economic disadvantage
2	~Settled YP 25-39 (not students)	(-.688)	2	Older people (not families, nor YP 25-39)
3	~Older people (not families)	(.845)		
4	Minority ethnic groups & students	(.751)	4	Minority ethnic groups & students (not OP)
6	~Mobile YP 18-29 & students	(.934)	3	~Mobile YP 18-29 (not families)
5	Rural, low density	(.964)	5	Rural, low density

Notes: YP – Young People; OP – Older People. The symbol '~' indicates that the factor score which is extracted has to be inverted to fit with this label. The loadings in the matrices below therefore have the opposite sign to what might be expected. Correlations are shown between the factors extracted by each analysis, with comparisons only for the most similar factors.

Table 5: Five factor solution

Communalities / **Pattern Matrix**

	Extraction	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5
% routine/never worked	.891	.931				
% unemployed	.781	.785				
% inactive	.864	.486	.653		.371	
% students (log)	.724				.761	
% no quals/level 1	.789	.820		.368		
% Social Rent	.681	.805				
% Private Rent (sqrt)	.704			-.802		
% LLTI	.823	.550	.597			
IMD Crime Depvn	.483	.473				
% age 18-24 (log)	.639			-.473	.417	
% age 25-29	.807		-.390	-.659		
% age 30-39	.760		-.742	-.320		
% age 40-49	.519		-.349	.447		
% single pensioner	.674		.714		-.309	
% couple pensioner	.840	-.378	.724			
% cple+kids	.732		-.561	.561		
% lone parent	.740	.816				
% other+kids (sqrt)	.623	.334			.627	
% Asian (sqrt)	.592				.668	
% Bl/Ch/Mx/Ot	.571				.460	
Gross turnover % (log)	.752			-.863		
Popln density (log)	.931					-.974
% Area Greenspace	.792					.917

Factor Correlation Matrix

Factor	1	2	3	4	5
1	1.00	.16	-.15	.15	-.31
2	.16	1.00	.05	-.13	.09
3	-.15	.05	1.00	-.26	.36
4	.15	-.13	-.26	1.00	-.26
5	-.31	.09	.36	-.26	1.00

Table 6: Six factor solution

Communalities / **Pattern Matrix**

	Extraction	Factor 1	2	3	4	5	6
% routine/never worked	.940	.944					
% unemployed	.785	.774					
% inactive	.858	.472	.593	-.310			
% students (log)	.735		.440		.318		-.492
% no quals/level 1	.817	.832			-.309		
% Social Rent	.722	.795					
% Private Rent (sqrt)	.696						-.773
% LLTI	.818	.529		-.549			
IMD Crime Depvn	.488	.470					
IMD Housing Depvn	.473				.653	.357	
% age 18-24 (log)	.795						-.783
% age 25-29	.802		-.495				-.536
% age 30-39	.812		-.758				
% age 40-49	.518			.407			.399
% single pensioner	.709			-.826			
% couple pensioner	.850	-.395	.315	-.563			.369
% cple+kids	.777			.734			.327
% lone parent	.745	.807					
% other+kids (sqrt)	.655	.336			.529		
% Asian (sqrt)	.591				.478	-.352	
% Bl/Ch/Mx/Ot	.849				.781	-.332	
Gross turnover % (log)	.800						-.902
Popln density (log)	.992					-.999	
% Area Greenspace	.769					.893	

Factor Correlation Matrix

Factor	1	2	3	4	5	6
1	1.00	.10	-.17	.12	-.29	-.15
2	.10	1.00	-.17	-.01	.01	.11
3	-.17	-.17	1.00	.14	-.07	.03
4	.12	-.01	.14	1.00	-.19	-.41
5	-.29	.01	-.07	-.19	1.00	.35
6	-.15	.11	.03	-.41	.35	1.00

Mix

Within the immediate neighbourhood, we can use the proportions of the population in each sub-group to create two kinds of measures of mix: entropy scores and cluster typologies. Entropy scores provide a single continuous measure of the extent to which different groups are equally present in a given area. Cluster typologies reflect both the level of diversity and the nature of diversity.

Entropy scores

A major review of measures of multigroup segregation by Reardon and Firebaugh (2002) recommended Theil's Information Theory index, H, as the best single measure. For the system of neighbourhoods as a whole, this is defined by:

$$H = \sum_{m=1}^{M} \sum_{j=1}^{J} \frac{t_j}{TE} \pi_{jm} \ln \frac{\pi_{jm}}{\pi_m} \qquad (1)$$

Where:

t_j = total number of individuals in neighbourhood j
T = total number of individuals (in all neighbourhoods)
π_m = proportion in group m (across all neighbourhoods)
π_{jm} = proportion in neighbourhood j in group m

And where E is a constant, given by:

$$E = \sum_{m=1}^{M} \pi_m \ln\left(\frac{1}{\pi_m}\right) \qquad (2)$$

The overall measure, H, is the population weighted sum (indicated by the term t_j/T in equation 1) of neighbourhood-level measures of mix. The entropy index, H, can range from zero (all neighbourhoods have the same proportion of each group as the national population) through to 1 (each neighbourhood has only one group present i.e. groups are wholly segregated). The neighbourhood contribution is given by:

$$H_j = \sum_{m=1}^{M} \frac{1}{E} \pi_{jm} \ln \frac{\pi_{jm}}{\pi_m} \qquad (3)$$

Hj has a minimum value of zero where the neighbourhood has the same population composition at the national average, but an upper limit in excess of 1. Scores greater than 1 occur where neighbourhoods have very large concentrations of groups with low prevalence nationally.

