

Measuring the Changing Pattern of Ethnic Segregation in England and Wales with Consumer Registers

Abstract

Analysis of changing patterns of ethnic residential segregation is usually framed by the coarse categorisations of ethnicity used in censuses and other large-scale public sector surveys and by the infrequent time intervals at which such surveys are conducted. In this paper, we use names-based classification of Consumer Registers to investigate changing degrees of segregation in England and Wales over the period 1997-2016 at annual resolution. We find that names-based ethnic classification of the individuals that make up Consumer Registers provides reliable estimates of the residential patterning of different ethnic groups and the degree to which they are segregated. Building upon this finding, we explore more detailed segregation patterns and trends of finer groups at annual resolutions and discover some unexpected trends that have hitherto remained unrecorded by Census-based studies. We conclude that appropriately processed Consumer Registers hold considerable potential to contribute to various domains of urban geography and policy.

Keywords: segregation; Consumer Registers; names classification; ethnicity; social integration; spatial analysis.

1. Introduction

Ethnic residential segregation has provided an enduring and debated focus for social investigation in the United Kingdom and elsewhere. Both the academic and public discussions are frequently dominated by anecdotal evidence because extensive, timely and detailed data on ethnic residential patterns are unavailable. Segregation researchers recognise that the lack of sufficiently granular data with respect to ethnic categories and temporal resolution are of paramount importance and impede progress in significant policy debates, such as migration and segregation in England and Wales (Harris & Johnston, 2018). Additional obstacles in understanding the patterns, causes and consequences of ethnic segregation arise from the “slippery” (Peach, 2009) nature of segregation measures, which continue to be contested.

In the United Kingdom, conventional data sources of ethnicity information are mainly drawn from decennial censuses of population. While they are broad in coverage, relying on census data creates significant gaps, as data are only collected every ten years. The delay to the release of the information means that, currently, the most recent data are eight years out of date and the next population-wide update cannot be expected before 2022. Moreover, census categorisations provide little flexibility in profiling particular ethnic communities of policy concern beyond the coarse, pre-defined ethnic categories released by the Census.

In this paper, we seek to demonstrate the feasibility of using Consumer Registers (Lansley et al., 2019) as an alternative population data source to official censuses, in order to develop a more granular analysis of recent segregation trends and patterns in England and Wales. We make annual estimates of ethnic segregation in England and Wales from 1997 to 2016 for all ethnic groups recorded by the Census as well as a selected number of finer ethnic categories. We develop two innovative manipulations: (1) we employ the algorithm developed by Kandt and Longley (2018) to infer probable ethnic origins for aggregations of individual names at

43 the Census Output Area level; and then (2) track annual segregation estimates for England
44 and Wales as a whole and for four case studies.

45 **2. Consumer Registers**

46 The Linked Consumer Registers (Lansley et al., 2019) are a set of annually assembled
47 databases of names, home addresses and postcodes of adult individuals spanning a period of
48 of nearly 20 years, 1997-2016. They are compiled from the public Electoral Register and a
49 range of private sector data sources. Based on individual names and addresses, it is possible
50 to match records between years and thus turn Consumer Registers into a powerful,
51 longitudinally linked data resource that can be aggregated to any convenient geography. Such
52 a resource can permit novel insights into a range of research and policy problems, including
53 segregation.

54
55 Yet, as is typical for consumer and other big data sources, Consumer Registers require
56 significant data cleaning and pre-processing before they can be deployed for research
57 purposes. A major challenge arises from the unknown provenance of individual records
58 because Consumer Register inherit records from the Electoral Register, which is known to
59 exclude non-electors and (post 2003) electors that have exercised their right to withdraw from
60 the public version of the Electoral Roll. Issues of residential mobility bring additional
61 concerns: recent research has estimated that 17% of eligible voters in Great Britain are not
62 correctly registered at their current address, representing as many as 9.4 million people
63 (Electoral Commission 2019). The same study revealed that 11% of register entries are
64 inaccurate, affecting up to 5.6 million people.

65 In order to attain wider coverage of non-electors, Electoral Register records have been
66 supplemented with records of other consumer data sources based on appropriate consents
67 from data subjects. By means of appropriate address standardisation methods, significant
68 parts of missing records could be reconstructed. Furthermore, Lansley et al. (2019) develop
69 procedures matching the individual names and addresses between registers of different years
70 through linkage to external data sources, such as the Ordnance Survey AddressBase list of all
71 addresses in Great Britain and Land Registry data for England and Wales and for Scotland.
72 Using the matched, annual records, they were able to fill gaps or correct inaccuracies that
73 may have arisen due to moves.

74 Although these cumulative additions are unlikely to bring the Consumer Registers to
75 universal adult coverage, Lansley et al (2019) demonstrate that population totals can be
76 brought in line with the 2001 and 2011 Census population counts and Office for National
77 Statistics (ONS) Mid Year Population Estimates. The total number of recorded adults in the
78 linked Consumer Registers is highly correlated (correlation coefficients $> .99$) with the Office
79 for National Statistics (ONS) Mid-Year Population Estimates over the period 1997 to 2016:
80 the average difference in estimated population sizes is just 1.8%.

81 Consumer Registers do not record any demographics but names can be used to infer age, sex
82 and ethnicity (Longley et al. 2015). Names-based classifications have been proved to be
83 viable for ethnicity related studies (Bakens & Pryce, 2018; Lan et al., 2018; Longley et al.,
84 2015; Nathan, 2015). In contrast to previous name-based segregation studies, the freely
85 available Ethnicity Estimator tool (indicators.cdrc.ac.uk/ethnicity-and-names/ee/) has been
86 calibrated with the 2011 Census population data and demonstrate a success rate of 88%
87 (Kandt and Longley, 2018), which performs better than other algorithmic comparators such
88 as Onomap (*ibid*). In what follows, however, we do use the Onomap algorithmic solution to

89 estimate numbers of adults within the blanket Census ‘Other White’ category, provided that
90 the higher level ethnicity estimator software assigns individuals to this blanket category.
91 Using these procedures, we infer ethnic group of each name bearer in the Consumer Registers
92 and are thus able to generate annual estimates of ethnic diversity and segregation.

93

94 **3. Background: Ethnic segregation in the British context**

95 Ethnic segregation first returned to the public and political spotlights after the 2001
96 disturbances. Policy reviews by governmental think-tanks and experts (Cantle, 2001) have
97 suggested that the disturbances resulted from inequality and self-segregation among minority
98 groups. Since then, continued efforts to monitor segregation in the country have demonstrated
99 that degrees of segregation have generally been decreasing, albeit with some variations
100 among ethnic groups (Cantle and Kaufmann, 2016; Casey, 2016). Simpson (2007) argues that
101 residential mixing increases as minority populations grow and spread more evenly across
102 localities. Likewise, Peach (2009) concludes that ethnic segregation levels in Britain have
103 been decreasing, challenging myths of US-style ghettos in English cities (Peach, 2010).
104 Similar evidence is presented by Finney and Simpson (2009). Catney (2016) analyses the
105 1991, 2001 and 2011 UK Censuses and charts the geography of ethnic diversity that has
106 evolved over the last two decades. Correspondingly, Harris and Owen (2018) define a
107 multilevel Index of Dissimilarity based on the 2001 and 2011 Census data to explore the
108 changing scales of residential segregation in England and Wales. They find that the Index of
109 Dissimilarity is falling even though patterns of segregation are emerging at less localised
110 geographical scales. Apart from the index-based segregation studies, another important strand
111 of work focuses on the ethnic composition and typologies of neighbourhoods using Census
112 data (Johnston et al., 2007; Poulsen et al., 2011; Johnston et al., 2016). Their findings show
113 the increase in both Type 2 (White predominate) and Type 4 (Non-White predominate)
114 neighbourhoods across the urban systems of England and Wales.

115 Although the evidence is broadly consistent, it remains at a coarse level as a result of the data
116 source that most studies employ, notably the UK Censuses. Census data are authoritative and
117 widely accessible but they are not without drawbacks. First, geographic boundaries of the
118 Census output units are subject to changes across censuses. For example, between the 2001
119 and 2011 UK Censuses, 4,354 of 175,434 Output Areas (2.4%) in England and Wales were
120 either split or merged. Such inconsistent boundaries lead to comparability issues (Simpson,
121 2007) when developing census-based segregation indices. Moreover, different Census
122 classifications of ethnicity have been created over time following stakeholder consultations
123 and changes in priorities mean that they are not consistent across the 1991, 2001 and 2011
124 UK Censuses. The current practice of defining ‘pan-ethnic’ classes (Aspinall, 2002)
125 combines distinctive minority groups in Britain and renders categories imprecise and
126 problematic. For example, many European immigrants, such as Polish, Germans, Italians,
127 French, Lithuanians, Portuguese and Romanians, are categorised collectively as ‘Other
128 White’ in Censuses, despite the fact that this comprises a diverse range of cultural origins that
129 likely pose distinctive challenges to social harmonisation. Consequently, the residential
130 segregation patterns of the various groups within the ‘Other White’ category may vary
131 markedly and contribute to a number of different policy concerns.

