Consumer Data Research Centre

An ESRC Data Investment

CACI

Representing Population Dynamics from Administrative and Consumer Registers

Guy Lansley, Wen Li and Paul Longley
University College London

Impacts

- Derives metrics of population dynamics through new forms of data.
- Utilises consumer and administrative data sources to create representative data on annual population changes at a small area level.
- Estimates household changes and migration flows from changes between annual registers.
- Provides a framework for the integration and validation of future consumer data products.

Project Background

There is an abundance of data on the population, which are routinely collected by public and private organisations. These data are crucial for decision-making processes in numerous areas such as urban management, retail, transport planning and informing government policies. However, they are rarely analysed or repackaged as public datasets. The decennial national Census seeks to attain universal coverage and is still considered the most reliable source of population data, yet it provides very infrequent snapshots of population change and is of limited use for understanding annual population dynamics.

This research seeks to derive representative metrics of population dynamics through new forms of data. Two composite Consumer Registers were obtained from CACI Ltd that contained the public version of the UK Electoral Register and individual records from a number of commercial organisations. These databases comprise of names and addresses for the vast majority of the UK adult population. Through appropriate data cleaning and linkage techniques, this study shows how it is possible to match addresses and record changes in their size and composition over a two-year period.

Applying these data for such purposes creates uncertainties, due to the unknown coverage achieved by different data collection procedures and inconsistent unique IDs between datasets leading to problematic linkage. Here, heuristics are presented demonstrating that it is possible to develop precise linkages at the levels of the individual and household in order to firstly, estimate household change and secondly, estimate migration flows from changes between annual registers. This research shows how data pooled from consumer and administrative sources may be used to create representative data on annual population changes at a small area level.

Data and Methods

Composite registers for 2013 and 2014 were acquired for this study, both representing a very high proportion of the adult population in the UK (see Table 1). The methodology adopted two assumptions. Firstly, that individual records matched by address and full names across both registers were the same individual and this person had not changed address. Secondly, if a composition of residents (as identified by their full names) were not recorded at the same address in the subsequent register, but a household comprising of identical named individuals appeared at a single other address, then these records pertained to the same household.

Year	Population	Households
2013	54,380,747	27,114,152
2014	55,397,463	27,371,755

Table 1. Counts from the Consumer Registers.

Household Changes

To model household changes in the adult population between 2013 and 2014, addresses from both registers were joined using their full addresses and postcodes. For each matched address, the number of persons who appeared in both years using the residents' full names was recorded and a key created to identify residents sharing full names within a household. This approach could not correctly account for individuals whose names may have changed or been misreported. Therefore, further techniques were applied to identify married women, such as using forenames to identify gender and matching the surname of a male occupant from the same household. Double-barreled surnames were also found to be inconsistently recorded in the data, therefore a filter was devised to identify one of their singular surnames in the other year's data.

Migration

To estimate migration flows from changes between the two registers, records that were not found at the same address in both registers were identified and three databases created ("Stable Residents", "2013 Movers" and "2014 Movers"). For each address within each

Representing Population Dynamics from Administrative and Consumer Registers

Guy Lansley, Wen Li and Paul Longley
University College London

register, a unique key was created to represent the full composition of names within households in alphabetical order. Only household compositions that were unique within both movers databases were considered to ensure definitive matches. Pooled data from the ONS, NISRA, NRS and The Migration Observatory (indicating deaths, those coming of age and moves into the UK) estimated between 8 and 9 million individual records to change each year.

Key Findings

The household matching analysis revealed that 73.7% of households in 2013 remained identical in composition in 2014. The results are presented in Figure 1.

