an introduction to

Microsimulation for demography

Nik Lomax* University of Leeds
Andrew P. Smith University of Leeds

*Corresponding author. Email: n.m.lomax@leeds.ac.uk. Address: School of Geography and Leeds Institute for Data Analytics, University of Leeds, Leeds LS2 9JT, United Kingdom

Paper received 18 August 2017; accepted 18 October 2017; published 20 November 2017

Abstract

Background

Microsimulation consists of a set of techniques for estimating characteristics and modelling change in populations of individuals.

Aims

To demonstrate how microsimulation can be used by demographers who want to undertake population estimates and projections.

Data and methods

We use data from the 2011 United Kingdom (UK) Census of population to create a synthetic population by age, sex and ethnic group. Static and dynamic microsimulations and the visualisation of results are undertaken using the statistical package R. The code and data used in the static and dynamic microsimulation are available via a GitHub repository.

Results

A synthetic population in 2011 by age, sex and ethnicity was produced for the East London Borough of Tower Hamlets, estimated from two Census tables. A population projection was produced for each of these age, sex and ethnicity groups to 2021. We used a projection of the Bangladeshi population to visualise population growth by Middle-layer Super Output Area (MSOA) and to produce a population pyramid of estimates in 2021.

Conclusions

We argue that microsimulation is an adaptable technique which is well suited to demography, for both population estimation and projection. Although our example is applied to the East London Borough of Tower Hamlets, the approach could be readily applied in Australia, or any other country.

Key words

Microsimulation; static; dynamic; R code; demography; data visualisation; population estimation.

1. Introduction

This paper outlines how microsimulation can be of use to demographers as a method for the estimation and projection of populations. We outline what microsimulation is and how it has been used in a range of applications, from broader social science to demography specific examples. We provide two examples of how to apply microsimulation (with accompanying R code), firstly creating a synthetic population, and then projecting this population forward in time. While the examples presented in this paper relate to the United Kingdom (UK), the methods could be readily applied to Australia. We offer some examples of how the data can be visualised and analysed and a discussion of how the examples could be further extended.

2. What is microsimulation?

Microsimulation is an approach used to estimate the characteristics of a population from a range of attribute-rich data sources. *Micro* refers to individual units (people, households, etc.) while *simulation* refers to the process of assigning attributes for those units. Microsimulation can, for example, be used to allocate characteristics from a sample survey dataset to a larger population enumerated in a census.

Microsimulation can be used to create static or dynamic models. A static model is a way of synthesising data to produce an entire population estimate of individuals. Dynamic microsimulation introduces time into the model and the ability to 'age' the static population. It is dynamic microsimulation which is used in population projections and to perform 'what if' scenario testing of policy or market interventions. The technique is often referred to as *spatial microsimulation*. Essentially, the spatial element is concerned with allocating some form of geography to your individual or unit. We provide examples of static and dynamic microsimulation later in the paper. Both are inherently spatial because they estimate people within specific areas.

Microsimulation has a wide range of utilities across a number of disciplines. For example, microsimulation has been used to model future elderly health care demand (Clark et al. 2017), project educational attainment (Nelissen 1991), commuting patterns (Lovelace, Ballas and Watson 2014) and population projections (Ballas et al. 2005). Many of the applications of microsimulation we see in practice today stem from the work of Orcutt (1957) and Orcutt et al. (1961). Spielauer (2011) provides a good overview of microsimulation applications for the social sciences (including the differences between static and dynamic models); Li and O'Donoghue (2013) provide a comprehensive discussion of dynamic microsimulation models implemented across a range of settings. For a comprehensive overview of spatial microsimulation see Birkin and Clarke (2011).

Ballas et al. (2005, p. 15) contend that microsimulation fundamentally is not that complicated an approach, noting that 'even calculating simple social statistics such as life expectancy is a method that is similar to microsimulation'. When it comes to more complicated dynamic applications, Bélanger and Sabourin (2017) observe that microsimulation is an approach which has become increasingly popular with improved computing power and better availability of survey microdata.

3. Why is microsimulation a useful technique for demographers?

As a technique, demographers can make use of microsimulation to estimate and project populations and their attributes. In an age where more information is becoming available from disparate sources (e.g. censuses, surveys, administrative or commercial data), there is a need to synthesise this information if we wish to estimate populations with a range of attributes. The need for this attribute-rich information stems from the fact that we can produce more accurate models and make better recommendations if we know more about each individual.

