Individual and Household Level Estimates Based on 2001 UK Human Population Census Data

Andy Turner

CSAP Seminar on Microsimulation: Problems and Solutions 2007-04-27

Outline

- MoSeS: A Brief Introduction to the Demographic Modelling work.
- Focus on Population Initialisation for 2001 based on 2001 UK Population Census data.
- Future Work

MoSeS: A Brief Introduction to the Demographic Modelling work.

- Population Initialisation for 2001 based on 2001 UK Population Census data.
- Dynamic Model for 2001 to 2031.
- Enrichment using health and social survey data.
- Making the results and methods available for researchers and policy makers.

Focus on Population Initialisation for 2001 based on 2001 UK Population Census data.

- The basic task is to select a well fitting set of records from the Individual and Household Sample of Anonymised Records (ISAR, HSAR) to constitute census areas.
- Mark Birkin has done some work to find a well fitting set using an Iterative Proportional Sampling (IPS) technique and I have done some work using Genetic Algorithms (GA).
- The fitness of a set is evaluated by a fitness function which compares aggregate estimates from the ISAR and HSAR, with those from published Census Aggregate/Area Statistics (CAS).

Current GA work

- Select from the HSAR the Household Population (HP) and the ISAR the Communal Establishment Population (CEP)
  - In theory this is OK,
    - Although the HSAR is only available for England and Wales.
- Control Constrain (CC) and Optimisation Constrain (OC)
  - CC are measures that *have to* be met in solutions.
  - OC are measures that are fitted to.

Control Constraints

- CC the CEP using CAS001 and CC the HP using CAS003.
- CC the HP using CAS003
  - Age of Household Reference Person (HRP)
  - CAS003 age is grouped as 0-19, 20-29, 30-39, 40-49, 50-59, 60-69, 70-79, 80+.
- CC on HP age is very hard!
  - Any Household can contain up to 12 individuals.
  - Consider the difficulty ensuring that the total number of individuals is correct.
    - Household records can be ordered by number of individuals.
  - To order by the exact age breakdown of the individuals is considerably more complex.

GA

1. An initial set of solutions is generated.
2. Solutions are bred:
   - This is done by mutation involving swapping a random number of HSAR and ISAR ensuring CC are met:
     - CC are met by ensuring any record swapped in has has age in the relevant age group of CC.
3. Each solution is measured for a goodness of fit.
4. The best fitting solutions are kept.
5. Steps 1 to 5 are repeated until convergence or until a fixed number of iterations are completed.

Some Results

- http://www.geog.leeds.ac.uk/people/a.turne

Future Work

- Next Steps
  - Analyse errors graphically and investigate outliers.
  - Produce and analyse Geographical Maps of the errors.
  - Run Belinda's dynamic model based on these data.
  - Develop a publication comparing the IPS and GA method and results.
    - Target GeoComp2007
  - Reproduce ISAR only results which can be used more readily by others.
  - Tidy and fully document the code.
    - The code is written in Java and is open source.
    - http://www.geog.leeds.ac.uk/people/a.turner/src/andyt/java/projects/MoSeS/
  - Enrich the ISAR and HSAR records by incorporating variables from the BHPS and other Survey and Health Databases.
    - Mark has some Masters students looking into this.

References

- MoSeS source code
  - http://www.geog.leeds.ac.uk/people/a.turner/sr
- MoSeS NCeSS Home Page
  - http://www.ncess.ac.uk/research/moses/
- Andy Turner's MoSeS Web Page
  - http://www.geog.leeds.ac.uk/people/a.turner/projects/

Acknowledgements

- Thanks to NCeSS colleagues and the ESRC for supporting this work.
- Thanks MoSeS project team for your collaboration.
  - In particular thanks to Mark Birkin, Martin Clarke and Phil Rees the demography experts.
- Thanks to the CCG, the School of Geography and the University of Leeds for your help and support over the years.