132 These vicissitudes have previously been overcome by using algorithmic classifications of
133 individual given- and family-name pairings to establish a consistent benchmark for analysis
134 (Mateos et al., 2011). Names-based classification of ethnicity has been applied in previous

135 research (Bakens & Pryce, 2018; Lan et al., 2018; Longley et al., 2015; Nathan, 2015).
136 However, unlike the classification that we use here, such tools do not validate their results
137 with reference to the self-assignments (e.g. Census ethnic groups) that individuals make
138 (Kandt and Longley, 2018), despite the evidence that such assignments are not invariably
139 stable or straightforward (Simpson et al., 2016).

140 Neighbourhood change may be rapid, rendering decennial measurements of segregation too
141 coarse to base propositions on segregation trends, especially when a key movement of ethnic
142 groups occurs between census years. For example, a large number of ‘Other White’ migrants
143 moved to the UK between the 2001 and 2011 Censuses following the 2004 and 2007
144 accessions of Eastern European countries to the European Union (EU). Although there exist
145 other population data sources for the intermediate years, they are not precise and sometimes
146 suffer from the underestimation of long-term immigration (Office for National Statistics,
147 2012). Thus most previous residential segregation studies have drawbacks of inconsistency,
148 imprecise ethnic categorisations and very coarse temporal granularity.

149 **4. Research Design**

150 Since the novel names-based classification tool, Ethnicity Estimator, is sensitive to ethnic
151 self-assignment in the British context (Kandt & Longley 2018). we use it to classify linked
152 records in Consumer Registers. This process results in Consumer Register records that are
153 coded up by granular ethnic group, geo-referenced and longitudinally linked between 1997
154 and 2016. Thus, we are able to define a number of segregation case studies that can be
155 benchmarked against the picture offered by conventional data sources. In so doing, we
156 demonstrate the value of Consumer Registers for segregation studies and outline new,
157 potential insights.

158 **4.1. Definition of case studies**

159 Using the Ethnicity Estimator as described above, we are able to follow the 2011 Census
160 ethnicity categorisation, which was developed by the ONS in consultation with key
161 stakeholders (Office for National Statistics, 2009), and to further isolate French, Greek and
162 Polish ethnicities from within the ‘Other White’ category of estimates. The longitudinal
163 structure of the linked Consumer Registers further enables us to attain an annual resolution by
164 which segregation may be measured. We further aggregate the population counts by ethnic
165 groups to 2011 Census Output Areas, which are the lowest geographical level in the UK
166 Census and are used by the ONS to maintain disclosure control for census outputs.

167 We first carry out some general validation of estimated ethnic composition in urban and rural
168 areas in England and Wales. We proceed with an estimate of segregation of ethnic groups in
169 those areas, and finally explore the different dynamics of segregation at annual resolution
170 between 1996 and 2007 in three urban study regions, Greater London, Greater Manchester,
171 and Birmingham and one rural region, Lincolnshire. In those regions, we go beyond Census
172 categories and explore trends related to three groups normally subsumed under ‘White
173 Other’: French, Greek and Polish. Our hypothesis is that the UK settlement patterns of
174 nationals of founding and accession EU member states are each geographically distinctive.
175 This distinction demonstrates the heterogeneity masked within conventional data sources for
176 ethnicity research.

177 **4.2. Measurement of segregation**

178 **5.**

179 Although numerous indices have been proposed to quantify the extent of ethnic segregation,
180 each of them has limitations and fails to capture all dimensions (Massey and Denton, 1988).
181 For example, a common critique voiced about the Index of Dissimilarity is that it is heavily
182 influenced by the population size, and it could overestimate segregation levels to different
183 degrees, depending on spatial unit sizes and minority proportions. More recent
184 methodological developments such as the Theil's entropy index overcome this issue;
185 however they mostly are undermined by the difficulties of interpretation. Despite this
186 critique, such indices are widely applied in academic and policy studies. Hence, in order to
187 ascertain the value of novel data sources in advancing segregation research, we examine
188 ethnic segregation using the Index of Dissimilarity and the Index of Isolation representing
189 evenness and exposure respectively for comparability and interpretability purposes.
190 Nonetheless, the use of Consumer Registers and other aspects of our analysis can be easily
191 extended to other segregation measurements.

192 The pairwise Index of Dissimilarity D is rewritten in Equation (1), where w_i represents
193 population counts of an ethnic group in the i^{th} Output Area, and W represents the total
194 population of that group in the whole region. Similarly, b_i denotes population counts of the
195 rest of the ethnic groups in the i^{th} Output Area, and B is the total population of the rest in the
196 whole region. The Index of Isolation P of a specified ethnic group is calculated using
197 Equation (2), where x_i denotes population counts of the specified group in the i^{th} Output
198 Area, and t_i denotes the total population of the i^{th} Output Area. The denominator X is the
199 total population of the specified group in the whole region. The isolation index values are
200 sensitive to the relative size or percentage of ethnic groups among the total population. For
201 this reason, comparisons of isolation indices can only be made between ethnic groups within
202 a city, rather than inter-city comparisons (Peach, 2009).

$$D = \frac{1}{2} \sum_{i=1}^n \left| \frac{w_i}{W} - \frac{b_i}{B} \right| \quad (1)$$

$$P = \sum_{i=1}^n \left[\left(\frac{x_i}{X} \right) \left(\frac{x_i}{t_i} \right) \right] \quad (2)$$

203 In addition, in order to evaluate the influence of randomness, we test the significance of the
204 Dissimilarity Index under the null hypothesis of no systematic segregation. Following the
205 randomisation tests of Boisso et al. (1994) and Carrington and Troske (1997), we generate
206 pseudo-sample distributions with 1000 repetitions by randomly allocating individuals from
207 different ethnic groups to 2011 Census Output Areas. In each repetition, random numbers of
208 the group population w_i and b_i are created under the multinomial distribution using the
209 restricted probabilities that the chance of a resident from either ethnic group being allocated
210 to a unit i equals the proportion of the population in unit i compared to the total population of
211 n units in the study area. Using 1000 repetitions of the random allocation process, we

212 calculate the mean Dissimilarity Index D^* and confidence intervals (CIs) to test the null
 213 hypothesis that the observed segregation level D is produced by randomness solely.

214 **6. Results**

215

216

217 **6.1. Ethnic composition in England and Wales**

218 We validate ethnic composition, as estimated from Consumer Registers, with the adult
 219 population (age 16 and over) of the Censuses in 2001 and 2011 and find a high level of
 220 consistency (see Table 1). The estimated proportions of ‘White British’, Indian, Pakistani,
 221 Bangladeshi and ‘Black African’ groups estimated from the Consumer Registers are similar
 222 to those found in the Census. The Chinese are underrepresented in comparison with the 2001
 223 Census and the ‘Black Caribbean’ groups are underrepresented in both years; whereas the
 224 ‘White Irish’ groups are overrepresented in Consumer Registers. Some possible explanations
 225 for the overestimation of Irish include mismatches between names-based ethnic origin and
 226 self-identification, given a high degree of common historical roots and cultural similarity
 227 between British and Irish people. Kandt and Longley (2018) describe how the Ethnicity
 228 Estimator software upweights probabilistic assignments of all ethnic minorities relative to
 229 White British, in order to improve successful identification of minority group membership: a
 230 consequence of this is the observed under-prediction of the White British group, largely
 231 corresponding to over-prediction of the White Irish group that may consider itself largely
 232 assimilated into the White British category. The underrepresentation of minority groups, such
 233 as the Chinese, may be due to their ineligibility to register to vote. Consequently, they are not
 234 well captured in the pre-2003 Consumer Registers which are largely based on the Electoral
 235 Roll. Such a situation appears to be mitigated in the 2011 Consumer Register by the
 236 compensation of consumer sources. In addition, the Black Caribbeans are more difficult to
 237 detect through names classification due to the disruption of traditional naming practices as
 238 one of the consequences of a family history of enslavement (Kandt and Longley, 2018).