Bélanger and Sabourin (2017, p. xix) argue that micro models are superior to macro models because 'the most powerful theoretical models for explaining human behaviour, such as decisions to have a child or migrate, operate at the level of the individual'. For a balanced assessment of microsimulation *vis-a-vis* macrosimulation models for population projection see van Imhoff and Post (1998, p. 134) who conclude that, while there are limitations (largely related to difficulty in implementation) 'microsimulation can do certain things that macrosimulation cannot'. One of the key advantages they identify is the ability of microsimulation to perform in large *state space* – 'the representation of the components of the system of interest' (van Imhoff and Post 1998, p. 102) – where a large number of individual attributes would not be handled well by a macro model. Wilson and Rees (2005), in a review of population projection methods, conclude that microsimulation should be considered when projecting populations with multiple attributes.

In the demographic literature, microsimulation has been used for a range of applications. For example, it was used by Ruggles (1992) to correct missing data bias in historical demography, and by Thomson et al. (2012) to estimate completed fertility in France when union formation and dissolution were varied in the simulation. Models have been implemented in population projections: see, for example, SMILE (developed for Ireland) (Ballas et al. 2005) and MOSES (developed for the UK) (Wu and Birkin 2012). Microsimulation models are also used to test the implications of policy change. Ballas and Clarke (2001) model the local impact of changes to national social policy in the UK; Brown and Harding (2002) provide an overview of a selection of microsimulation models examining policy decisions in Australia.

Microsimulation is a technique which is useful outside the constraints of academia. Increasingly microsimulation is being adopted to undertake official projections by national statistics agencies. MOSART is used by Statistics Norway to project education, labour force and public pension benefits (Fredriksen and Stolen 2007); Statistics Canada's projection model Demosim uses microdata from the 2011 National Household Survey to produce estimates of future ethnocultural composition, Aboriginal populations and labour force (Bélanger and Sabourin 2017); PENSIM is used by the UK Department for Work and Pensions to estimate the future distribution of pensioner incomes (Emmerson, Reed and Shephard. 2004); and SESIM is used by the Swedish Ministry of Finance for similar purposes (Sundberg 2007).

Notwithstanding their utility, there are challenges and limitations associated with implementing microsimulation models which must be considered. First, a certain amount of technical expertise is required, although we explain in the next section that there are a range of options for implementing microsimulation models which are suitable for different levels of expertise. Second, microsimulation can be computationally intensive. This was a key problem for early implementations of micro models;

however, advances in computing means that this is less of a limitation now than it once was. Spielauer (2011, p. 18) outlines a third, more conceptual criticism: 'randomness caused by accumulated errors and biases'. Because microsimulation draws from a range of data sources, there is the potential for errors or biases in those data being extrapolated because of the level of detail required (or expected) from the microsimulation model. This randomness, similarly identified by van Imhoff and Post (1998) and Wilson and Rees (2005), is more acute in dynamic models because they are stochastic: results of a projection, for example, are randomly drawn from an expected distribution. Brown and Harding (2002, p. 1) highlight that microsimulation models are limited by design, assumptions, algorithms and data requirements but the solution is to 'make these explicit and then interpret the results within the models' limitations and capacities'. This is advice which should be applied to any model, be it micro or macro in design.

4. How can microsimulation be applied?

There are a multitude of ways to undertake microsimulation, from freely available custom software packages to writing your own model in a development language. Bespoke packages are probably the easiest place to start. One excellent example is a program called the Flexible Modelling Framework (FMF), created specifically to perform spatial microsimulation (Harland 2013). The FMF has a Graphical User Interface (GUI) which allows the user to select input files and specify required options and outputs, circumnavigating any requirement for the user to write their own code. The model is, however, only capable of static microsimulation.

In the programming language R, there are several packages which have been developed to undertake microsimulation. The MicSIM package is one example which provides the required functionality to undertake microsimulation. For a comprehensive overview of how to use R to undertake spatial microsimulation, see Lovelace and Dumont (2016). An interim step between writing a program from scratch and an off-the-shelf GUI is offered by Statistics Canada (Bélanger and Sabourin 2017). Their framework, called Modgen (model generator), implements an extension of the C++ language to help researchers build a model to their own specification, but does require some development skills to implement properly.

For the examples presented in this paper we have chosen to write the code in R. This is because it offers the flexibility to incorporate data and functions of our choosing. Also, because the source code is freely available, it offers researchers the option of adapting the code to their specific needs. A slightly longer initial outlay in terms of time and learning yields a more flexible and adaptable approach.