In situations where one group is quite dominant nationally (e.g. with the White majority ethnic group in the UK), neighbourhoods dominated by the majority group never differ that much from the national average and can never have high entropy scores. High entropy scores are therefore confined to places with high concentrations of minority groups. In other words, a neighbourhood composed solely of the White majority group would have a much lower entropy score than one composed solely of a minority ethnic group. For our work, this makes little sense: the former may be much more common but both are equally ‘unmixed’.

We therefore use an alternate measure of entropy where each neighbourhood is measured against the situation where groups are equally prevalent:

$$H_j = \frac{1}{\ln(M)} \sum_{m=1}^{M} \pi_{jm} \ln \frac{1}{\pi_{jm}} \qquad (4)$$

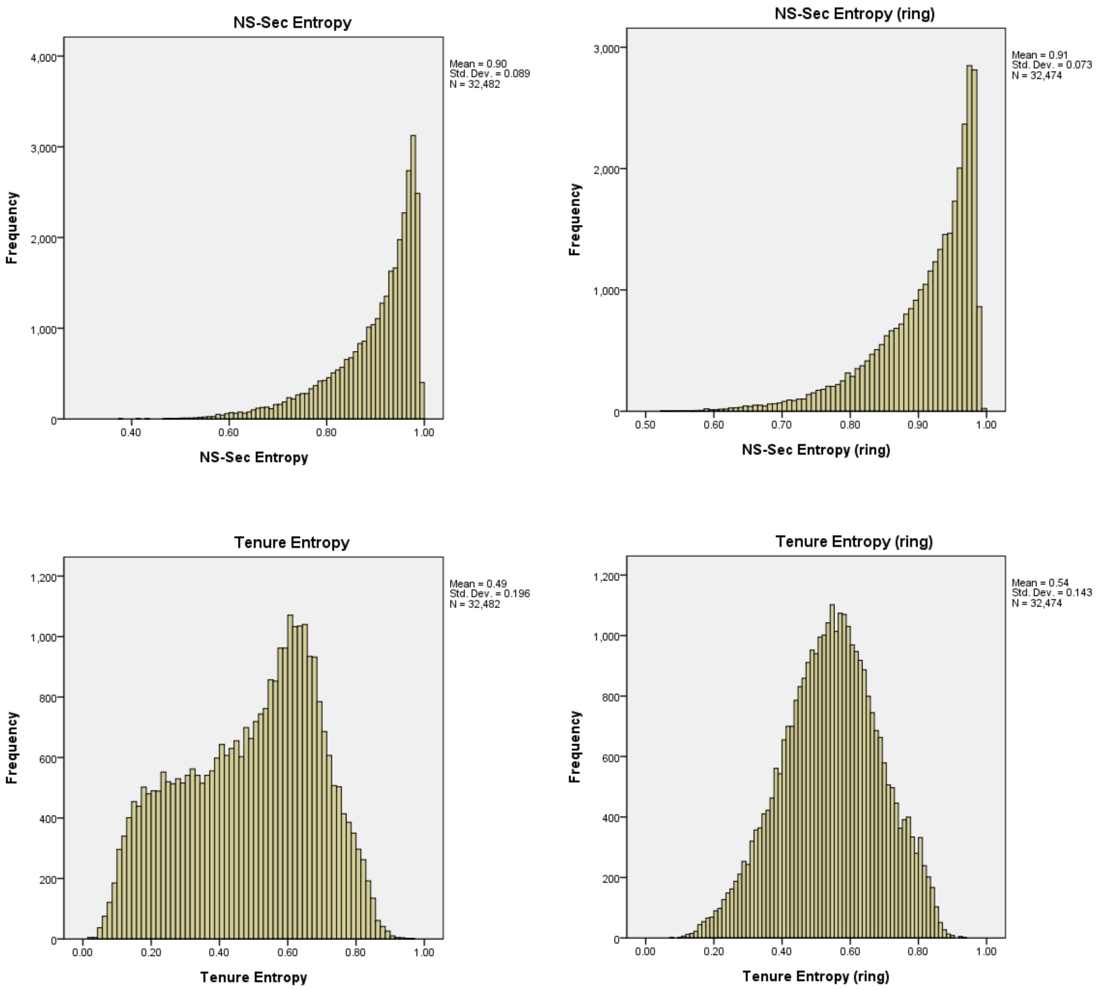
These measures are calculated for the variables which are of most central interest analytically: NS-Sec and tenure to reflect socio-economic status; and household type and ethnicity to reflect socio-demographic differences. Descriptive statistics are shown in Table 7 and histograms in Figure 1; statistics and figures are also shown for the entropy scores for the surrounding rings, discussed below.