Table 1 Ethnic composition in percent found in Consumer Registers compared with the adult
 population (16 and over) 2001 and 2011 Censuses (italics);
 (source: Author’s own calculations with the 2001 and 2011 Censuses)

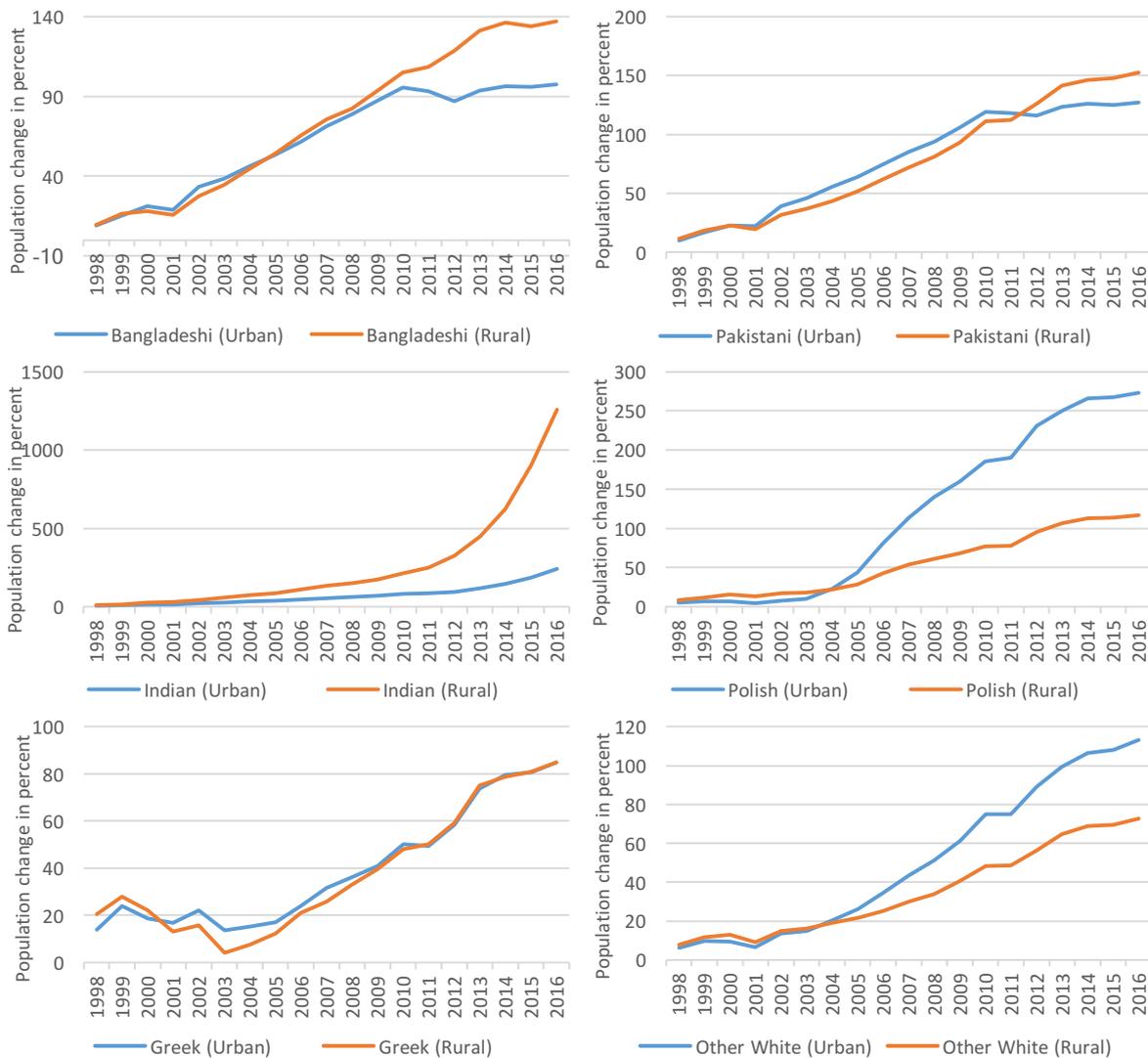
Year	White British	White Irish	Other White	Indian	Paki-stani	Bangla-deshi	Chinese	Other Asian	Black African	Black Caribbean	Any Other
2011	79.6	5.1	5.4	3.4	2.3	0.6	0.5	0.7	1.4	0.4	0.5
	<i>83.0</i>	<i>1.1</i>	<i>4.8</i>	<i>2.5</i>	<i>1.7</i>	<i>0.7</i>	<i>0.8</i>	<i>1.4</i>	<i>1.5</i>	<i>1.1</i>	<i>1.3</i>
2001	85.4	5.2	3.4	2.1	1.4	0.4	0.3	0.3	0.6	0.4	0.4
	<i>88.2</i>	<i>1.5</i>	<i>2.8</i>	<i>1.9</i>	<i>1.1</i>	<i>0.4</i>	<i>0.9</i>	<i>0.4</i>	<i>0.8</i>	<i>1.1</i>	<i>0.9</i>

239 In sum, the inferred proportions of ethnic groups in Consumer Registers largely align with
 240 those in the Censuses, and hence Consumer Registers offer a valid resource to estimate ethnic
 241 composition in more details. To do so, we divide Output Areas into urban and rural Output

242 Areas using the ONS 2011 Rural-Urban Classification¹ for England and Wales. Defining
243 1997 as the base-year, we calculate yearly population counts for each ethnic group and divide
244 these by population counts for the corresponding groups in 1997 (see Figure 1 and
245 Supplementary material Table S1).

246 Both urban and rural areas have experienced population growth by all ethnic groups over the
247 last 20 years. Growth is particularly pronounced for Indians in rural areas. In addition, two
248 trends can be identified from annual population growth rates by selected ethnic groups, as
249 shown in Figure 1. Growth in numbers of some ethnic groups is increasingly divergent in
250 urban and rural areas, but the nature of this divergence differs among some ethnic groups.
251 While growth of Bangladeshi, Pakistani and Indian populations has accelerated in rural areas
252 and stagnated in cities since the last Census, ‘Other White’ populations have grown more in
253 cities. Furthermore, Consumer Registers document the marked increase in the urban Polish
254 population following the Polish EU accession and further stagnating of this trend post 2014.
255 A different trend emerges for the Greek population, which shrank between 1998 and 2004,
256 only to then grow steadily in urban and rural areas alike. A similar trend can be traced for
257 French name bearers although stronger growth occurred in urban areas. If the trends
258 estimated from Consumer Registers are reliable, we can expect to see in the next Census
259 notable demographic shifts in rural areas by influx from Asian groups and modestly
260 decelerating growth of some European groups in cities.

¹ <http://www.ons.gov.uk/ons/guide-method/geography/products/area-classifications/2011-rural-urban/index.html>



261

262

263

264

265

Figure 1. Annual population change in percent relative to 1997 for selected ethnic groups in urban and rural areas estimated from Consumer Registers.

266 **6.2. Changes in ethnic segregation in England and Wales**

267 While ethnic segregation is mainly studied in metropolitan areas, trends in ethnic
 268 composition suggest that the movement of ethnic minorities and migrants into rural localities
 269 in recent decades deserves increasing attention (see also Catney, 2016; Finney and Simpson,
 270 2009). Again, we validate segregation estimates from Consumer Registers with those derived
 271 from the Censuses of 2001 and 2011. We estimate Dissimilarity Indices for urban and rural
 272 areas at the national level and calculate their confidence intervals.

273 We first show the Dissimilarity Indices D and ranks of the urban Output Areas from
 274 Consumer Registers in 2011 and 2001 and compare to the equivalent statistics derived from
 275 the Censuses (Table 2). Taking values of D from 2011 Consumer Registers as an illustration,
 276 ethnic minorities (e.g. Pakistanis, Bangladeshis and Africans) are among the most segregated
 277 groups while the least segregated groups are the White groups (British, Irish and Other
 278 White). The equivalent D values derived from the 2011 Census suggest that the D values may
 279 be underestimated for all groups except the Chinese (which tend to be underestimated in
 280 Consumer Registers, as noted above). The ranking by D value across the 2011 Consumer

281 Registers and Census remains unchanged for Bangladeshis (1), Pakistanis (2), Africans (4)
282 and 'White Irish' (10), with minor differences for other groups.