5. Two examples: static and dynamic microsimulation

We have produced an R package to demonstrate the techniques of both static and dynamic micro-simulation with some example visualisations of the results. This package and associated documentation can be accessed via a GitHub repository (see https://github.com/virgesmith/demographyMicrosim). In this section, we give a description of the data and methodology, and provide some example illustrations from the results. Detailed instructions on how to install and use the package are given in the README.md file at the GitHub link. The R package is self-contained in that it contains all the code and input data needed for the microsimulation.

In the first example we use static microsimulation to create a base human population for the East London Borough of Tower Hamlets from 2011 UK Census data, estimating the population by geographical location, gender, age and ethnicity. In the second example we take this synthetic population and undertake a projection with dynamic microsimulation using detailed ethnicity-specific fertility and mortality data. We demonstrate ways to visualise the estimates using a map and population pyramid.

5.1 Case study area

The East London Borough of Tower Hamlets makes an interesting case study area because it is one of the most diverse areas in the UK, both in terms of its ethnic composition and its socioeconomic structure. This part of London has seen successive waves of immigration and onward migration for centuries, including Huguenots, Jews, Irish, Bangladeshis and, more recently, migrants from Eastern Europe, Somalia, Eritrea and Yemen. The borough is also a place of economic extremes: areas of poverty such as Shadwell, Whitechapel and Poplar contrast with the financial district of Canary Wharf and the high-end 'Docklands' riverside housing developments which have replaced the old working docks.

Our estimates are produced at Middle-layer Super Output Area (MSOA) spatial resolution. MSOAs are a census geography where zones are aggregations of Output Areas, the smallest geography at which census data are released, and designed to provide consistent population counts across the country. MSOAs contain an average of just under 8,000 people. Tower Hamlets is split into 32 MSOAs and the borough's total population is recorded as just over 250,000. The method outlined here could be adapted and applied to any other geography in the UK or elsewhere, so long as the data are available. The Australian equivalent to these hierarchical census geographies are Statistical Areas (SAs). SA2s have an average population of about 10,000 so are the nearest equivalent to the UK's MSOA geography.

5.2 Input data

The input data consist of two datasets, which are used for the static and dynamic microsimulation. We also provide geographical boundaries for the visualisation of results. The specific data used in our example are described in more detail in the following sections. However, in general terms, to undertake a static microsimulation as described in this paper the user will need one or more datasets containing full enumeration of the population of interest (e.g. from a census), disaggregated by some attributes. Subsequent tables don't need to provide full enumeration and some attributes need to match if drawing from a number of tables. To undertake dynamic microsimulation the user will additionally need demographic rates for the population of interest (e.g. from vital registration data). These need not be at the same spatial scale, as shown in our example. To replicate the visualisation of results, our code can be adapted to deal with any spatial system if shapefiles are available.

5.2.1 Aggregate population data

We use data from the UK 2011 Census of population to create the microsimulated synthetic population. The 2011 Census provides counts of people by area (in this case MSOA), cross-tabulated by a number of characteristics. Census data used here are sourced from Nomis official labour market statistics (https://www.nomisweb.co.uk/) and contain public sector information licensed under the Open Government Licence v3.0.

We have used the following tables:

- DC2101EW: Ethnic group by sex by age. After processing we have named this file sexAgeEth.csv in the GitHub repository. It provides a count of persons by MSOA by sex by age band by ethnicity.
- DC1117EW: Sex by single year of age. After processing we have named this file sexAgeYear.csv. It provides a count of persons by MSOA by sex by single year of age.

The data tables contain the following fields, where 'Persons' is the count of each cross tabulation:

- MSOA: identification code for the 32 MSOAs within Tower Hamlets
- Sex: M (Male) and F (Female)
- Age Band: 0-4, 5-7, 8-9, 10-14, 15, 16-17, 18-19, 20-24, 25-29, 30-34, 35-39, 40-44, 45-49, 50-54, 55-59, 60-64, 65-69, 70-74, 75-79, 80-84, 85+
- Age: individual years up to 84, then a single 85+ category
- Ethnicity: BAN (Bangladeshi), BLA (Black African), BLC (Black Caribbean), CHI (Chinese), IND (Indian), MIX (Mixed), OAS (Other Asian), OBL (Other Black), OTH (Other), PAK (Pakistani), WBI (White British), WHO (White Other). The categories for Ethnicity have been aggregated from the original Census categories in order to be consistent with the fertility and mortality data discussed below.