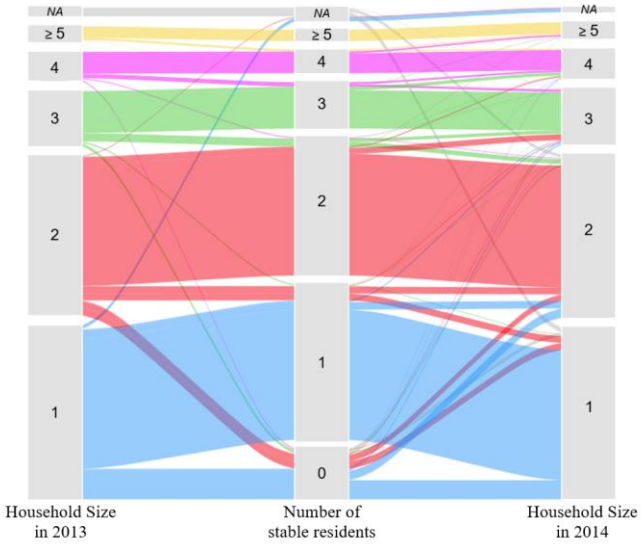

Figure 1. Alluvial plot of household changes between 2013 and 2014 (NA refers to addresses that were not in one of the datasets).

Figure 2 visualises the key trends in internal migration at the local authority level across Great Britain as estimated by the consumer registers. These analyses estimated approximately 46.1 million stable residents, 8.2 million present in 2013 but not 2014 and 9.2 million present in 2014 but not 2013. There are distinctive flows of migration between more populous districts and also many neighbouring areas. Over 490,000 of the moves were within the same local authority and 810,000 within the same region.

Despite only representing the first stage of data linkages, the model definitively identified the origin and destinations of 1.93 million individuals. This is a high figure considering that children are excluded from the data and further stages to improve the match rate have been devised (such as running partial joins using incomplete combinations of household members).

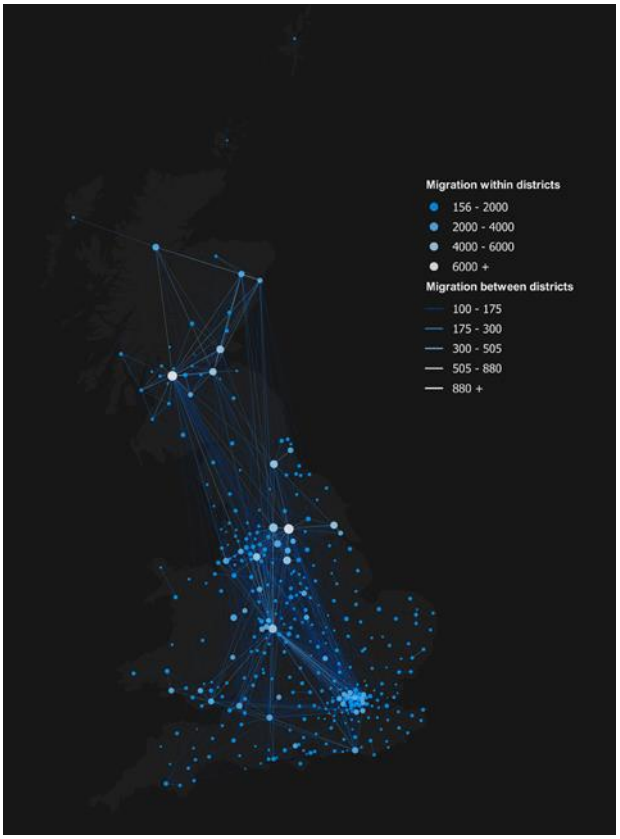

Figure 2. A representation of internal migration within Great Britain. Migration within local authorities are displayed as proportional circles whilst flows of over 100 persons between districts are represented as lines.

In addition, just over 78,000 households moved within the same postcode. The researchers consider that many of these occurrences could arise due to properties changing name and therefore not address matching correctly.

Future Directions

Whilst this only presents the preliminary stages of matching procedures, future research will aim to increase the proportion of internal migration flows that can be successfully matched between the registers. In addition, research is aiming to assign duplicated household name keys into origin-destination pairs based on the heuristics identified from the modelled results, and perform linkages to other data, such as price paid data from the Land Registry in England and Wales.