283 With respect to temporal changes, segregation levels from Consumer Registers and Censuses
284 have both decreased in 2011 with the exception of a slight increase observed for the 'White
285 British' using Consumer Registers. There are some upward and downward changes in the
286 ranks of D values from 2001 to 2011. For example, the Pakistanis and Chinese, 'White
287 British' and 'Other White' swap their ranks of D based upon Consumer Register data. Results
288 for the rural Output Areas are reported in the supplementary material (Table S2). In rural
289 Output Areas, segregation levels by ethnic groups are high relative to values recorded for
290 urban Output Areas for ethnic minorities; however, the 'White' groups are more evenly
291 distributed in rural Output Areas where segregation levels are generally lower.

292 As outlined in Section 3, we test the significance of the Dissimilarity Index D under the null
293 hypothesis of random allocation. The point estimates D^* and Confidence Intervals (CIs) from
294 the 1000 random tests are presented in Table 2. Dissimilarity Indices D from both Consumer
295 Registers and Censuses lie outside the confidence intervals of the point estimates D^* , thus
296 rejecting the null hypothesis segregation levels are random and confirming the existence of
297 systematic segregation. However, the ratios of D^* to D are generally higher in the rural
298 Output Areas than in the urban Output Areas, since a larger component of the observed
299 segregation D in rural areas may arise because of the relatively small population size of
300 ethnic minorities.

301 It is observed from Table 2 and Table S2 that the correspondence degrees of D values
302 between Consumer Registers and Censuses vary among individual ethnic groups. In order to
303 present a summary metric of the conformity of the Dissimilarity Indices obtained from
304 Consumer Registers with those from the Censuses, we conduct correlation analysis of the D
305 values and ranks of the 10 ethnic groups from Table 2 and Table S2 for both urban and rural
306 areas. Results of the analysis are summarised in Table 3. Here, the coefficients suggest that
307 there are strong and positive correlations (coefficients $> .8$) between Consumer Registers and
308 Censuses in terms of the index values and ranks at the 99% confidence level (p -value $< .01$).
309 All in all, the correlations suggest that Consumer Registers offer a broadly accurate picture of
310 the ethnic structure of segregation in the country and can be used a source of information
311 when examining segregation during the intercensal period. At the same time, the group-
312 specific segregation indices should be viewed with caution.

Table 2 Consumer Register and Census Dissimilarity Indices (D) and associated ranks, showing changes since 2001 and point estimates D* by random allocations with Confidence Intervals (CI) for urban Output Areas in 2001 and 2011.

	2011 Consumer Register			2011 Census			2001 Consumer Register			2001 Census		
	D	D* (CI)	Rank	D	D* (CI)	Rank	D	D* (CI)	Rank	D	D* (CI)	Rank
Bangladeshi	.732	.29889 (.29886-.29893)	1 (0)	.775	.22964 (.22961-.22967)	1 (0)	.809	.37048 (.37044-.37052)	1	.855	.28342 (.28339-.28345)	1
Pakistani	.667	.15386 (.15384-.15388)	2 (+1)	.734	.14563 (.14561-.14564)	2 (0)	.715	.20233 (.2023-.20235)	3	.769	.17891 (.17889-.17893)	<u>2</u>
Chinese	.629	.35112 (.35108-.35116)	3 (-1)	.587	.2507 (.25067-.25072)	<u>6</u> (0)	.718	.42931 (.42926-.42936)	2	.593	.23182 (.23179-.23185)	<u>6</u>
African	.596	.19488 (.19486-.19491)	4 (0)	.634	.15572 (.1557-.15574)	4 (-1)	.701	.30135 (.30131-.30139)	4	.749	.21829 (.21826-.21832)	<u>3</u>
Caribbean	.569	.37377 (.37373-.37381)	5 (0)	.651	.20025 (.20023-.20027)	<u>3</u> (+1)	.617	.40356 (.40351-.40361)	5	.691	.20158 (.20156-.2016)	<u>4</u>
Indian	.518	.13141 (.1314-.13143)	6 (0)	.612	.13145 (.13144-.13147)	<u>5</u> (0)	.615	.16674 (.16672-.16676)	6	.666	.14992 (.1499-.14994)	<u>5</u>
Any Other	.508	.21006 (.21003-.21008)	7 (0)	.448	.09433 (.09432-.09434)	<u>9</u> (0)	.585	.29077 (.29074-.29081)	7	.472	.15574 (.15572-.15576)	<u>9</u>
White British	.401	.06046 (.06045-.06046)	8 (+1)	.551	.05353 (.05352-.05354)	<u>7</u> (0)	.374	.07088 (.07087-.07089)	9	.557	.06506 (.06505-.06507)	<u>7</u>
White Other	.373	.10723 (.10722-.10725)	9 (-1)	.450	.10207 (.10205-.10208)	<u>8</u> (0)	.375	.13846 (.13844-.13848)	8	.478	.13874 (.13872-.13875)	<u>8</u>
White Irish	.243	.11252 (.11251-.11253)	10 (0)	.423	.2207 (.22067-.22072)	10 (0)	.252	.11417 (.11416-.11418)	10	.412	.19658 (.19655-.1966)	10

313
314

Table 3 Statistics of the correlation analysis of the Dissimilarity Indices and their ranks from Consumer Registers and Censuses

		Pearson Correlation on the Dissimilarity Index D		Spearman Correlation on the ranks	
		Coefficient	p-value	Coefficient	p-value
2011	Urban	.867	.001	.879	.001
	Rural	.843	.002	.867	.001
2001	Urban	.836	.003	.830	.003
	Rural	.891	.001	.842	.002

315

316 **6.3. Segregation dynamics in Greater London, Greater Manchester, Birmingham and**
317 **Lincolnshire between 1997 and 2016**

318 Set against the backcloth of rapid change in the ethnic composition of England and Wales, we
319 investigate the detail of change in annual increments. It is evident from Table S1 in the
320 supplementary material that the Polish group has experienced quite a different process of
321 demographic changes in terms of both amount and pace, compared with other ‘Other White’
322 sub-groups, such as the French or Greek. Such divergent patterns remain concealed within
323 the aggregated ‘Other White’ in the Census. To illustrate such divergent patterns, we take
324 advantage of our granular, names-based ethnicity estimation and explore segregation for
325 Polish, French and Greeks separately from ‘Other White’, measuring ethnic segregation in
326 Greater London, Greater Manchester, Birmingham and Lincolnshire. The first three cases are
327 traditional immigrant entry and settlement points. Lincolnshire is chosen as an example of a
328 rural and agricultural setting for increased inward, international migration.

329 Table S3 in the supplementary materials presents the Indices of Dissimilarity and Isolation in
330 the most recent year 2016 from Consumer Registers as an example to show the absolute
331 levels of segregation across ethnic groups. There are patterns of high, moderate and low
332 segregation levels for various ethnic groups concerning evenness. The White majority,
333 including ‘White British’, ‘White Irish’ and ‘Other White’, have lower segregation levels
334 compared to the other ethnic groups. The Bangladeshis in London are one of the most
335 segregated groups, and they are highly concentrated in particular areas in East London. The
336 Greek group appears to be the most segregated group in each of the four areas. However,
337 unlike the Bangladeshi community, the high segregation index values of the Greeks can be
338 mainly attributed to the small size of the Greek community, as the distribution of smaller
339 populations is more prone to randomness in a statistical sense as we demonstrate in the
340 previous section and Table 2. Given their small population size, similar arguments can be
341 applied to segregation levels of the French group in Greater Manchester and Birmingham, as
342 well as all of the ethnic minorities in rural areas like Lincolnshire.

343 Despite these statistical concerns, the relative, temporal changes of segregation indices
344 remain meaningful for each group (Simpson 2004). We plot the changes in Dissimilarity
345 Index values by ethnic groups each year relative to 1997 in Figure 2. We can ascertain that
346 although ethnic diversity has been increasing with respect to the proportions of the ethnic
347 minorities, the Dissimilarity Indices of most of the ethnic groups have been dropping, except
348 for ‘White British’ and ‘White Irish’ in the three urban areas and “Other White” in
349 Lincolnshire. Such a decrease suggests that these minority groups are more evenly distributed
350 and less segregated than before. In contrast to these groups with gently declining
351 Dissimilarity Indices, the Indian, Black African and Polish communities have experienced a
352 dramatic fall in the segregation levels in terms of the evenness dimension.