Figure 1 provides an extract from the two constraint tables. We can see that both tables contain useful information but are not consistent with each other.

```
"MSOA";"Sex";"AgeBand";"Ethnicity";"Persons"
"E02000864";"M";"0-4";"WBI";95
"E02000864";"M";"0-4";"WHO";8
"E02000864";"M";"0-4";"MIX";30
"E02000864";"M";"0-4";"IND";2
"E02000864";"M";"0-4";"PAK";2
"E02000864";"M";"0-4";"BAN";129
"E02000864";"M";"0-4";"CHI";1
"E02000864";"M";"0-4";"OAS";2
"E02000864";"M";"0-4";"BLA";19
"E02000864";"M";"0-4";"BLC";16
"E02000864";"M";"0-4";"OBL";19
"E02000864";"M";"0-4";"OTH";5
"E02000864";"M";"5-7";"WBI";29
"E02000864";"M";"5-7";"WHO";7
```

a

```
"MSOA";"Sex";"Age";"Persons"
"E02000864";"M";0;72
"E02000864";"F";0;57
"E02000864";"M";1;70
"E02000864";"F";1;62
"E02000864";"M";2;61
"E02000864";"F";2;52
"E02000864";"M";3;67
"E02000864";"F";3;79
"E02000864";"M";4;58
"E02000864";"F";4;61
"E02000864";"M";5;59
"E02000864";"F";5;68
"E02000864";"M";6;51
"E02000864";"F";6;58
```

b

Figure 1: Extracts from (a) table sexAgeEth.csv and (b) table sexAgeYear.csv.

5.2.2 Fertility and mortality rate data

There is substantial variation in the rates of mortality and fertility for different ethnic groups (Rees et al. 2009; Norman et al. 2014), and it is important that our microsimulation captures this. We use ethnic-specific fertility and mortality rates produced as part of the NewETHPOP project. NewETHPOP has produced population projections for UK local authorities, disaggregated by age, sex and ethnic group, that capture the variation in demographic components which is mostly missed by other (non-ethnic disaggregated) projections. For more information on the project see Rees et al. (2017). Projection results can be downloaded from www.ethpop.org.

Rates by ethnicity and single year of age are available for the entire Borough of Tower Hamlets. These rates are not differentiated at any smaller geographical scale. We use this information in the microsimulated population projection. Note that while ethnic-specific migration data are used in the NewETHPOP project, we do not use this in our paper. Our aim is to keep the examples we present simple and reproducible.

We provide the following files, which provide the ethnic-specific information, in the GitHub repository:

- TowerHamletsFertility.csv, which is the ethnic and single year of age specific fertility rate for Tower Hamlets.
- TowerHamletsMortality.csv, which is the ethnic and single year of age specific mortality rate for Tower Hamlets.

5.2.3 Geographical boundaries

The code to map the results for the Tower Hamlets projection is also provided at GitHub. This includes MSOA boundaries made available by the Office for National Statistics under the Open Government Licence v3.0. Background map data are sourced from Carto through OpenStreetMap under the Open Data Commons Open Database License.

5.3 Methodology

In the following examples, microsimulation is necessary because we have information from different data sources which are not entirely consistent. While we have counts of people by single year of age by sex in one census table, the ethnic disaggregated counts from the other census table are only available for grouped age bands. The information on fertility and mortality rates is available for single year of age and sex, but is not geographically disaggregated below borough level. The aim here is to use static microsimulation to produce a population of individuals by sex, ethnicity, single year of age and MSOA. This base population is then aged on using the ethnic-specific fertility and mortality rates in the dynamic microsimulation.

5.3.1 Static microsimulation

Using microsimulation, we can generate a synthetic population that enumerates both ethnicity and single year of age for every individual. The sum of this population is entirely consistent with the input data: it will add up to both the ethnicity totals (by age band) and the population totals (by single year of age) for each geographical area and sex.

For the microsimulation we use the humanleague R package that generates a population using quasirandom sampling of the marginal data (Smith et al. 2017). Iterative proportional fitting (IPF) achieves a similar goal, minimising statistical significance at the expense of not always producing a whole-number population in each (age, ethnicity) state, and thus requires some form of adjustment. See Lomax and Norman (2016) for a step-by-step guide to IPF.

In our example, in any given area, for each gender and age band, we know the total persons of each ethnicity, and the total persons of each (single year of) age. We fill a table with people in such a way that: (1) there is a whole-number population for each age and ethnicity; (2) the totals for each age and ethnicity are correct; and (3) the population has a low statistical significance, i.e. age and ethnicity are not correlated.

The algorithm has three broad functional components:

1. Load the input data and compute various data that will be required later (such as the categories, and a match between age band and age).
2. Perform microsimulation for each geographical area and insert into the population.
3. Perform checks on the synthesised population to ensure it is consistent with the input data.

