Making a spatially and temporally consistent country of birth dataset for Great Britain

Dan Olner[*1] and Gwilym Pryce[†1]

[1]Sheffield Methods Institute, University of Sheffield

January 11th 2018

Summary

We present a novel linkage of Census Data for Great Britain, designed to maximise the spatial and temporal consistency of country of birth analysis. In this paper, we first discuss the process of creating the dataset before examining two analyses the data has so far been used for: spatial stability of migrant location choice over time, building on the work of Geoff Meen; and the effect of cultural distance between countries on migrant location and segregation.

KEYWORDS: Census, country of birth, migration, data linkage, cultural distance

1. Introduction

We present a novel linkage of Census Data for Great Britain, designed to maximise the spatial and temporal consistency of country of birth analysis. In this paper, we first discuss the process of creating the dataset before examining two analyses the data has so far been used for:

1. Spatial stability of migrant location choice over time, building on the work of Geoff Meen (2016)
2. The effect of cultural distance between countries on migrant location and segregation.

2. Creating the dataset

The project's dataset has two parts: a three-Census tranche (1991 to 2011) containing thirty-three consistent country-of-birth categories and a five-Census tranche (1971 to 2011) with nine country of birth categories.

The aim of this project was to create as consistent a dataset for country of birth as possible, both spatially and over time. This requirement determined choices made about how to link data between Censuses. The first goal was to identify a list of countries or country groups that can be made consistent between decades. No two Censuses use exactly the same 'country of birth' fields.

The 'foundation' for this linkage is the country of birth table from the 1991 'Local Base Statistics' (LBS) data, because it has a much larger number of country of birth categories than the other available 1991 Small Area Statistics (SAS) table - so in order to maximis possible matches in other decades, this is the logical choice. However, LBS tables have some zone values set to zero and their counts added to contiguous zones if their count is below a certain threshold. To address this, we identified where these zones had their counts re-assigned by comparing to the correct SAS values at the smallest scale. This allowed us to create larger joined zones with the correct counts - in the process, creating a newer version of the LBS geography for Great Britain with a slightly smaller number of zones but no zero counts (figure 1). For Scotland, the LBS table data uses postcode sectors. In England and Wales, wards are

[*] d.olner@sheffield.ac.uk
[†] g.pryce@sheffield.ac.uk

used. We join these and process the zero count zones to produce a single geography for Great Britain.

Figure 1: example of collating 1991 'Local Base Statistics' zones to deal with zero-set zones, for central Glasgow postcode sectors. Green zones with re-assigned counts are combined with the correct zero-count (lighter green) neighbours to make larger zones.

This new 1991 LBS geography was then used as the common zone for all other decades. For other decades, smaller zones are aggregated into these larger LBS geographies, with the majority being entirely contained within larger zones.

3. Spatial stability

We have been using the dataset to ask - building on Geoff Meen's work in London - how migrant populations in previous decades determine later migrant location patterns, at present with a focus on the difference between richer and poorer countries of birth. The model used is described in equation (1): the proportion of people from country of birth **i** in zone **j** at time **t** (2011 in our analysis) is a function of the same proportion for an earlier time period; an average of this proportion from neighbouring zones (via W, a weights matrix); the sum of all other non-UK people in **j** minus the count of **i**; and a vector of control variables including unemployment, dwelling density and, for Scotland, house prices back to 1991 via Registers of Scotland data.

$$x_{ijt} = \gamma_0 + \gamma_1 x_{ijt-k} + \gamma_2 W x_{ijt-k} + \gamma_3 \sum_{r \neq i}^{I} x_{rjt-k} + \gamma_4 \mathbf{Z}_{jt} + \epsilon_{ijt} \qquad (1)$$

Results so far indicate a stronger spatial stability for people from poorer countries of birth, compared to richer.

4. Cultural distance

Data from Inglehart and Welzel's World Values Survey (e.g. Inglehart 2003) has been used to analyse the effect of cultural distance on migration (e.g. Tadesse & White 2010). Inglehart and Welzel create two principle components from their survey data to produce a two-dimensional map of countries' cultural 'location' relative to each other. Cultural distance is then simply the euclidean distance between these cultural map points.

There is a twenty country match between the cultural distance data and our Great Britain country of

birth dataset. For these twenty countries, we are analysing the impact of cultural distance on migrant location choice. Early results of this linkage suggest cultural distance may be a useful lens through which to examine spatial location choices. Figure 2 illustrates this for London: here, cultural distance of each non-UK-born person to the UK is found in each ward and the average taken. Darker blue zones are culturally closer to the UK, dark red are culturally more distant. The dark blue cluster is a strongly European area of London.