353 In particular, we observe pronounced decreases in the Dissimilarity Indices for the Poles
354 across the four areas from 2004. This trend is consistent with the national trend in urban and
355 rural areas presented in the supplementary Table S1. The changing pattern of the Polish
356 residents is quite different from the other communities from the EU, particularly the French.
357 We may speculate that the apparent dispersion of Polish residents in Lincolnshire is a result
358 of their settlement in areas of agricultural labour market shortages since the 2004 Polish EU
359 accession. The Indian group appears to be distributed more evenly across Output Areas in the
360 four urban regions, particularly after 2011. Such observations would not be possible with the
361 Census population data until the next Census in 2021.

362 Measuring the exposure dimension of residential segregation, we find varying levels of
363 Isolation among ethnic groups (Table S3). The south Asian groups – Bangladeshis, Pakistanis
364 and Indians – seem to be more isolated in the four case study areas, which indicates that they
365 tend to live in spatial clusters with less likelihood of meeting people from different ethnic
366 communities in their neighbourhoods. ‘Black Caribbean’, Chinese, Greek and French remain
367 at relatively low levels of Isolation, which may be partly due to their small overall population
368 sizes. Regarding the temporal trend of the Isolation Indices (see Figure 3), some of these
369 groups have become less segregated along the exposure dimension, for instance, the
370 Bangladeshis. Others have experienced increasing levels of Isolation, most notably the ‘Black
371 Africans’ in urban areas and the Polish, Indian and ‘Other White’ groups in Lincolnshire. The
372 Greek and French also exhibit almost identical stability in levels of segregation; conversely,
373 there has been increased Isolation of Poles since 2004, which may reflect the sensitivity of
374 Isolation indices to relative population size. Combining the two measurements, Simpson
375 (2007) has suggested that the Index of Dissimilarity and Index of Isolation can change in
376 opposite directions in response to significant streams of immigration. Such contrary changes
377 are observed in our study as well. For example, the ‘Other White’ group appears to have
378 dispersed with the influx after the EU accession, while earlier generations of the ‘Other
379 White’ group may have moved away from existing concentrations. These trends are reflected
380 in the decreasing Dissimilarity Indices and increasing Isolation Indices.

381

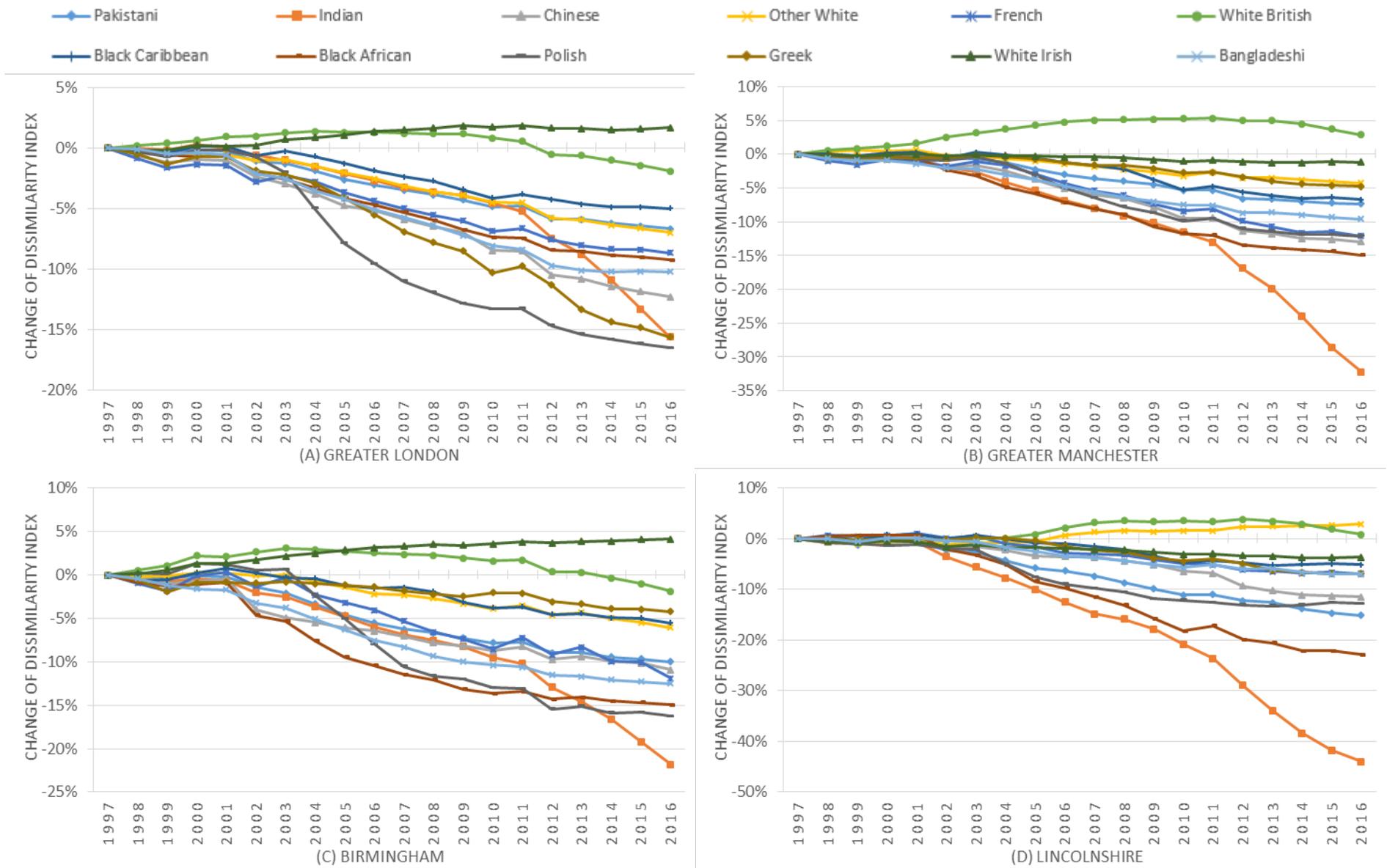


Figure 2. Change of the Dissimilarity Indices with respect to 1997 by ethnic groups in the case study areas, 1998-2016

