

Using Council Tax data extraction for mapping migration patterns

by Bob Line , independent housing consultant/researcher, and working with De Montfort and Nottingham Trent Universities.

Draft article for academic housing, GIS and local authority research journals

Local authority operational databases can be a rich source of useful information to help develop housing strategy, as they hold data on specific changes and transactions for large proportions of local populations and households.

For example Council Tax (CTax) databases, the operational part of the local authority computer systems used for tracking CTax accounts and payments, have records for something like 95% + of households in a district. This can be used to show, amongst other things, when households have moved, because when they do so their CTax account is ended or amended.

Extracting a query for accounts that have ended between two dates can give a good indication of how many households have moved, although care needs to be taken to avoid double counting because several operational changes may be carried out for just one move.

In addition, CTax accounts hold the current address and postcode of the household, and, because the CTax department is interested in pursuing any outstanding charges, in many cases they will also hold a forwarding address and postcode. This means that with an appropriate query both the origin and destination of the moving household can be obtained for mapping in a Geographic Information System.

Forwarding addresses generally seem to be recorded in around 50% of cases where a CTax account has been ended. It might be thought initially that the non recorded cases would be more likely to be poorer and less established households, but consultations with CTax operational staff suggests that it is much more random across social and economic circumstances. This perhaps partly helps support the economists' 80:20 rule of thumb – that 80% of people are honest and responsible, and 20% are less so, across all walks of life.

CTax forwarding postcodes will, however, unavoidably miss non CTax payers – including people leaving parental homes or institutional accommodation, so it will only give data on movements of existing households. Students, although they do not pay CTax, are often nevertheless recorded on the database as exempt and can be included in extracts..

Dirty data

However, there are difficulties and dangers. Other causes of changes, such as name changes or the death of the account holder, also need to be taken into account and cleaned out of the data. Another distortion is when CTax liability reverts to a Landlord when a property falls empty, - but most of these can be cleaned by stripping out liabilities periods of less than two months , which operational staff suggest will find most of the Landlord liability periods.

All other causes of dirty data also need to be carefully considered, and often some ingenuity is required to find ways to clean them out of the data. If they cannot be cleaned out operational staff can often make a general estimate of the extent of these distortions, and the amounts, if not the individual records, can be discounted.

Handling the data

Origin postcodes typically create a rash of dots across a whole district, but sometimes with clear concentrations in a particular local area . To make better visual sense of it the data can be aggregated into larger areas, such as postcode sectors or wards, in GIS, but it is important to note that administrative boundaries are not 'real' in terms of housing markets, neighbourhoods or sub markets , and often new boundaries created on the GIS map can give new and better insights into the patterns on the ground. (Map 1)

Forwarding postcodes show the locations that household from an area are most likely to move to, giving an indication of the housing market area, or sub areas, for that district.

Data Protection

It is crucial that Data Protection is taken very seriously in this approach, but for strategic analysis the full postcode is usually adequately detail. A full postcode is typically around 15 addresses, which gives sufficient anonymity, and when mapped in GIS the chances of identifying any individual is pretty remote. Extracting a unique reference number is useful for combining multiple counted changes, but at no stage in the data extraction are names or full addresses required at all.

When the areas selected get smaller Data Protection needs to be considered even more carefully, as the chances of recognising particular households could increase. While the data may be usable within the local authority for strategic purposes, disclosure, for example by publication, may not always be possible.

The Information Commissioner recognises and accepts the use of data for statistical and historical data for social research and planning in this way, and the 1998 Data Protection Act states:-

Research, History and Statistics

2.6.1 Section 33 of the Act provides for various exemptions in respect of the processing (or further processing) of personal data for research purposes (including statistical or historical purposes) provided that the processing (or further processing) is exclusively for those purposes and, also, that the following conditions are met:-

- *that the data are not processed to support measures or decisions relating to particular individuals, and*
- *that the data are not processed in such a way that substantial damage or substantial distress is, or is likely to be, caused to any data subject.*

The Compliance Manager for the Information Commissioner has confirmed again in 2003 that this is still correct.

By linking movers data with incomes and other demographic factors, through a database query at a geographically detailed level, the characteristics of moving households can also be extrapolated. This will depend on geographical homogeneity and variance within these small areas, as the datasets are not directly linked at an individual or household level. But we commonly use much cruder averages to reach policy conclusions, and increasing the level of detail can surely only improve the accuracy and relevance.

Migration patterns

The number of movers within and between areas is a first step towards an estimate for the gross flow of households, which is a key component in a more sophisticated method of estimating housing demands, supply and needs. (Whitehead and Kleinman, 1992).