The result is a population of 254,096 individuals who have some combination of age, sex, ethnicity and MSOA location. This is the base population which will be used for the dynamic projection.

5.3.2 Dynamic microsimulation

In this example we project the base population from 2011 to 2021. The projection uses a Monte-Carlo simulation that assign births and deaths to the population based on the age- and ethnicity-specific fertility and mortality rates for the Borough of Tower Hamlets. This type of methodology is described in van Imhoff and Post (1998). In this example, for each iteration (i.e. year) we draw two uniform independent random variates for each eligible person. We compare these values to the fertility and mortality rates for the appropriate age and ethnicity: if the first random variate is lower than the fertility rate, a birth is assigned to this person; if the second is lower, a death is assigned. Only females have nonzero fertility rates, which are strongly age dependent.

If the fertility rate is 0.1, this will result in on average one in ten of the population drawing a random lower number and having a birth assigned to them. For a mortality rate of 0.01, only 1 per cent of the population will on average draw a number lower than this. The simulation is discrete in that it operates in one-year intervals, and the user specifies the number of years to run the projection.

In this relatively simple example, our population is 254,096 persons. Each person exists in one of 66,048 possible states (32 [MSOAs] × 86 [ages] × 2 [genders] × 12 [ethnicities]). In a *macro*simulation we would need to keep track of the number of people in each state (i.e. 66,048 values). However, in a *micro*simulation, we need to keep track of individuals and the states that they occupy. This requires storing about 1,000,000 values, since there are four separate categories.

As more categories are added the number of states grows *exponentially*, while the storage required for individuals only grows *linearly*. It would only take the addition of one or two more categories before the number of states would exceed the population, and microsimulation would be a more efficient approach in terms of storage.

Note that the following assumptions are made in the model, which are reflected in the source code in the 'microsimulate' function (see lines 71–82 in microsimulation.R):

- only single births occur (i.e. we assume that multiple births are factored into the fertility rate)
- newborns have an equally probable chance of being male or female
- the ethnicity and MSOA of the newborn is the same as their mother's
- births occur before deaths – thus a newborn will survive if a parent dies within the same year
- migration is not taken into account.

See the final section of this paper for discussion about developing the example further.

The algorithm can be described as follows:

1. Load the ethnicity-specific fertility and mortality rates.
2. Randomly assign births and deaths to members of the population in a manner that is consistent with the fertility and mortality rates.
3. Age the population by one year.
4. Add newborns (aged zero) and remove the deceased from the population.
5. Repeat from step 2 until the target year is reached.

