Spatial microsimulation techniques for constructing a spatially disaggregated population micro dataset

Kazumasa HANAOKA
Associate prof., Department of Geography Ritsumeikan University
Deputy director, Institute of Disaster Mitigation for Urban Cultural Heritages, Ritsumeikan University
Aim and Outputs

• The aim of this research is to create spatially disaggregated microdata (synthetic microdata) of population in a spatial microsimulation approach.

• Such synthetic microdata allows us to understand population characteristics at the small area level for the whole country (Japan).

• Our general purpose microdata can be used for spatial analysis on provision of public service, retailing, disaster risk management, etc.
Backgrounds

• In social science, disaggregated approaches using microdata (data of decision-making units) have gained increasing interests since 1950s.

• Microsimulation model was proposed by an economist (Guy Orcutt) in the late 1950s. It considers heterogeneity of individuals by using microdata.

• The model has been applied in wide disciplines but each discipline applied it in slightly different ways. The meaning of “microsimulation” are thus varied by discipline nowadays.
Microsimulation means:

- In economics and demographics
  - Population projection and simulation based on an individual dataset
  - Applications: Population projection, tax transfer, pension

- In transportation studies
  - Traffic simulation (behavior of individual vehicles)
  - Applications: optimization of traffic light, congestion

- In geography/geographic information science
  - Small-area population estimation method and policy applications
  - Application: health mapping, demand analysis in retailing
Microsimulation in Australia

- Example: Developments by the National Centre for Social and Economic Modelling (NATSEM)
- Non-spatial and spatial microsimulation: STINMOD+, APPSIM, SpatialMSM
Microsimulation in Geography

- Microsimulation was introduced in geography in the 1980s. Geographers tried to add geographical element in the simulation.

- However, spatially disaggregated microdata sets were not readily available (even at municipality level) for researchers in many countries.

- Thus, methods to create such microdata were studied in 1980-90s in the UK and they are often called as “spatial microsimulation”.
Spatial microsimulation

- A method to create spatially disaggregated microdata by combining various data sources such as census tables and survey samples

- Two major approaches
  - (1) Iterative proportional fitting: A method to estimate table cell counts based on marginal totals of benchmark tables
  - (2) Reweighting by combinatorial optimization algorithm: A method to estimate a new combination of survey samples which agrees to marginal totals of benchmark tables at the small area level

- Others: regression type etc.
Illustrative explanation of combinatorial optimization algorithm (Simulated annealing method)

Seed microdata (survey samples)

Feedback to resampling

Resampling

Census tables

Benchmark tables

Goodness-of-fit score

tabulation

Estimated microdata

Area A

Area B

Area C

Area D
Datasets

- Census samples for seed microdata
  - 1% anonymized samples of Japanese population census 2000 (approx. 1 million individuals)
  - Resampling by household unit (no change in household members)

- Census tables for benchmark (marginal totals)
  - Small-area statistics of population census 2010 (Approx. 200,000 neighborhood areas)

<table>
<thead>
<tr>
<th>Individual level</th>
<th>Household level</th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) sex * age</td>
<td>(7) building type</td>
</tr>
<tr>
<td>(2) sex * nationality</td>
<td>(8) housing tenure</td>
</tr>
<tr>
<td>(3) sex * marital status</td>
<td>(9) floor size</td>
</tr>
<tr>
<td>(4) sex * type of industry</td>
<td>(10) household type</td>
</tr>
<tr>
<td>(5) sex * occupation</td>
<td>(11) household size</td>
</tr>
<tr>
<td>(6) sex * work/school place</td>
<td></td>
</tr>
</tbody>
</table>

Total categories: 161
Outline of our estimation method

Example: Area i in Tokyo

Select census samples of Tokyo only

Small-area statistics

Benchmark tables for Area i

Area i

Area j

Area k

Stack all micro data sets

Synthetic microdata of the whole Japan

= 120 million individuals
Goodness-of-fit measures

• Absolute Error (AE)
  • \(= |\text{census count} - \text{estimated count}|\) for cell \(i\)

• Total Absolute Error (TAE)
  • \(= \text{sum of absolute error}\) for table \(j\)

• Squared Error (SE)
  • \(= (\text{census count} - \text{estimated count})^2\) for cell \(i\)

• Total Squared Error (TSE)
  • \(= \text{sum of squared error}\) for table \(j\)

For sample replacement only
Distribution of absolute error

- The 77% of all cells in estimated tables agreed with those of benchmark tables. The 96% are within ±1.

Cell counts by attribute

<table>
<thead>
<tr>
<th>Household size</th>
<th>1 person</th>
<th>2</th>
<th>3</th>
<th>⋮</th>
<th>7+</th>
</tr>
</thead>
<tbody>
<tr>
<td>Benchmark table</td>
<td>6</td>
<td>10</td>
<td>12</td>
<td></td>
<td>1</td>
</tr>
<tr>
<td>Estimated table</td>
<td>5</td>
<td>10</td>
<td>12</td>
<td></td>
<td>0</td>
</tr>
<tr>
<td>Absolute error</td>
<td>1</td>
<td>0</td>
<td>0</td>
<td></td>
<td>1</td>
</tr>
</tbody>
</table>

Absolute Error = |Oi - Ei|
Goodness-of-fit measure

- Average scores of TAE among all neighborhood areas

![Bar chart showing goodness-of-fit measures for individual and household levels.](chart.png)
Mapping results

- Average age of household head at the neighborhood level

Tabulations by any census variables are virtually possible
Application

• In March 2011, the Great East Japan Earthquake gave devastating impacts on millions of people by its huge tsunami and radioactive contamination.

• In recent, natural disasters such as earthquake, typhoon, heavy snow, tornado, landslide, and volcano eruption occur almost every year in Japan.
Disaster impact assessment

• Overlay analysis of population distribution and natural hazard distribution to present vulnerable areas.
Non-three generation households living in owned house
and the household head aged 65 years old and over