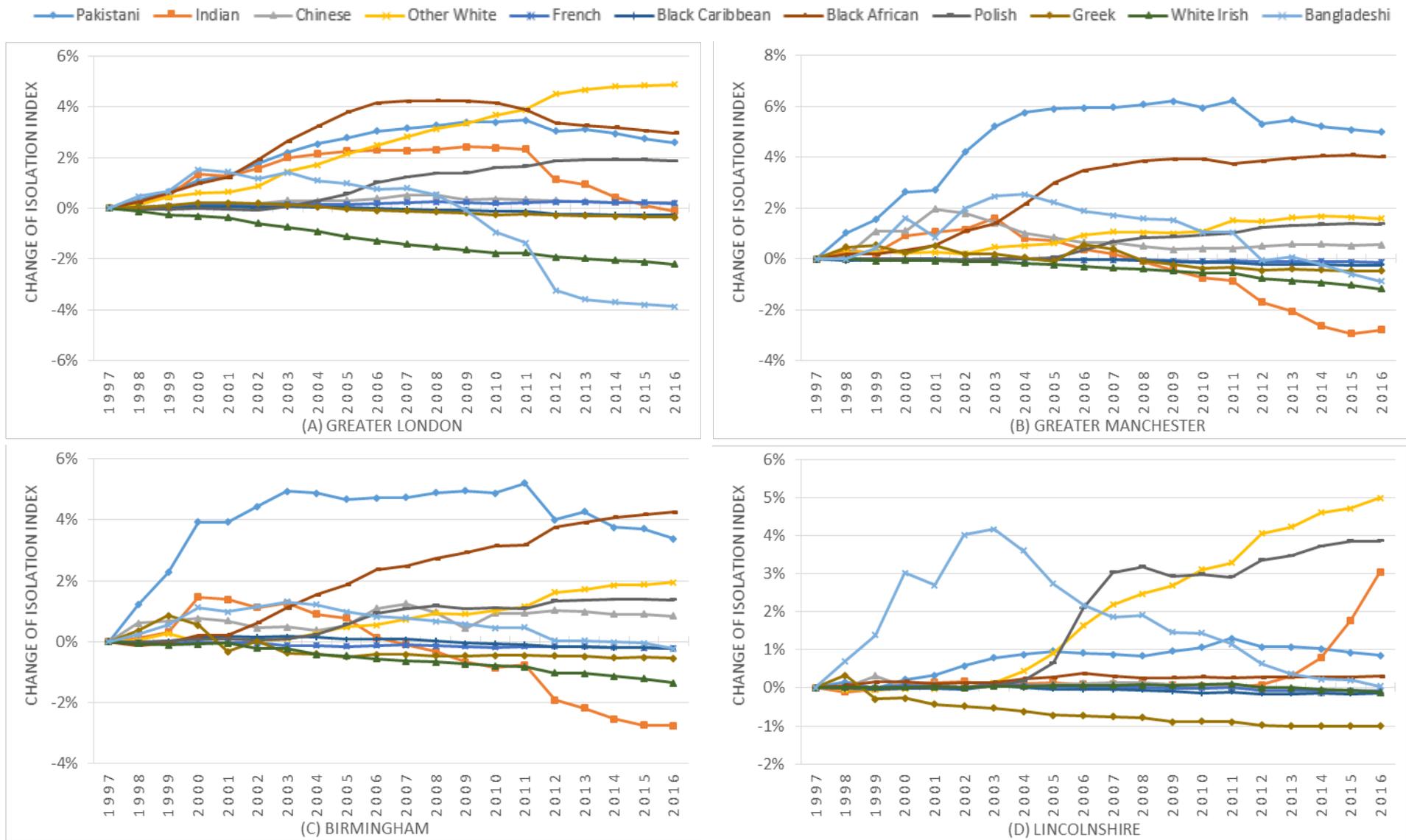


Figure 3. Change of the Isolation Indices with respect to 1997 by selected ethnic groups of the case study areas, 1998-2016

382 We use 2001 and 2011 Census population figures broken down by ethnic groups as
 383 benchmarks and examine the changes in the composition of the Consumer Registers in the
 384 intervening years (see Figure 4). These changes in the four case study areas generally
 385 correspond to the national trend of the ethnic composition summarised in Table 1 but with
 386 local variations. Mostly, the proportions of ‘White British’ and ‘White Irish’ have decreased
 387 in the four case study areas while others have increased. Increases in the proportions of
 388 ‘Other White’ groups in London, Manchester and Lincolnshire, are very apparent, as are
 389 those for Pakistanis in Manchester and Birmingham. The intercensal population changes by
 390 ethnic groups are largely mirrored well in Consumer Registers in terms of ethnic composition
 391 change between 2001 and 2011, except for the anomalous result for the Chinese in London
 392 and Caribbeans in both London and Birmingham.

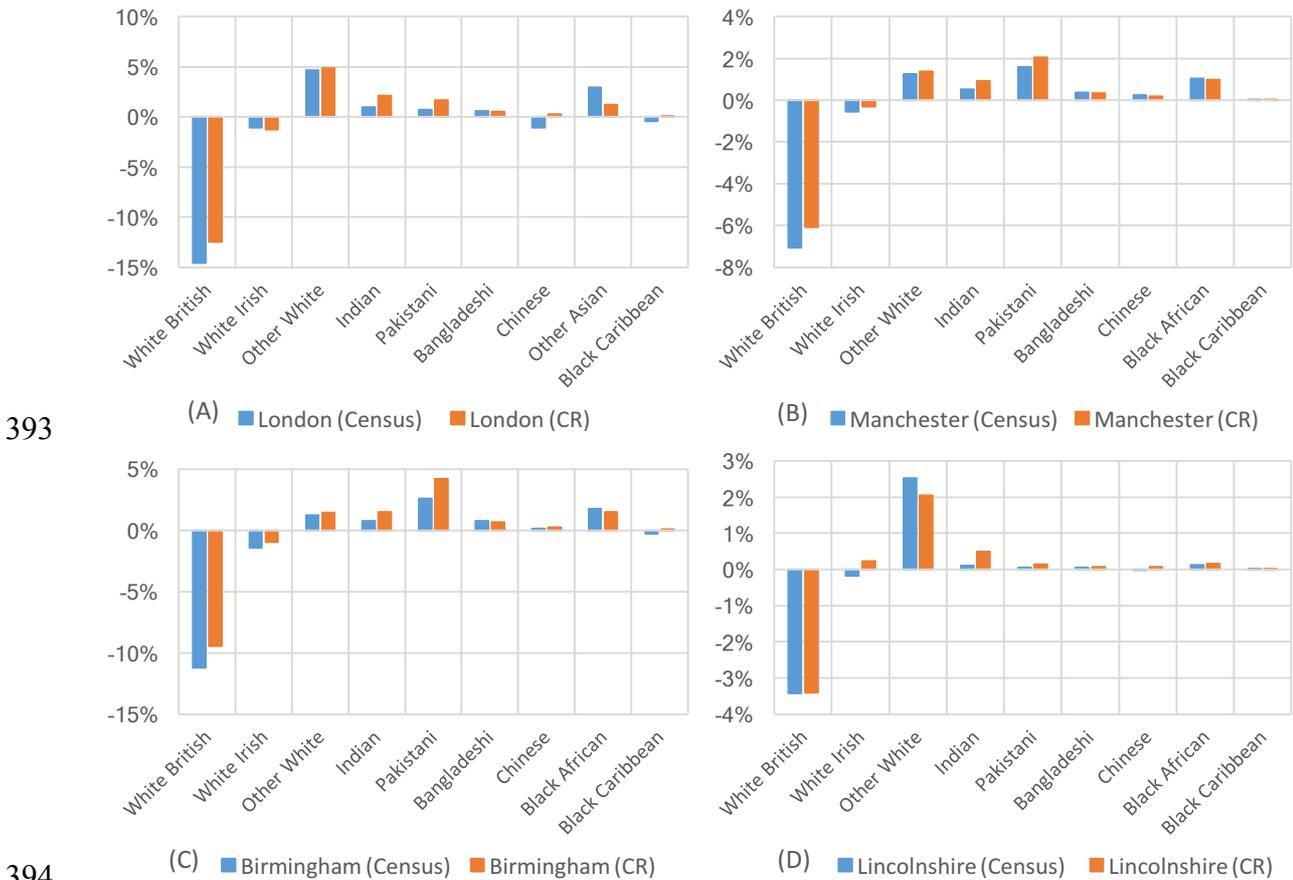
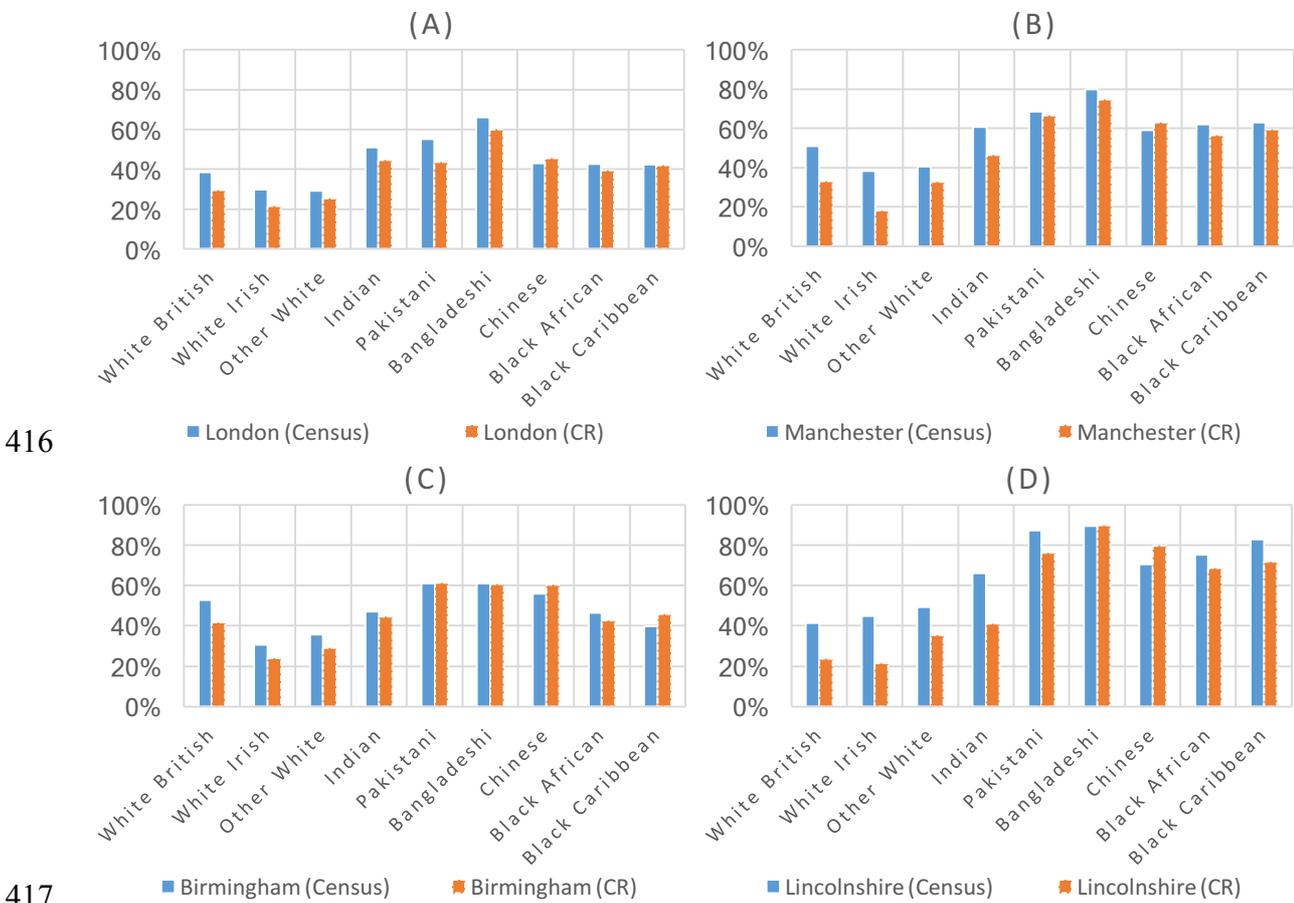