Neighbourhoods are most mixed in relation to NS-Sec, and least mixed in relation to ethnicity. Mix in relation to household types is also quite high, with tenure showing the greatest spread. Tenure mix has the highest correlation with measures of deprivation but, contrary to much policy rhetoric, mixed neighbourhoods tend to be more deprived. Neighbourhoods with higher ethnic mix (implicityly, those with larger proportions of minority ethnic groups) also tend to be more deprived as do those with more mix in relation to NS-Sec. This reflects the fact that the largest group in each case (Whites and those in professional or managerial occupations) are also economically better off.

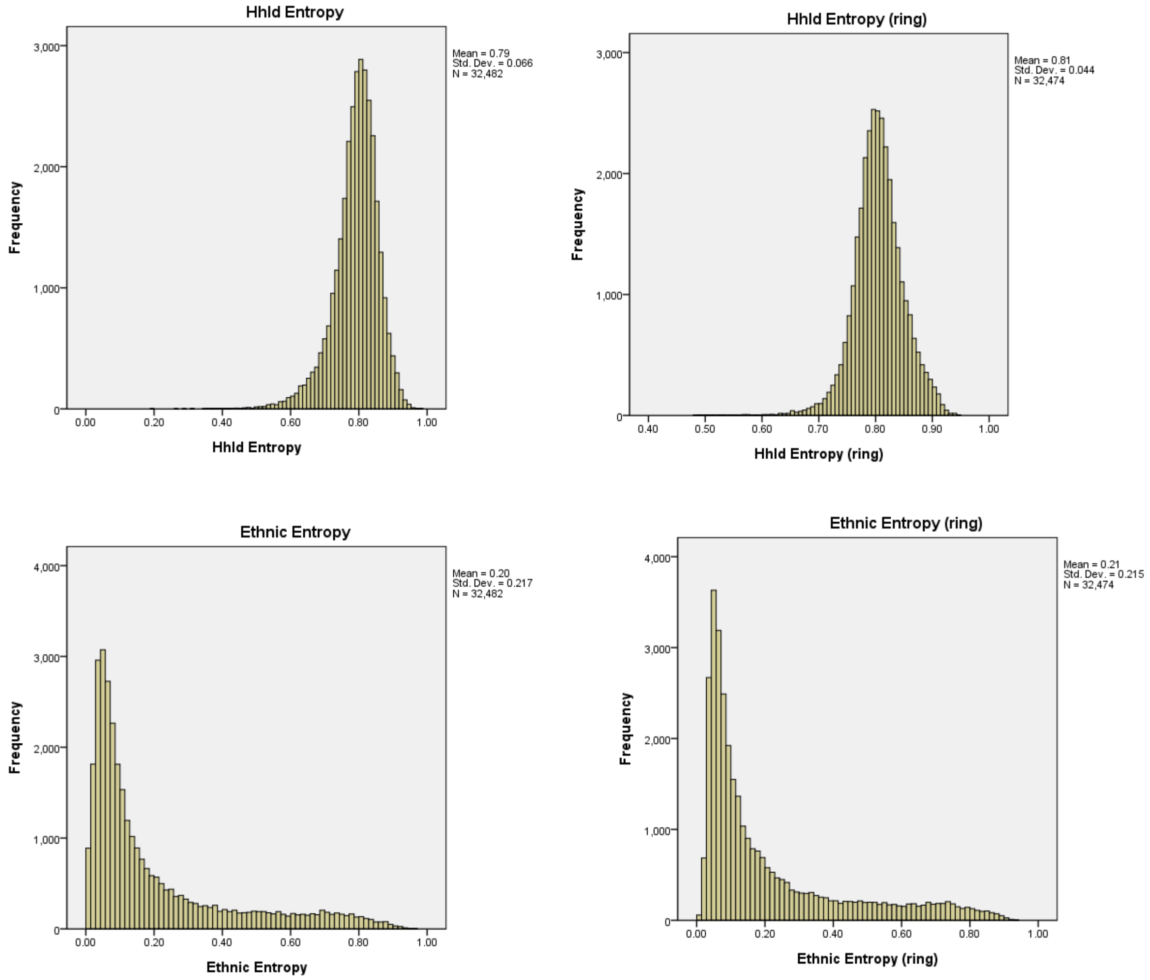
Table 7: Summary statistics for Entropy scores – immediate neighbourhood and surrounding ring

	N	Minimum	Maximum	Mean	Std. Deviation	Skewness		Kurtosis	
	Statistic	Statistic	Statistic	Statistic	Statistic	Statistic	Std. Error	Statistic	Std. Error
NS-Sec Entropy	32482	.38	1.00	.8981	.08905	-1.434	.014	1.914	.027
NS-Sec Entropy (ring)	32474	.52	1.00	.9100	.07334	-1.441	.014	2.170	.027
Tenure Entropy	32482	.02	.96	.4883	.19621	-.275	.014	-.917	.027
Tenure Entropy (ring)	32474	.08	.93	.5438	.14336	-.147	.014	-.326	.027
Hhld Entropy	32482	.20	.98	.7913	.06569	-1.113	.014	3.202	.027
Hhld Entropy (ring)	32474	.48	.95	.8053	.04391	-.474	.014	2.803	.027
Ethnic Entropy	32482	.00	.97	.2022	.21668	1.524	.014	1.314	.027
Ethnic Entropy (ring)	32474	.00	.94	.2115	.21510	1.486	.014	1.163	.027

Figure 1: Histograms for entropy scores – immediate neighbourhood and surrounding ring



[Figure 1 – continued]



Cluster typologies

Entropy scores measure the degree of mixing but not the nature of the mix: the all-White and the all-Asian neighbourhoods have the same entropy score, for example. Cluster typologies were therefore formed as well to capture the nature of the mix for the same four variables. In each case, five clusters appeared to give a reasonable level of detail without any one type becoming too small.

