Linking aspatial survey data to geodemographics classification: Using geodemographics classification as the local model for Spatial Microsimulation

# Introduction

The generation of synthetic population estimates through Spatial Microsimulation has been a popular technique in recent years. Static spatial microsimulation creates a synthetic population (a population built from anonymous sample data at the individual level) which realistically matches the observed population in a geographical zone for a given set of criteria. There is a diverse set of research and policy applications that use synthetic populations in a spatial setting, including: health (Brown & Harding 2002; Smith, Pearce, & Harland 2011; Tomintz, Clarke & Rigby 2008), transportation (see, for example, Beckman, Baggerly ,& McKay 1996; McFadden, Cosslett, Duguay, & Jung 1977) and water demand estimation (Williamson & Clarke 1996).

Much modelling in these disciplines takes an aggregate or meso-scale approach to the issues the spatial resolution. That is, characteristics of individuals or households are summed to provide zonal population that can be as large as entire cities and regions. Models built on these more aggregate data sets are widespread and have proved very fruitful in many areas of policy analysis. However, such modelling techniques often provide very little information concerning the inter-dependencies between household structure or type and their lifestyles. It also has long been argued that the most powerful theoretical models for explaining human behaviours operate at the individual person level, with emergent higher-level properties giving the best opportunity to understand the economic and spatial systems at all level (Orcutt, 1957). For social policy evaluation such micro models allow analysts to monitor the effects of changes in taxation, family credit, property or council tax, pensions, social security payments, etc. (the actions of local and national governments) at the household level (and hence at any more aggregate spatial scale). For area-based policy evaluation such models allow differential impacts between and within areas to be analysed more effectively. The necessity of predicting the impacts of social and area-based policies at the local or micro-level has also been emphasized by Openshaw (1995, p.60). ‘Governments need to predict the outcomes of their actions and produce forecasts at the local level’.

There are many national surveys, which provide detailed population data for countries, regions or other large geographic areas. However, given the fact that microdata is often not available to academics due to respondents’ confidentiality and collection issues, Spatial Microsimulation presents an opportunity to simulate individual or household behaviours at small-area level that can then simulate what-if type changes in order to test the impacts of certain policy. For these reasons Wilson (2000, p. 98) identified microsimulation as one the most important methods in regional science modelling: ‘Simulation is a critical concept in the future development of modelling because it provides a way of handling complexity that cannot be handled analytically. Microsimulation is a valuable example of a technique that may have increasing prominence in future research’.

Whilst this technique has become more popular, it is also becoming increasingly apparent that the geographical patterns produced by such techniques can look ‘too flat’ for certain variables, with not enough discrimination across a city or region. This can come about for three major reasons. First, the model is built by linking two data sets (survey and census data, for example) based on selected matching variables—so-called constraint variables. Variables not used as constraints (unconstrained variables) are thus likely to be less well matched. Second, household types that are least likely to be in the surveys (perhaps households of extreme affluence or poverty) are very difficult to reproduce across a city or region using reweighting techniques. In contrast, the attributes of households in ‘average’ type census wards tend to be cloned very effectively. Third, there may be geographical explanations for why key variables may be very different in similar kinds of housing areas. Thus, for example, two housing estates of similar type may have different car ownership levels if one has local services that are more easily reached on foot (the area thus enjoys a greater degree of ‘walkability’) or by public transport. However, little progress has been made to date in relation to the third problem.

The aim of this paper is thus to review the general issue of cloning household attributes in reweighting techniques and to begin to offer some solutions in relation to geographical issues which might explain real-life differences in the attributes of households which appear to be very similar through incorporating geodemographic classifications in Spatial Microsimulation. In addition, this study also aims to bridge the identified data gap between the published regional level estimates of expenditure and known drivers of local level variation by testing the model and situating the research in real-world practice. In the first section I will provide more depth into the problems of trying to reproduce spatial patterns on a small area level through providing an overall view on Small Area Estimation Methods. In the second section, I will provide the step-by-step procedure on the methods of Spatial Microsimulation, specifically on Iterative Proportional Fitting (IPF), with the real example from the data. For the third section, I will discuss my rationale of formatting microdata and constraints. In section four, I will present and compare microsimulation modelling and model fits between the spatial microsimulation with and without Geodemographics Classifications, which I then offer a new methodology for enhancing the performance of microsimulation models by adding geodemographic detail. Finally, I will discuss and analysis the internal and external validation of the model fits.

# Small Area Estimations Overview: Synthetic Reconstruction and Spatial Microsimulation

Small Area Estimation (SAE) has become an important technique in survey sampling due to the demand of providing reliable small area statistics for the target population and for any sub-population (Ghosh & Rao,1994). Traditionally, there are two types of small area estimation – Direct

# Iterative Proportional Fittings: How does it generate new weights?

# Explicit Numerical Solution with Iterative Proportional Fittings