Figure 4. Changes in ethnic composition between 2001 and 2011 as estimated from the Census and Consumer Registers in London (A), Manchester (B), Birmingham (C) and Lincolnshire (D). (Source: Authors’ calculations using the 2001 and 2011 Census release.)

395 We also calculate Dissimilarity Indices of the four case study areas using the adult population
 396 from the Consumer Registers and compare them to indices calculated from the whole
 397 population recorded in the 2011 and 2001 Censuses (see Figure 5 and Figure S1 in the
 398 supplementary material respectively). The magnitude and trend of the segregation levels of
 399 each ethnic group largely correspond between Consumer Registers and Censuses with some
 400 nuanced differences. The datasets consistently show that some ethnic minorities, for instance,
 401 Bangladeshis, Pakistanis, ‘Black Africans’ and ‘Caribbeans’ are more segregated than ‘White
 402 British’, ‘White Irish’ and ‘Other White’. A comparison of Figure 5 and Figure S1 in the

403 supplementary material suggests that segregation levels of most groups have been decreasing
 404 in both data sources.

405 Despite the overall correspondence in segregation measurements from Consumer Registers
 406 and from the Census, we notice the degree of correspondence varies among ethnic groups and
 407 locations. Segregation levels of most ethnic groups have been underestimated in the
 408 Consumer Registers, compared with the two Censuses, except for the Chinese. In addition,
 409 the Dissimilarity Indices for both datasets agree more in urban than rural areas. The apparent
 410 discrepancies between the segregation indices calculated from the Consumer Registers and
 411 those from the Censuses can be partly attributed to the non-identical populations under
 412 investigation. Direct comparison between Consumer Registers and published small area
 413 census statistics is not possible since individuals aged under 17 cannot be omitted from the
 414 latter: however, with this caveat, the degree of correspondence between the Dissimilarity
 415 Indices from the two different data sources is generally high.



416
 417
 Figure 5. Comparison of Dissimilarity Indices in 2011 as estimated from Census and Consumer Registers for Greater London (A), Greater Manchester (B), Birmingham (C) and Lincolnshire (D).
 (source: Author’s own calculations with the 2011 Census data)

418 **7. Discussion: Uncovering nuanced segregation dynamics**

419 The analysis reported in this paper breaks new ground in both methodological and
 420 substantive terms. The refined, names-based ethnic classifications of the spatially precise and
 421 longitudinal linked Consumer Registers combined generates a powerful resource for
 422 segregation studies.
 423

424 7.1. Advantages of Consumer Registers

425 Our case studies
426 demonstrate the flexibility and necessity of accessing finer granular categorisations, as
427 illustrated by the noticeable disparities of segregation levels and trends among French, Greek,
428 Polish and ‘Other White’ groups. Ethnic segregation levels depend on the granularities of the
429 ethnic classifications under investigation. Therefore, similar heterogeneities can be
430 anticipated for other Census groups, for example, ‘Other Asian’, ‘Other Black’ or ‘Mixed’.
431 Such groups are unlikely to share close cultural affinity and hence common clustering
432 tendencies.

433
434 Consumer Registers and the name-based classifications in this study can largely lift
435 restrictions of pre-defined ethnic categorisations based on evidence-based research by
436 offering more flexibility in aggregating or disaggregating ethnic groups. The census ethnic
437 groups were developed for the main users of census data, and are not versatile enough to
438 target specific ethnic groups of interest for all academic researchers and policymakers. For
439 instance, Coombes et al. (2007) have investigated whether the early A8² immigrants to
440 England went to areas of labour shortage using the Worker Registration Scheme dataset.
441 They have also acknowledged the data limitations and pointed out there was no way to
442 estimate the numbers of A8 immigrants in these areas at one point in time.

443 In substantive terms, the temporal granularity of the analysis illustrates how trends of
444 segregation may differ over the short-term, and these differences can uniquely be detected in
445 the Consumer Registers. For instance, Consumer Registers clearly show the accelerated
446 increase in the Polish population and the decrease in the Dissimilarity Index immediately
447 after the 2004 European Union enlargement. Another finding is the increasingly even
448 distribution for the Indian population since 2011, which might otherwise lay undiscovered
449 until the results of the 2021 Census are published.

450 Apart from the more nuanced segregation trends and patterns revealed by our analysis of
451 Consumer Registers, our findings further support those who assert that Britain is *not*
452 experiencing an increase in ethnic segregation as a corollary to increasing ethnic diversity.
453 On the contrary, we find that segregation levels are steadily reducing for most ethnic groups
454 and that England and Wales have therefore become both more ethnically diverse and more
455 residentially mixed over time. The populations of most ethnic groups are growing in both
456 urban and rural areas. Larger proportions of ethnic minorities, for instance the Indians, are
457 now observed in rural areas, which suggests that there is a process of dispersion beyond
458 metropolitan regions.

459 This latter finding resonates with the chain migration process (Catney, 2015), which denotes
460 the process by which earlier immigrants begin to move away from metropolitan gateway
461 areas while subsequent immigration continues to settle in a wider set of urban cores.
462 Observations from the exposure dimension suggest that increased evenness for some ethnic
463 groups does not necessarily accompany increased exposure. Contrary changes in the two
464 measurements of individual ethnic groups, such as those observed in London, may indicate
465 significant immigration involving one or more ethnic groups.

² The eight countries joined the European Union during the 2004 enlargement: Czech Republic, Estonia, Hungary, Latvia, Lithuania, Poland, Slovakia, and Slovenia.

466 7.2. Uncertainties and limitations

467 We note the nuanced inaccuracies arising from measuring segregation using Consumer
468 Registers by validating our ethnicity estimates with reference to 2001 and 2011 Census data.
469 We identify two major sources of mismatches: the dataset representativeness (e.g. age bias
470 and voting registry eligibility); and systematic bias arising from the use of the Ethnicity
471 Estimator algorithm (e.g. the underestimation of Caribbeans). Despite these discrepancies,
472 values of the Dissimilarity Index and their ranks from Consumer Registers exhibit strong
473 positive correlation with those obtained using Census data. The randomisation tests suggest
474 that all of the Dissimilarity Indices are significant set against the null hypothesis of
475 randomness. With the caveats of certain bias attributable to the ways in which they are
476 assembled, Consumer Registers appear to be a promising supplementary source to, rather
477 than a substitute for, Census data.

478 Additional uncertainty of ethnicity estimates arises because the provenance of the different
479 consumer data sources used to augment the Electoral Registers with non-voters over the 20
480 year period is unknown, and the potential sources and operation of bias arising from opt out
481 from the public Electoral Roll post 2003 is also unknown. The methods developed by
482 Lansley et al (2019) promise to address this in parts, but more research is necessary to fully
483 establish the extent of bias.

484
485 Another broad issue is that ethnicity is only inferred from given- and surname pairings, albeit
486 in part using procedures that are more sensitive to the vagaries of self-assignment of identity
487 than purely algorithmic procedures. The merits of names-based analysis would be much
488 reduced were the focus of analysis upon segregation of individuals from ‘New World’
489 countries. However, since naming conventions here bear a less clearly identifiable
490 correspondence with geographic origins. Set against these issues, the use of algorithmic
491 procedures to disaggregate the ethnic categories used in UK censuses allows consideration of
492 more classes than is possible through census analysis. Consumer Registers also bring greatly
493 enhanced temporal granularity, in that they are updated in real time, crystallised into annual
494 incremental updates.