For each variable, the same four or five groups used to estimate entropy scores are used as the basis of the clusters. The variables and the final cluster centres are shown in Table 9 below along with the mean entropy score in each case; a separate cluster analysis was conducted for the set of “surrounding rings” as well as discussed below.

Short labels are used to describe each cluster (Table 8). With ethnicity and tenure, the largest group forms the first part of the label, with other groups accounting for more than 20 per cent of the total shown in brackets. With NS-Sec, the first group (professional/managerial workers) is so large that it forms the largest group in every cluster except the fifth. Since these clusters form a clear hierarchy with declining proportions of professional/managerial workers and increasing proportions of routine manual /never worked, they are simply labelled "NS 1" to "NS 5". With household type, households with children dominate in a similar manner but there is no obvious hierarchy here. In this case, labels indicate which groups are most over-represented compared with the national average. For example, in the "WA (Ot)" type of area, there is an over-representation of households composed solely of working-age adults (the largest group) and an over-representation of other household types. The second largest group is households with children but they are found less commonly in this type of area than in any other and so are not mentioned in the label. The "Ch (WA)" type does not quite fit this rule but, given how important households with children are in this kind of area, it made sense to mention them in the label.

Table 8: Cluster labels

Hhld cluster	
Ch 1	Hhlds with children 1
Ch 2	Hhlds with children 2
Ch (WA)	Hhlds with children (Working-age adult hhlds)
WA (Ot)	Working-age adult hhlds (and others)
WA (OP)	Working-age adult hhlds (Hhlds with older people)
Ethnic cluster	
Wh 1	White 1
Wh 2	White 2
As (Wh)	Asian (White)
Wh (As)	White (Asian)
Wh (Bl)	White (Black)
Tenure cluster	
OO 1	Owner-occupier 1
OO 2	Owner-occupier 2
OO (SR)	Owner-occupier (Social renter)
SR (OO)	Social renter (Owner-occupier)
OO (PR)	Owner-occupier (Private renter)
NS-Sec cluster	
NS 1	NS-Sec 1 (most advantaged)
NS 2	NS-Sec 2
NS 3	NS-Sec 3
NS 4	NS-Sec 4
NS 5	NS-Sec 5 (most disadvantaged)

Table 9: Cluster centres – immediate neighbourhood and surrounding ring

Immediate neighbourhood

Hhld cluster	***Ch 1***	***Ch 2***	***Ch (WA)***	***WA (Ot)***	***WA (OP)***	***Total***
% pensioner hhld	7	12	15	10	25	14
% WA adult only	25	32	36	46	35	34
% hhlds with children	63	52	45	30	36	47
% other hhlds	4	4	4	14	3	5
Hhld entropy	0.68	0.77	0.81	0.85	0.85	0.79
Number	3506	11105	12100	1930	3841	32482
Ethnic cluster	***Wh 1***	***Wh 2***	***As (Wh)***	***Wh (As)***	***Wh (Bl)***	***Total***
% White	97	81	27	56	57	91
% Asian	1	9	62	31	9	5
% Black	0	5	8	8	26	2
% Mixed/Chinese/Other	1	5	4	5	8	2
Ethnic entropy	0.10	0.46	0.66	0.72	0.76	0.20
Number	25540	4066	651	1055	1170	32482
Tenure cluster	***OO 1***	***OO 2***	***OO (SR)***	***SR (OO)***	***OO (PR)***	***Total***
% Own	90	74	55	29	48	71
% Social Rent	4	13	36	61	15	18
% Private Rent	5	11	7	7	35	9
% Rent Free	1	2	2	3	2	2
Tenure entropy	0.28	0.55	0.67	0.65	0.74	0.49
Number	12498	9366	5367	3119	2132	32482

Surrounding ring

Hhld cluster	***Ch 1***	***Ch 2***	***Ch (WA)***	***WA (Ot)***	***WA (OP)***	***Total***
% pensioner hhld	8	12	15	10	22	14
% WA adult only	26	33	35	41	34	34
% hhlds with children	59	51	46	35	40	47
% other hhlds	6	4	3	13	4	5
Hhld entropy	0.74	0.78	0.81	0.88	0.85	0.81
Number	2117	10630	13957	2326	3444	32474
Ethnic cluster	***Wh 1***	***Wh 2***	***As (Wh)***	***Wh (As)***	***Wh (Bl)***	***Total***
% White	97	83	32	60	60	91
% Asian	1	9	54	29	9	4
% Black	0	4	9	7	23	2
% Mixed/Chinese/Other	1	4	5	5	8	2
Ethnic entropy	0.11	0.44	0.73	0.69	0.74	0.21
Number	24934	4442	635	1113	1350	32474
Tenure cluster	***OO 1***	***OO 2***	***OO (SR)***	***SR (OO)***	***OO (PR)***	***Total***
% Own	85	73	60	34	53	72
% Social Rent	7	16	31	50	17	17
% Private Rent	7	9	7	13	28	9
% Rent Free	1	2	2	2	2	2
Tenure entropy	0.39	0.57	0.66	0.74	0.75	0.54
Number	11626	11434	5094	2081	2239	32474