The Universities of Newcastle upon Tyne, Leeds and the Greater London Authority (GLA) have also recently published research on the Development of a Migration Model for the ODPM, which uses data of migration flows between the 98 Family Health Service Authority (FHSA) areas of England and Wales recorded by the National Health Service Central Register (NHSCR).

The Migration Model develops the conceptualisation and mathematical modelling of migrations to a higher level of sophistication, using 100 origin/destination areas. However it may ultimately be limited by the level of detail of the data used on migrations to quite large geographical areas only. CTax data can give much more detail, but will require a more detailed and possibly different type of model, and because of the potential number of interactions between locations probably quite considerable computing power to run it.

A comparison between the Health Service data used in the Migration Model and the extracted CTax data for movements between adjacent authorities in Nottingham showed a considerably better than might be expected agreement between the two, with the proportions of movers to surrounding local authority areas from CTax data agreeing within two percentage points with the Health Service data.

2. What is a housing market area ?

Defining what actually constitutes a housing market area is by no means widely agreed or at all easy. Travel to Work areas are frequently proposed as representative of housing market areas. The DETR good practice guide (2000) suggests that a "Housing Market Area is the geographical area in which a substantial majority of those employed both live and work, and where most of those changing house without changing employment choose to stay". Jones (2001) proposes that it should be an area of self containment, within which a certain proportion of moves take place:- "(A) at least 50 per cent internal migration; or (B) in-migration from an adjacent HMA equivalent to less than 5 per cent of the market."

Another definition from the USA in the 70's (Suttles, 1972) was that a housing market 'is an area in which properties compete with each other'. This also revives the complex but intuitively compelling idea that housing markets are defined by the choices and trade offs of home seeking household; and, consequently, that sub areas within a wider geographical area may actually be excluded from 'their' market.

A more recent refinement of this sub market concept categorises housing 'product groups' as competing with each other, or having 'substitutability'. This leads to a conceptualisation of housing sub markets as dispersed and non contiguous. So households in specific groups, - which may be socio-economic, racial, or some other commonality -, will often consider living in several different sub areas, which may perhaps be some considerable distance apart, while excluding other sub areas nearby, based on the trade-off of factors which they consider most important.

Housing sub market patterns will therefore be dispersed and interlocking, often with fuzzy boundaries between them, within the wider market areas. These fuzzy boundary areas could also be of particular importance, as it may be in these that policy interventions have the greatest chance of success.

On this conceptualisation, defining housing markets may require spatial market search analysis, as argued by MacLennan (1992) . A common objection to this is that it is simply too complex to model, and that data on choices is not available and, moreover just cannot realistically be obtained. The failure of survey methods so far to fully capture housing market patterns and behaviour might appear to confirm this.

A possible methodological alternative may be to look at the dynamic patterns in the *detailed* data to try to identify substitutable product groups. Perhaps, for example, *changes* in house prices in sub markets over quite long periods would show similar sub market areas. The danger would be, however, that these changes are more driven by general changes, or run in parallel simply by coincidence. A further prerequisite for attempting this would be that existing, distorting, administrative boundaries - like wards - should be discarded for this purpose.

However, if the aggregated effects of myriad choices to define broad sub market areas can be researched - perhaps by data analysis or by interviews, questionnaires, focus groups, home seeker diaries and similar techniques, this may enable some progress towards combining spatial market search analysis with more empirical data modeling methods. Perhaps neither a purely subjective spatial search approach nor empirical data analysis alone can fully capture and understand how housing markets operate.

The CTax data mining technique does not resolve any of these definitional issues, but may give a more detailed source of empirical evidence with which to make progress on them. (See Map 2).

The technique can also be extended in GIS to show 'arrows' joining the origin and destination postcodes, to show the flows of movers. In MapInfo, this requires a couple of extra MapBasic tools - CoordinateExtractor.MBX and cre8line.mbx - both free from the web.

At postcode level this initially results in an unintelligible spaghetti-like mish-mash. (See Map 3). However, by selecting specific areas for the origins or destinations of movers, this overwhelming mass of data can be limited so that the underlying trends are clearer. (Maps 4 and 5).

The level of detail the technique produces can enable very localised patterns to be seen – who moves to this sheltered scheme, for example; or where do movers out of this unpopular estate generally go to. For local authority level housing, renewal and regeneration strategy it is often the very detailed patterns of migrations, influenced by local choices, preferences and especially avoidance of particular areas, which can be key to successful interventions. When visible in this way, policy and resource allocation can be much better targeted.

Local authorities have traditionally tried to assess demands and needs in their area by general 'housing needs surveys', effectively trying to count the number of households in need by a questionnaire with a postal or interview sample. While surveys will still have their place, these data mining and mapping techniques can provide an alternative, complementary and much more comprehensive, dynamic view of the patterns and flows in housing markets, which, because the data is already within the local authority, are also substantially cheaper to obtain and update than surveys.