495

496 8. Concluding remarks

497 The motivation of this study has been to offer a more granular and comprehensive picture of
498 recent segregation trends and to demonstrate the feasibility of revisiting the topic of ethnic
499 segregation using a novel data source: Consumer Registers. Names-based ethnic
500 classifications applied to consumer data offers an innovative and powerful way to identify
501 nuanced patterns of and trends in segregation. Names remain an under-exploited resource in a
502 variety of applications. In particular, the flexibility of defining finer categories of ethnicity
503 produces detailed representations of the widely established two dimensions of ethnic
504 residential segregation. This method can be extended to explicitly spatial investigations of
505 segregation in future research and has the potential to enhance our understanding of ethnic
506 segregation change both in space and over time (see e.g. Lan et al 2019). Removing the
507 constraint of aggregation to Census Output Areas, future research could reconceive
508 segregation as a problem of point pattern analysis subject to restrictions of disclosure control.
509 In view of the high degree of correspondence of segregation patterns with the Census,
510 Consumer Registers are promising resources to uncover new and nuanced dynamics of the
511 complex phenomenon of segregation.

512 **References:**

- 513 Aspinall PJ. (2002) Collective Terminology to Describe the Minority Ethnic Population: The
514 Persistence of Confusion and Ambiguity in Usage. *Sociology* 36: 803-816.
- 515 Bakens J and Pryce G. (2018) Homophily horizons and ethnic mover flows among homeowners in
516 Scotland. *Housing Studies*: 1-21.
- 517 Boisso D, Hayes K, Hirschberg J, et al. (1994) Occupational segregation in the multidimensional
518 case: Decomposition and tests of significance. *Journal of Econometrics* 61: 161-171.
- 519 Cantle T. (2001) Community Cohesion: A Report of the Independent Review Team. In: Home Office
520 (ed). Home Office, United Kingdom.
- 521 Cantle T and Kaufmann E. (2016) *Is segregation on the increase in the UK?* Available at:
522 [https://www.opendemocracy.net/wfd/ted-cantle-and-eric-kaufmann/is-segregation-on-](https://www.opendemocracy.net/wfd/ted-cantle-and-eric-kaufmann/is-segregation-on-increase-in-uk)
523 [increase-in-uk](https://www.opendemocracy.net/wfd/ted-cantle-and-eric-kaufmann/is-segregation-on-increase-in-uk).
- 524 Carrington WJ and Troske KR. (1997) On Measuring Segregation in Samples with Small Units.
525 *Journal of Business & Economic Statistics* 15: 402-409.
- 526 Casey DL. (2016) The Casey Review: A review into opportunity and integration. In: Department for
527 Communities and Local Government (ed). London.
- 528 Catney G. (2015) Exploring a decade of small area ethnic (de-)segregation in England and Wales.
529 *Urban Studies* 53: 1691-1709.
- 530 Catney G. (2016) The Changing Geographies of Ethnic Diversity in England and Wales, 1991–2011.
531 *Population, Space and Place* 22: 750-765.
- 532 Coombes M, Champion T and Raybould S. (2007) Did the Early A8 In-migrants to England go to
533 Areas of Labour Shortage? *Local Economy* 22: 335-348.
- 534 Electoral Commission. (2019). *2019 report: Accuracy and completeness of the 2018 electoral*
535 *registers in Great Britain*. Retrieved from London:
536 [https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-](https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/our-research/accuracy-and-completeness-electoral-registers/2019-report-accuracy-and-completeness-2018-electoral-registers-great-britain)
537 [research/our-research/accuracy-and-completeness-electoral-registers/2019-report-accuracy-](https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/our-research/accuracy-and-completeness-electoral-registers/2019-report-accuracy-and-completeness-2018-electoral-registers-great-britain)
538 [and-completeness-2018-electoral-registers-great-britain](https://www.electoralcommission.org.uk/who-we-are-and-what-we-do/our-views-and-research/our-research/accuracy-and-completeness-electoral-registers/2019-report-accuracy-and-completeness-2018-electoral-registers-great-britain)
- 539 Finney N and Simpson L. (2009) *'Sleepwalking to segregation'? Challenging myths about race and*
540 *migration*: Policy Press at the University of Bristol.
- 541 Harris R and Johnston R. (2018) Measuring and modelling segregation – New concepts, new methods
542 and new data. *Environment and Planning B: Urban Analytics and City Science* 45: 999-1002.
- 543 Harris R and Owen D. (2018) Implementing a Multilevel Index of Dissimilarity in R with a case study
544 of the changing scales of residential ethnic segregation in England and Wales. *Environment*
545 *and Planning B: Urban Analytics and City Science* 45: 1003-1021.
- 546 Johnston R, Poulsen M and Forrest J. (2007) The Geography of Ethnic Residential Segregation: A
547 Comparative Study of Five Countries. *Annals of the Association of American Geographers*
548 97: 713-738.
- 549 Johnston R, Poulsen M and Forrest J. (2016) Ethnic Residential Patterns in Urban England and Wales,
550 2001–2011: A System-Wide Analysis. *Tijdschrift voor economische en sociale geografie* 107:
551 1-15.
- 552 Kandt J and Longley PA. (2018) Ethnicity estimation using family naming practices. *PLOS ONE* 13:
553 e0201774.
- 554 Lan T, Kandt J and Longley P. (2018) Ethnicity and Residential Segregation. In: Longley P, Cheshire
555 J and Singleton A (eds) *Consumer Data Research*. London: UCL Press, pp. 71-83.
- 556 Lan, T., Kandt, J. & Longley, P. (2019). *Geographic scales of residential segregation in English*
557 *cities*, *Urban Geography* Aug 2019, 1-21.
- 558 Lansley G, Li W and Longley PA. (2019) Creating a linked consumer register for granular
559 demographic analysis. *Journal of the Royal Statistical Society: Series A (Statistics in Society)*.
560 doi:10.1111/rssa.12476

- 561 Longley P, Adnan M and Lansley G. (2015) The Geotemporal Demographics of Twitter Usage.
562 *Environment and Planning A* 47: 465-484.
- 563 Massey DS and Denton NA. (1988) The Dimensions of Residential Segregation. *Social Forces* 67:
564 281-315.
- 565 Mateos, Pablo, Longley, Paul A., & Sullivan, David O. (2011). Ethnicity and Population Structure in
566 Personal Naming Networks. *PLOS ONE*, 6(9).
- 567 Nathan, Max. (2015). Same Difference? Minority Ethnic Inventors, Diversity and Innovation in the
568 UK. *Journal of Economic Geography*, 15(1), 129-168.
- 569 Office for National Statistics. (2009). *Final recommended questions for the 2011 Census in England*
570 *and Wales: Ethnic group*. Retrieved from
571 https://www.ons.gov.uk/file?uri=/census/2011census/howourcensusworks/howweplannedthe2011census/questionnairedevelopment/finalisingthe2011questionnaire/final-recommended-questions-2011-ethnic-group_tcm77-183998.pdf
573
- 574 Office for National Statistics. (2012) Methods used to revise the national population estimates for
575 mid-2002 to mid-2010. London: Office for National Statistics.
- 576 Peach C. (2009) Slippery Segregation: Discovering or Manufacturing Ghettos? *Journal of Ethnic and*
577 *Migration Studies* 35: 1381-1395.
- 578 Peach C. (2010) 'Ghetto-Lite' or Missing the G-Spot? A Reply to Johnston, Poulsen and Forrest.
579 *Journal of Ethnic and Migration Studies* 36: 1519-1526.
- 580 Poulsen M, Johnston R and Forrest J. (2011) Using Local Statistics and Neighbourhood
581 Classifications to Portray Ethnic Residential Segregation: A London Example. *Environment*
582 *and Planning B: Planning and Design* 38: 636-658.
- 583 Simpson L. (2004) Statistics of Racial Segregation: Measures, Evidence and Policy. *Urban Studies*
584 41: 661-681.
- 585 Simpson L. (2007) Ghettos of the mind: the empirical behaviour of indices of segregation and
586 diversity. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* 170: 405-
587 424.
- 588 Simpson L, Jivraj S and Warren J. (2016) The stability of ethnic identity in England and Wales 2001–
589 2011. *Journal of the Royal Statistical Society. Series A, (Statistics in Society)* 179: 1025-1049.
590
591