[Table 9 – continued]

	Immediate neighbourhood						Surrounding ring					
NS-Sec cluster	***NS 1***	***NS 2***	***NS 3***	***NS 4***	***NS 5***	***Total***	***NS 1***	***NS 2***	***NS 3***	***NS 4***	***NS 5***	***Total***
% manager/prof	59	45	34	25	16	36	57	45	36	28	20	36
% intermed/small emp	21	25	24	20	15	22	21	24	24	21	16	22
% supervisory/technical	5	8	11	12	11	9	5	8	10	11	11	9
% semiroutine	8	12	16	20	22	16	8	12	16	19	21	15
% routine/never worked	7	10	15	24	36	17	9	11	15	22	32	17
NS-Sec entropy	0.72	0.86	0.94	0.98	0.93	0.90	0.75	0.86	0.94	0.97	0.96	0.91
Number	4147	8227	8906	7016	4186	32482	3188	8117	9497	7967	3705	32474

Notes: The following abbreviations are used in cluster names:
Households: Ch – Households with Children; WA – Working-age Adults only; OP – Older People only; Ot – Other household types.
Ethnicity: Wh – White; As – Asian; Bl – Black.
Tenure: OO – Owner-occupier; SR – Social Renter; PR – Private Renter.

Summary

For the immediate neighbourhood, we therefore have the following set of 33 variables:

- six direct measures or estimates of population in ‘poverty’ (four from the IMD and two estimates from Fahmy et al’s work);
- seven measures for socio-economic status (covering NS-Sec, employment status and tenure);
- seven measures of demographic groups (covering household type and ethnicity);
- four entropy scores (NS-Sec, employment status, tenure, household type and ethnicity);
- four cluster typologies for the same dimensions; and
- five factor scores.

3. Surrounding areas

Defining the surrounding area

We have said that we will work with neighbourhoods defined at two scales: the LSOA (c. 1500 population); and the ring of adjacent LSOAs. This follows broadly the work of Suttles (1972) who argued that neighbourhoods have multiple levels or meanings for individuals, but also stems from the specific theories we have about the mechanisms by which attitudes may be shaped.

Where LSOAs border Wales or Scotland, the adjacent LSOAs would be outside England. As we have not extracted data for these areas, we therefore exclude them from the analysis (69 LSOAs or 0.2 per cent of the total for England). The Isles of Scilly LSOA is also excluded. This gives a set of 32,412 LSOAs to work with.

We used GIS to identify which LSOAs touch or neighbour each other. In the initial file, some LSOAs were regarded as adjacent to others on the opposite side of major rivers such as the Thames or the Mersey due to the way that LSOA boundaries are drawn to cover the whole territory of the UK. This does not seem appropriate in our work, since the rivers act as major barriers to interaction and observation. We removed a significant number of such connections (around 600 or 0.3 per cent of all adjacents). Of course, other major barriers to connection remain such as motorways, other major roads or railway lines, inland water etc.. We do not attempt to take those into account at this stage.

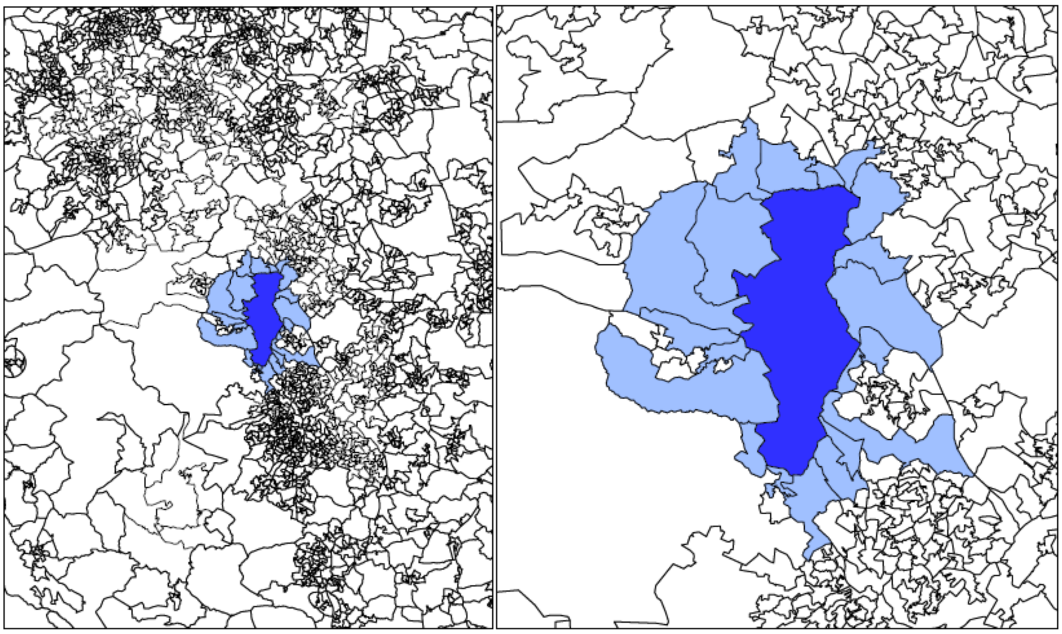
On average, each LSOA is bordered by 5.9 LSOAs, giving an average population of around 9000 in the ring. There is quite a range, however, with some LSOAs being bordered by 20 or more other LSOAs (Table 10). These occur typically where the core LSOA covers a rural area between two or more urban areas (see Figure 1 for an example). At the opposite extreme, some LSOAs have only one adjacent LSOA, where they lie at the end of a peninsular, for example. The great majority (93 per cent) have between 3 and 9 adjacent LSOAs.