5.4 Visualisation

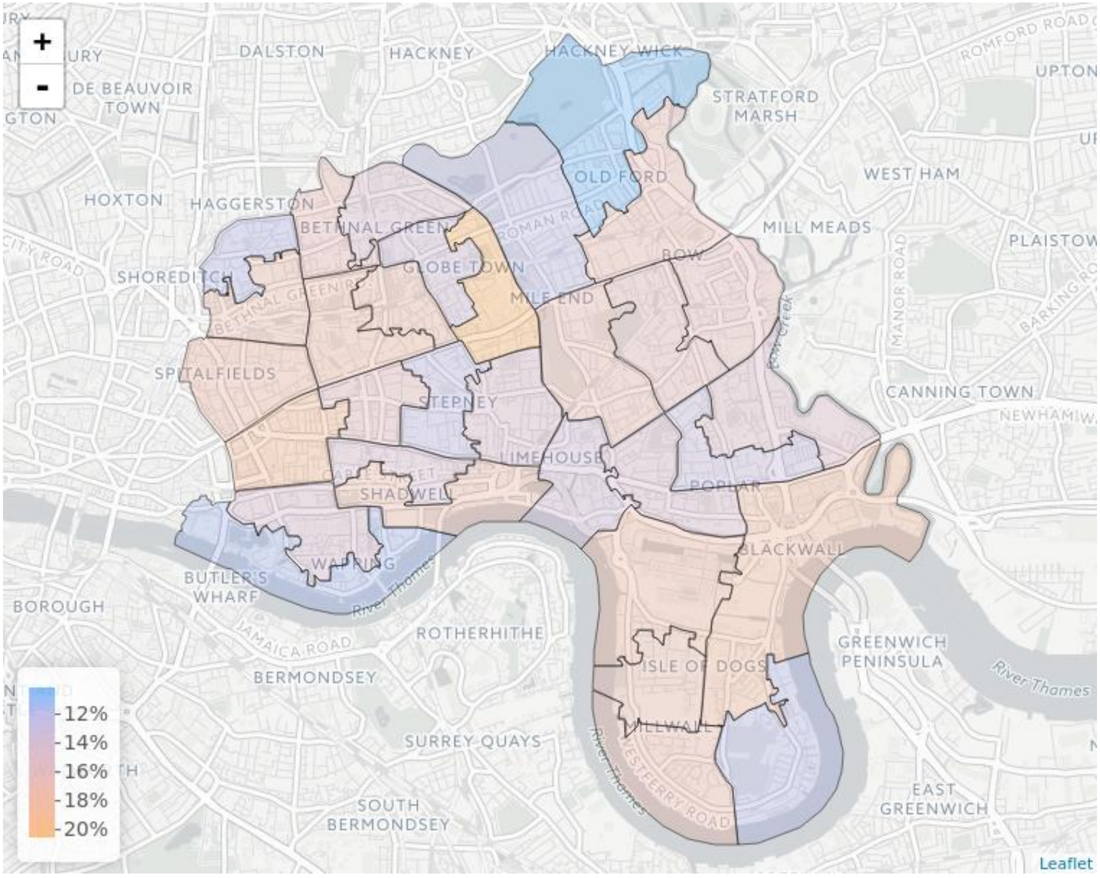

Figure 2: Map of projected percentage population growth, 2011–2021

Source: Map tiles by Carto under CC BY 3.0. Data by OpenStreetMap under ODbL. *Note*: Low growth is blue and higher growth orange.

To interpret the results, it is important to be able to visualise them effectively. The package provides functionality to calculate summary measures for visualisation. Firstly growth (g_T),

$$g_T = \frac{P_T}{P_0} - 1$$

where P_0 is the initial population and P_T the final one. The value will be negative if the population shrinks, zero if unchanged, and positive if the population increases. The value is not annualised. Secondly diversity,

$$d = 1 - n.\mathrm{var}(\mathbf{p})$$

where n is the number of categories (i.e. ethnicities), $\mathbf{p}$ a vector of populations in each category, and var is the sample variance. While numerous measures for diversity are used to explain population distributions (e.g. Rees and Butt 2004 use the Diversity Index), this measure has the useful properties that the value is: zero when the population is all in one category; one (1) when the populations in each category are equal; and essentially independent of the number of categories.

We include code to produce maps of these data in the GitHub repository. Figure 2 (previous page) provides an example of total population growth between 2011 and 2021.

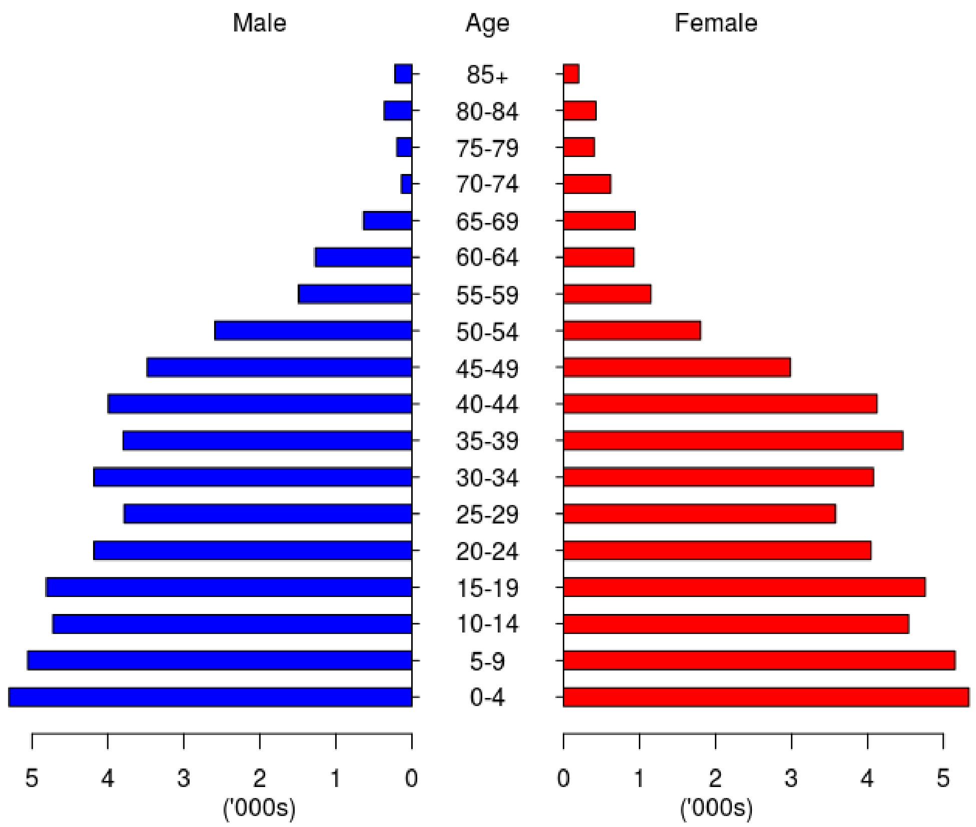

Figure 3: 2021 Projected Bangladeshi population pyramid, 2021
Source: Authors' projections.