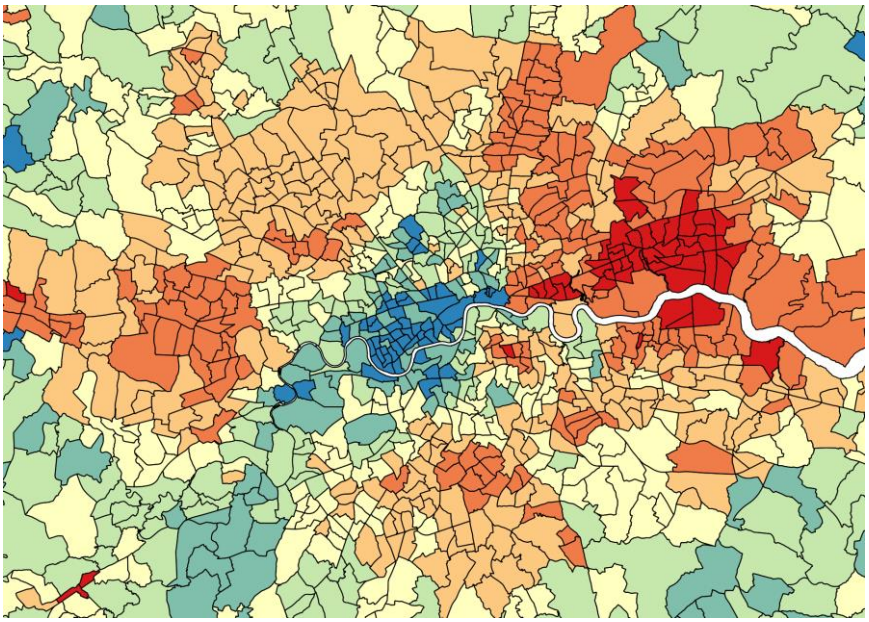

Figure 2: average cultural distance from UK residents in London LBS zones. Red zones have the largest cultural distance, blue zones the smallest. The blue cluster is a strongly European area of London.

In order to examine whether cultural distance has a role in spatial location choices, we compare the cultural distance between every country pair with the index of dissimilarity (DI) for those two countries. DI values give a proportion between 0 and 1: they indicate what proportion of either country would have to move zone in order to create a perfectly even mix across all zones. Thus higher values indicate greater segregation.

Figure 3 gives an illustrative result from this, for Australia. The y axis shows other countries' cultural distance from Australia; for the same pairs, the left axis shows the dissimilarity index between them. This positive linear pattern is particularly prominent for European countries, suggesting that the use of cultural distance to think about spatial patterning of migration is a useful avenue to pursue.

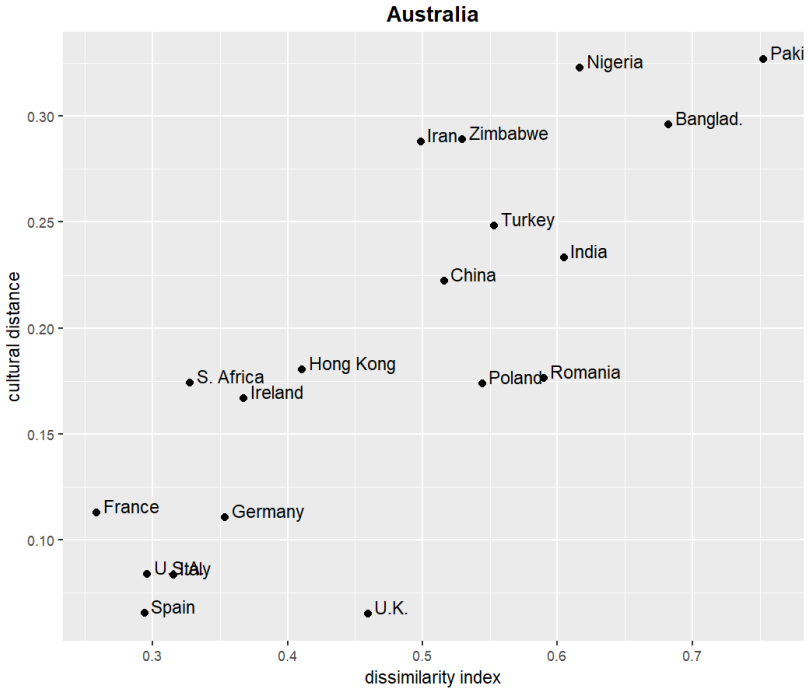

Figure 3: Australia's cultural distance from other countries (y axis) compared to the index of dissimilarity between those two countries. As spatial differentiation increases, so does cultural distance.

We are now in the process of adding cultural distance to the spatial stability model; early results suggest an expected outcome - that countries more culturally distant from the UK tend to be more spatially stable than others. This is using just cultural distance to the UK: next steps include using average cultural distance for each person, to every other zone neighbour, both including and excluding the UK, to test whether non-UK-born residents have a stronger effect than each zone's total population.

Challenges remain for the country of birth dataset. It requires more work to define the bounds of geographical consistency (cf. e.g. Norman *et al* 2003). Further development is needed of the underlying model of internal and external migration to better explain location choice given the data only describes the count at each Census point. Nevertheless, this is a unique new data linkage shaped around the need to analyse country of birth over time in Great Britain.

The dataset will be made publicly available via the Urban Big Data Centre and Sheffield University's ORDA service in the near future.

5. Acknowledgements

We acknowledge the ESRC's funding for the Urban Big Data Centre, Glasgow University, who fund this project.

6. Biography

Dan Olner is a researcher for the Urban Big Data Centre in Glasgow, based in the Sheffield Methods Institute at Sheffield University. With a background in spatial economics and politics, Dan has recently worked on projects including examining the effect of windfarms on Scottish house prices and modelling

migration.

Gwilym Pryce is Co-Director of the ESRC CDT Data Analytics and Society and Professor of Urban Economics and Social Statistics in Urban Studies and Planning at Sheffield University. His core research interests are in urban economics, and most of his research publications have been on housing and mortgage markets.

References

Meen, G., Gibb, K., Leishman, C., 2016. *Path Dependence, the Spatial Distribution of Immigrant Communities and the Demand for Housing*, in: Housing Economics: A Historical Approach. Palgrave Macmillan, London.

Inglehart, R., 2003. *Human Values and Social Change: Findings from the Values Surveys*. Brill, Leiden.

Norman, P., Rees, P., Boyle, P., 2003. Achieving data compatibility over space and time: creating consistent geographical zones. International Journal of Population Geography 9, 365–386

Tadesse, B., White, R., 2010. *Cultural distance as a determinant of bilateral trade flows: do immigrants counter the effect of cultural differences?* Applied Economics Letters 17, 147–152.