Table 10: Number of adjacent LSOAs

		Frequency	Percent	Valid Percent	Cumulative Percent
Valid	1	92	.3	.3	.3
	2	586	1.8	1.8	2.1
	3	2025	6.2	6.2	8.3
	4	4928	15.2	15.2	23.5
	5	7527	23.2	23.2	46.8
	6	7170	22.1	22.1	68.9
	7	4563	14.1	14.1	83.0
	8	2557	7.9	7.9	90.9
	9	1263	3.9	3.9	94.8
	10	697	2.2	2.2	96.9
	11	382	1.2	1.2	98.1
	12	242	.7	.7	98.8
	13	142	.4	.4	99.3
	14	79	.2	.2	99.5
	15	45	.1	.1	99.6
	16	46	.1	.1	99.8
	17	20	.1	.1	99.9
	18	14	.0	.0	99.9
	19	12	.0	.0	99.9
	20	8	.0	.0	100.0
	21	4	.0	.0	100.0
	22	4	.0	.0	100.0
	23	4	.0	.0	100.0
	24	1	.0	.0	100.0
	25	1	.0	.0	100.0
	Total	32412	100.0	100.0	

Figure 2 also highlights that, in some cases, the use of the adjacency approach leads to a rather irregularly shaped surrounding ring. As the right-hand pane shows, some LSOAs quite close to the immediate neighbourhood (dark blue) are not included in the (light blue) surrounding ring (e.g. to the south east in this example). Johnston et al (2004) use two alternative approaches, based on fixed distance from the centre of the core LSOA and based on adding the nearest LSOAs up to a give population threshold. The former would tend to give a more regular shape in general, but would also lead to much large populations in more dense urban areas than in more sparse rural areas. The latter would give a more regular shape and a more consistent population but is more demanding computationally. Both of Johnston et al's approaches could lead to LSOAs being joined across major rivers unless additional checks are implemented.

Figure 2: Example of LSOA with many adjacent LSOAs



Levels and mix

Having defined the set of LSOAs that make up the broader scale for each LSOA, the challenge is again to decide how to measure the characteristics of these rings.

For the immediate neighbourhood, we have three different kinds of measure:

- population percentages;
- entropy and factors scores; and
- cluster typologies.

For the surrounding ring, population percentages can be derived directly from weighted averages for constituent LSOAs. Entropy and factor scores could be crudely estimated in the same way but these estimates may be misleading since scales are not linear. They are therefore re-estimated directly from the aggregated data.

Entropy scores for the surrounding rings are summarised in Table 7 and Figure 1 above. The distributions of the scores are generally very similar to those for the immediate neighbourhood. Entropy for the surrounding ring tends to be slightly higher on average and with slightly less variation, as we might expect with larger areal units. Correlations between entropy scores for the immediate neighbourhood and the surrouding rings vary, from .50 for tenure and household type, to .68 for NS-Sec and .93 for ethnicity.

For the factor analysis, we first reproduced the same set of 23 variables for the surrounding ring as had been used to derive factor scores for the immediate neighbourhood. Most have very similar mean values and slightly reduced spread, as would be expected. Two variables differed significantly in their mean values, however: density and the proportion of land given to greenspace. Mean density for the surrounding ring tended to be much lower while the mean area of greenspace tended to be much higher. This is because, when adding areas together, larger, more rural areas dominate. When factor scores were produced using these versions of the density and greenspace indicators, they differed quite significantly from the factors scores for the immediate neighbourhoods.

Alternative density and greenspace measures were therefore estimated, using population weighted averages. This not only produces mean values much closer to the original but also leads to factor solutions with more similar structures and similar correlations between factors in the oblique rotated solution.

Table 11 shows the correlations between the constituent variables and the five factor scores at each scale. Overall, the rotated solutions are very similar. There is very little difference in the loadings for the first two factors. With the third factor, the surrounding ring factor loads slightly more strongly on students and minority ethnic groups than at the neighbourhood level, suggesting that these kinds of areas are often in close proximity even where groups do not coexist at the level of the LSOA. The rural/low density factor is markedly stronger on almost all its loadings at the scale of the surrounding ring.

Looking at correlations between factors scores within each scale (Table 12), correlations do tend to be slightly greater for the surrounding rings, as Johnston et al (2004) found.

Correlations across the two scales were relatively high but perhaps lower than those implied by Johnston et al (2004). This may reflect the fact that we define the surrounding ring to exclude the core neighbourhood itself.