We also provide a function to produce population pyramids of subsets of the base or projected population. Figure 3 shows one of these population pyramids for the age and sex structure of the Bangladeshi population in 2021. The code can be adapted to produce figures for other ethnic groups.

6. Adaption and extension of the example

The examples outlined in the previous section provide the building blocks needed for researchers to implement both static and dynamic microsimulation. In order to use the example in different contexts, researchers will want to adapt the data inputs to take in to account their own research interests. For example, fertility and mortality rates for different groups (e.g. stratified by socioeconomic status or health status) could replace ethnicity, and different geographic systems could replace the MSOAs used here.

There are also improvements which would make the microsimulation models more sophisticated. The first obvious extension is to include migration (both internal and international) in the model. This could be done by incorporating migration rates in a similar way to those used for fertility and mortality, but it is important to note that the model complexity and run time would be increased. For an overview of how complex and multifaceted a projection with dynamic microsimulation can be, see the work of Holm et al. (2008) who describe the SVERIGE model which not only incorporates migration, but also education, marriage, leaving home, divorce, employment and earnings.

An individual simulated population cannot be considered an accurate prediction of the future. It is essentially an extrapolation that is subject to biases and uncertainties in the input data, the model assumptions and the random noise that is inherent in any Monte-Carlo simulation, which may amplify the initial biases. The second improvement, which would help to negate this problem, would be to perform a number of iterations of the model run and from these compute not a single value but a confidence interval for aspects of the final population. This would go some way towards alleviating the criticisms about randomness discussed in Section 3. Multiple runs of the model would give the researcher an idea about the variability of results, providing some indication of uncertainty which could be attached to projection estimates. This significantly increases the amount of computation required, but there are various statistical techniques available (outside the scope of this paper) which are more efficient than a brute-force approach.

A third improvement is to test the sensitivity of the model to the input data or model assumptions. If there is a suspicion that some value or values in the input data are inaccurate or biased, perturbing these values (i.e. changing them by a small amount) and re-running the microsimulation can establish the sensitivity of the model to these inputs. If the projected population (or some aspect of it) is very similar the model can be considered insensitive to the input and the input's accuracy may not be a major issue. Conversely, if the projected population differs significantly the model is highly sensitive to the input, and perhaps the results cannot be accepted with much confidence. A final important consideration is the validation of model results. Limitations in the amount of space available negate a discussion of model validation in this paper but for a good summary see Ballas et al. (2005).

7. Conclusion

This paper has provided an overview of microsimulation as a technique which should be considered by demographers who are interested in estimating individuals within a population (through static models) and undertaking projections of those individuals (through dynamic models). We have argued that microsimulation is a technique which is useful for demographers and one which is becoming increasingly used as data, computing power and user support increases. Our example is applied to the East London Borough of Tower Hamlets but the approach could be readily applied in Australia, or any other country for that matter, so long as appropriate data are available. Useful next steps would be to look at the full R code provided for the examples outlined in this paper and have a go at implementing the model. The code can be adapted to suit the researchers' requirements.

References

Ballas D and Clarke G (2001) Modelling the local impacts of national social policies: a spatial microsimulation approach. *Environment and Planning C: Government and Policy* 19(4): 587–606.

Ballas D, Clarke G, Dorling D, Eyre H, Thomas B and Rossiter D (2005) SimBritain: a spatial microsimulation approach to population dynamics. *Population, Space and Place* 11(1): 13–34.

Bélanger A and Sabourin P (2017) *Microsimulation and Population Dynamics: An Introduction to Modgen 12*. The Springer Series on Demographic Methods and Population Analysis 43. Switzerland: Springer.

Birkin M and Clarke M (2011) Spatial microsimulation models: a review and a glimpse into the future. In: Stillwell J and Clarke M (eds) *Population Dynamics and Projection Methods*. Netherlands: Springer; 193–208.

Brown L and Harding A (2002) Social modelling and public policy: application of microsimulation modelling in Australia. *Journal of Artificial Societies and Social Simulation* 5(4), http://jasss.soc.surrey.ac.uk/5/4/6.html.