References

Department of Environment, Transport and the Regions; (2000) 'Local Housing Needs Assessment: a guide to good practice'. Produced for the DETR by Heriot-Watt University.

Jones, C.; The Definition of Housing Market Areas and Strategic Planning; (2002); Urban Studies, Volume 39, No. 3, pp 549-564

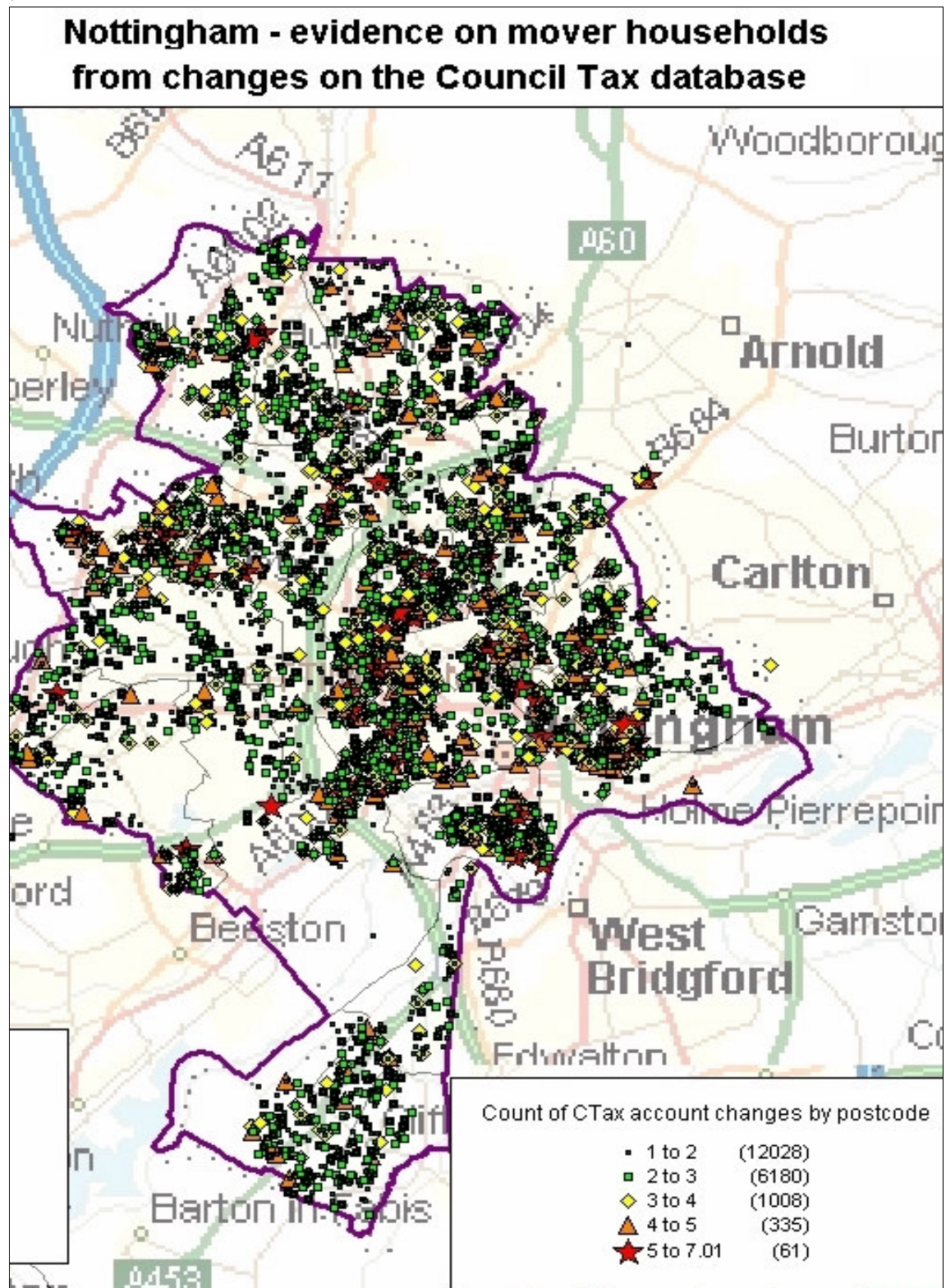
MacLennan, D. (1992). Housing search and choice in a regional housing system; new housing in Strathclyde. A report to the Housing Research Foundation for the Scottish House Builders Federation, University of Glasgow

Suttles G.D. (1972) The Social Construction of Communities. Chicago: Phoenix Press