F51 - .685
F52 - .475
F53 - .675
F54 - .739
F55 - .625

Table 11: Correlations between variables and factor scores – immediate neighbourhood and surrounding ring

	Immediate neighbourhood					**Surrounding ring**				
	F1: Socio-economic Disadv.	**F2: Older People (not families, nor YP 25-39)**	**F3: Mobile YP 18-29 in PRS (not families)**	**F4: Minority Ethnic Groups & Students (not OP)**	**F5: Rural, Low Density**	**F1: Socio-economic disadv.**	**F2: Older people (not families, nor YP 25-39)**	**F3: Mobile YP 18-29 in PRS (not families)**	**F4: Minority ethnic groups & students (not OP)**	**F5: Rural, low density**
% Routine/Never Worked	0.95	0.18	0.13	0.24	-0.28	0.95	0.17	0.13	0.29	-0.29
% Unemployed	0.85	0.02	0.34	0.35	-0.42	0.87	-0.03	0.38	0.36	-0.54
% Inactive	0.63	0.72	0.18	0.28	-0.12	0.65	0.67	0.31	0.27	-0.18
% students (log)	0.02	-0.04	0.46	0.83	-0.29	0.08	-0.01	0.74	0.68	-0.40
% no quals/level 1	0.76	0.26	-0.30	-0.26	-0.07	0.71	0.26	-0.44	-0.18	0.01
% Social Rent	0.83	0.11	0.23	0.14	-0.31	0.78	-0.08	0.30	0.18	-0.44
% Private Rent (sqrt)	-0.06	-0.14	0.83	0.37	-0.21	-0.04	-0.13	0.86	0.28	-0.30
% LLTI	0.64	0.71	0.06	-0.29	-0.15	0.71	0.62	-0.05	-0.31	-0.16
IMD Crime Depvn	0.60	0.01	0.37	0.30	-0.49	0.69	-0.10	0.38	0.29	-0.66
% age 18-24 (log)	0.28	-0.19	0.66	0.63	-0.43	0.29	-0.14	0.79	0.55	-0.49
% age 25-29	0.14	-0.54	0.75	0.28	-0.51	0.19	-0.58	0.72	0.31	-0.64
% age 30-39	0.02	-0.81	0.31	0.06	-0.33	0.07	-0.86	0.28	0.14	-0.45
% age 40-49	-0.44	-0.30	-0.59	-0.12	0.38	-0.44	-0.24	-0.66	-0.16	0.51
% Single Pensioner	0.18	0.71	0.22	-0.40	-0.14	0.13	0.69	0.13	-0.53	-0.19
% Couple/All Pensioner	-0.33	0.75	-0.36	-0.52	0.36	-0.38	0.72	-0.44	-0.58	0.46
% cple+kids	-0.36	-0.43	-0.72	-0.07	0.37	-0.39	-0.21	-0.78	-0.01	0.56
% Lone Parent	0.85	-0.05	0.22	0.20	-0.37	0.86	-0.10	0.25	0.21	-0.50
% other+kids (sqrt)	0.46	-0.08	0.17	0.73	-0.35	0.48	-0.15	0.33	0.80	-0.50
% Asian (sqrt)	0.18	-0.16	0.24	0.78	-0.41	0.23	-0.18	0.34	0.84	-0.50
% Bl/Ch/Mx/Ot	0.23	-0.27	0.45	0.66	-0.50	0.23	-0.39	0.56	0.61	-0.62
Gross turnover % (log)	0.12	-0.25	0.88	0.28	-0.26	0.10	-0.20	0.92	0.21	-0.34
Popln density (log)	0.33	-0.15	0.26	0.31	-0.98	0.37	-0.20	0.36	0.30	-0.92
% Area Greenspace	-0.23	0.09	-0.26	-0.30	0.90	-0.33	0.15	-0.41	-0.32	0.96

Table 12: Correlations between rotated factors scores – immediate neighbourhood and surrounding ring

	Immediate neighbourhood				Surrounding ring			
	F1: Socio-economic Disadv.	**F2: Older People (not families, nor YP 25-39)**	**F3: Mobile YP 18-29 in PRS (not families)**	**F4: Minority Ethnic Groups & Students (not OP)**	**F1: Socio-economic disadv.**	**F2: Older people (not families, nor YP 25-39)**	**F3: Mobile YP 18-29 in PRS (not families)**	**F4: Minority ethnic groups & students (not OP)**
F2: Older People (not families, nor YP 25-39)	0.17				0.13			
F3: Mobile YP 18-29 in PRS (not families)	0.17	-0.06			0.16	-0.08		
F4: Minority Ethnic Groups & Students (not OP)	0.16	-0.20	0.29		0.17	-0.23	0.30	
F5: Rural, Low Density	-0.33	0.10	-0.37	-0.28	-0.38	0.15	-0.47	-0.25

The cluster typologies were also re-estimated, based on the same equivalent variables for the surrounding rings. Results were reported above (Table 7). Although the cluster typologies are very similar in structure, patterns of membership are slightly different for the surrounding rings. As these are based on aggregated areas, the spread of values on each variable is lower so that rings are slightly more likely to be members of more ‘central’ clusters than more extreme ones.

In summary, we end up with a set of measures for ‘levels’ in each surrounding ring which mirrors the set calculated for the core LSOA – a further 33 variables. We will also attach a variable indicating the number of LSOAs in the ring to permit us to remove cases with very few or very many LSOAs.