Clark S, Birkin M, Heppenstall A and Rees P (2017) Using 2011 Census data to estimate future elderly heath care demand. In: Stillwell J and Duke-Williams O (eds) *The Routledge Handbook of Census Resources, Methods and Applications: Unlocking the UK 2011 Census.* London: Routledge; 305–319.

Emmerson C, Reed H and Shephard A (2004) *An assessment of pensim2*. London: IFS Working Papers, Institute for Fiscal Studies (IFS) No. 04/21, doi: 10.1920/wp.ifs.2004.0421.

Fredriksen D and Stolen N (2007) Model 1: MOSART (Dynamic Cross-Sectional Microsimulation Model). In: Gupta A and Harding A (eds), *Modelling Our Future: Population Ageing, Health and Aged Care.* Bingley: Emerald Group Publishing Limited; 433–437.

Harland K (2013) *Microsimulation model user guide: flexible modelling framework*. UK: National Centre for Research Methods (NCRM), NCRM Working Paper No. 06/13.

Holm E, Holme K, Lindgren U and Mäkilä K. (2008) *The SVERIGE Spatial Microsimulation Model*. Sweden: Department of Social and Economic Geography, Umeå University, Discussion Paper No. 595.

Li J and O'Donoghue C (2013) A survey of dynamic microsimulation models: uses, model structure and methodology. *International Journal of Microsimulation* 6(2): 3–55.

Lomax N and Norman P (2016) Estimating population attribute values in a table: "get me started in" iterative proportional fitting. *The Professional Geographer* 68(3): 451–461.

Lovelace R, Ballas D and Watson M (2014) A spatial microsimulation approach for the analysis of commuter patterns: from individual to regional levels. *Journal of Transport Geography* 34: 282–296.

Lovelace R and Dumont M (2016) *Spatial Microsimulation with R*. Florida: Chapman and Hall, CRC Press.

Nelissen J (1991) Household and education projections by means of a microsimulation model. *Economic Modelling* 8(4): 480–511.

Norman P Rees P and Wohland P (2014) The use of a new indirect method to estimate ethnic-group fertility rates for subnational projections for England. *Population Studies* 68(1): 43–64.

Orcutt G (1957) A new type of socio-economic system. *Review of Economics and Statistics* 39(2): 116–123. Reproduced in *International Journal of Microsimulation* 2007 1(1): 3–9.

Orcutt G, Greenberger M, Korbel J and Rivlin A (1961) *Microanalysis of Socioeconomic Systems: A Simulation Study*. New York: Harper and Row.

Rees P and Butt F (2004) Ethnic change and diversity in England, 1981–2001. *Area 36*(2): 174–186.

Rees P, Wohland P and Norman P (2009) The estimation of mortality for ethnic groups at local scale within the United Kingdom. *Social Science and Medicine* 69(11): 1592–1607.

Rees P, Wohland P, Norman P and Clark S (2017) Population projections by ethnicity: challenges and solutions for the United Kingdom. In: Swanson D (ed.) *The Frontiers of Applied Demography*. Switzerland: Springer; 384–408.

Ruggles S (1992) Migration, marriage, and mortality: correcting sources of bias in English family reconstitutions. *Population studies* 46(3): 507–522.

Smith A, Lovelace R and Birkin M (2017) Population synthesis with quasirandom integer sampling. *Journal of Artificial Societies and Social Simulation* 20(4): 14, doi: 10.18564/jasss.3550.

Spielauer M (2011) What is social science microsimulation? *Social Science Computer Review* 29(1): 9–20.

Sundberg O (2007) Model 5: SESIM (Longitudinal Dynamic Microsimulation Model). In: Gupta A and Harding A (eds), *Modelling Our Future: Population Ageing, Health and Aged Care*. Bingley: Emerald Group Publishing Limited; 453–460.

Thomson E, Winkler-Dworak M, Spielauer M and Prskawetz A (2012) Union instability as an engine of fertility? A microsimulation model for France. *Demography* 49(1): 175–195.

van Imhoff E and Post W (1998) Microsimulation methods for population projection. *Population: An English Selection* 10(1): 97–138.

Wilson T and Rees P (2005) Recent developments in population projection methodology: a review. *Population, Space and Place* 11(5): 337–360.

Wu B and Birkin M (2012) Agent-based extensions to a spatial microsimulation model of demographic change. In: Heppenstall A, Crooks T, See L and Batty M (eds) *Agent-based Models of Geographical Systems*. Netherlands: Springer; 347–360.