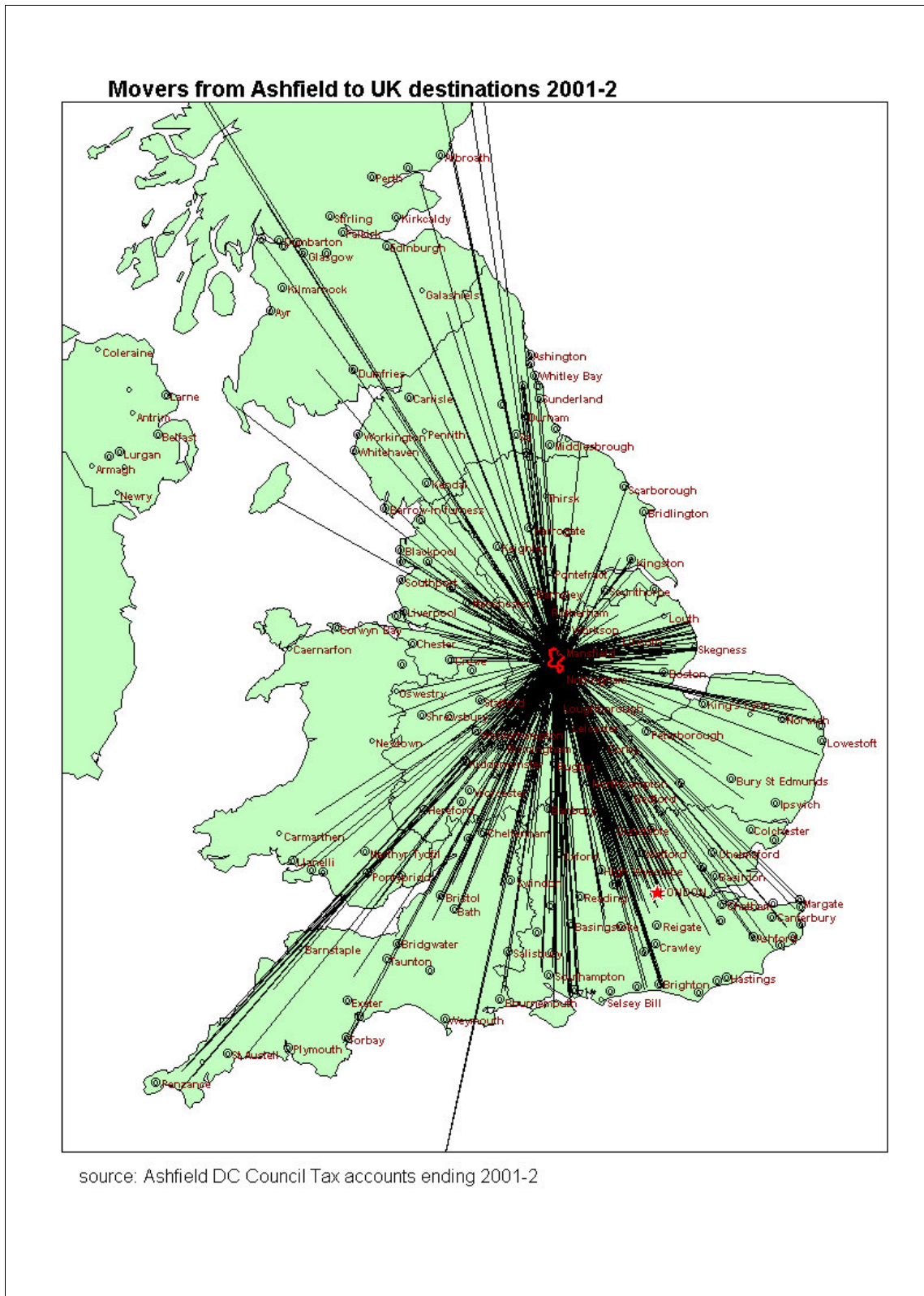
Universities of Newcastle upon Tyne, Leeds and Greater London Authority/London Research Centre (2002); Development of a Migration Model, London: Office of the Deputy Prime Minister. ODPM Housing Research Summary No.167 - <http://www.housing.odpm.gov.uk/hrs/hrs167/index.htm>. Full report <http://www.housing.odpm.gov.uk/information/migration/pdf/model.pdf>

Whitehead, C. and Kleinman, M. (1992); A review of housing needs assessments; London: Housing Corporation.

Map1 : Evidence on moving households from Council Tax account changes in Nottingham 2001-2.



Map 3 . UK destinations of movers from Ashfield DC 2001-2, derived from Council Tax accounts ending forwarding postcodes



Map 4. Movers from Nottingham City to Rushcliffe Borough area 2001-2

Movers to Rushcliffe form Nottingham City 2001-2