Patterning

As well as measuring context at two different scales, we have said we will try to capture aspects of patterning. By this, we mean that we want to be able to identify the presence of ‘extreme’ LSOAs within the surrounding ring, not merely the average characteristics of the ring. One issue is the ‘cut-off’ value that should be used to identify such extremes. For now, we can avoid decisions on that point by attaching actual values.

Our approach to patterning is to identify the most and least deprived neighbourhood (LSOA) in the ring, and to attach the full set of scores for each to the immediate neighbourhood. We use the factor score for the socio-economic disadvantage factor to do this since it is the composite of a number of underlying variables and it explains the largest proportion of the variance. Using the IMD income deprivation score gives very similar results; in three quarters of cases, the neighbourhood with the highest factor score also has the higher income deprivation score, while in two thirds of cases, the neighbourhood with the lowest factor score also has the lowest income deprivation score.

In this way, a further 66 variables are added to the database – 33 for the most deprived LSOA in the surrounding ring, and 33 for the least deprived LSOA in the ring.

4. Summary

We are constructing quite a novel database on neighbourhood context, which captures multiple dimensions of each place at two scales, and reflects the patterning of poverty and affluence within each area.

The variables are summarised in Tables 13 and 14.

References

Johnston, R., Jones, K., Burgess, S., Propper, C., Sarker, R., and Bolster, A. (2004) Scale, factor analyses and neighborhood effects, *Geographical analysis* 36 (4): 350(19).

Reardon, S. F. and Firebaugh, G. (2002) Measures of multigroup segregation, *Sociological Methodology* 32: 33-67.

Suttles, G. (1972) *The social construction of communities.* Chicago: Chicago UP.

Table 13: Number and type of variable by level

	Immediate neighbourhood			**Surrounding ring – average for all LSOAs**			**Surrounding ring – most deprived LSOA**			**Surrounding ring – least deprived LSOA**		
	Level	Entropy	Cluster	Level	Entropy	Cluster	Level	Entropy	Cluster	Level	Entropy	Cluster
Scalar												
IMD	4			4			4			4		
Poverty	2			2			2			2		
NS-Sec	2	1	1*	2	1	1*	2	1	1*	2	1	1*
Employment status	3			3			3			3		
Tenure	2	1	1*	2	1	1*	2	1	1*	2	1	1*
Household type	4	1	1*	4	1	1*	4	1	1*	4	1	1*
Ethnicity	3	1	1*	3	1	1*	3	1	1*	3	1	1*
Scaleless												
Factor scores	5			5			5			5		
Total	25	4	4	25	4	4	25	4	4	25	4	4

Notes: All variables are continuous except ‘*’ which are categorical. In addition to the 132 variables measuring neighbourhood characteristics, we include: an LSOA identifier (lsoa) which is used only to make the connection with the individual data; and nlsoa which identifies the number number of LSOAs in the surrounding ring.

Table 14: Details for the individual variables

Dimension (denominator)	***Variable***	***Var name***	***Source, basis***
IMD (popall)	% income deprived - all	imdinc	IMD07, Single variable
(popch)	% income deprived - children	imdincch	IMD07, Single variable
(popop)	% income deprived – older people	imdincop	IMD07, Single variable
(popwa)	% income deprived – working-age adults	imdincwa	IMD07, Single variable
Poverty	% “Breadline poor”	poorb	Census, Combination of variables
(poor_hhld)	% “Core poor”	poorc	Census, Combination of variables
NS-Sec	% Professional, Managerial/Intermediate	nssec12	Census, Single variable
(nssec_pop)	% Routine Manual/Never Worked	nssec78	Census, Single variable
	Entropy score	nssecent	Census, Five NS-Sec variables
	Cluster type	nsseccl	Census, Five NS-Sec variables
Employment status	% Unemployed	ecounemp	Census, Single variable
(econact_pop)	% Inactive	ecoinact	Census, Single variable
	% Students	students	Census, Single variable
Tenure	% Private Rent	tenpr	Census, Single variable
(tenure_pop)	% Social Rent	tensr	Census, Single variable
	Entropy score	tenent	Census, Four tenure variables
	Cluster type	tencl	Census, Four tenure variables

[Table 14 continued]

Dimension (denominator)	***Variable***	***Var name***	***Source, basis***
Household type	% pensioner households	hhop	Census, Single variable
(hh_pop)	% working age adult household	hhwa	Census, Single variable
	% households with children	hhch	Census, Single variable
	% lone parent households	hhlp	Census, Single variable
	Entropy score	hhent	Census, Four household variables
	Cluster type	hhcl	Census, Four household variables
Ethnicity	% Asian	ethas	Census, Single variable
(ethnic_pop)	% Black	ethbl	Census, Single variable
	% Other	ethot	Census, Single variable
	Entropy score	ethent	Census, Four ethnicity variables
	Cluster type	ethcl	Census, Four ethnicity variables
Factor scores	F1: socio-economic disadvantage	f51	Census, Combination of variables
(n/a)	F2: Older people (not families, nor 25-39 yr olds)	f52	Census, Combination of variables
	F3: Mobile YP 18-29 in PRS (not families)	f53	Census, Combination of variables
	F4: Minority ethnic groups & students (not OP)	f54	Census, Combination of variables
	F5: Rural, low density	f55	Census, Combination of variables

Note: ‘Single variable’ can be combination of several Census table cells.