Estimating Spatial Distributions of Earnings at the Small Area Level: A Spatial Microsimulation Approach

K. Hanaoka[1]

[1] Department of Geography, Ritsumeikan University, 56-1 Tojiin-kita-machi, Kita-ku, Kyoto, 603-8577, JAPAN
Email:kht27176@fc.ritsumei.ac.jp

1. Introduction

Recently, spatial inequalities have attracted wide attentions in social sciences. However, the largest obstacle to conduct geographical analysis on earnings is a lack of earning information in census tables (Green 1998). The government-sponsored surveys are also the primary candidates for data sources in terms of accuracy and the coverage of populations. Nevertheless, they are not readily available especially at the small area level due to privacy issues, and customized questionnaire surveys are too expensive to carry out. In recent, spatial microsimulation approaches have been used for various socio-economic phenomena and validated as a useful approach to estimate the spatial distributions at the small area level (for example, Ballas and Clarke 2001), which cannot be understood with single dataset.

Therefore, this article attempts to estimate the spatial distributions of incomes, especially the earnings of employees, by combining census datasets and a non-census micro dataset in a spatial microsimulation approach. This study has two aims: they are first, to propose a method of creating a synthetic micro dataset from the Person Trip Survey (PTS) and to add individual earnings information from the Basic Survey on Wage Structure (BSWS); and second, to present a socio-economic application of a spatial microsimulation through analysing the spatial inequalities of earnings and the relationship to crimes. Spatial inequalities referring here mean uneven distributions of people and goods between small areas. Small area means lower geographical units than municipality. The city of Kyoto, Japan is selected as an area for study.

2. Methodology

A spatial microsimulation framework and the datasets for estimating spatial distributions of earnings are presented in Figure 1.

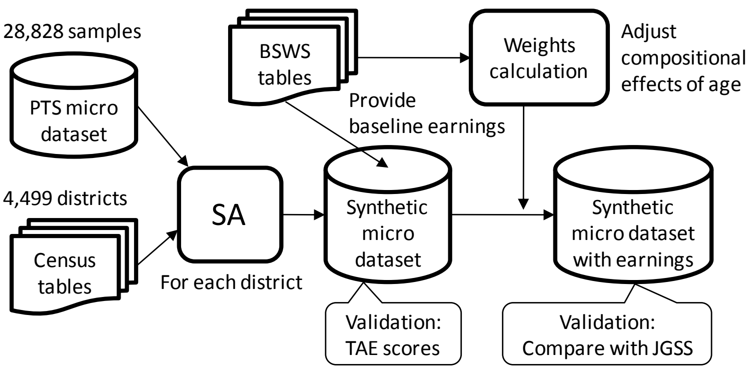


Figure 1. A spatial microsimulation framework for estimating spatial distributions of earnings.

2.1 Datasets

The BSWS are selected for estimating baseline earnings. We consider the annual earnings of standard employees. This includes scheduled cash earnings and annual special cash earnings (bonuses and term-end allowances).

The 2005 Population Census of Japan is used for the constraints for generating a synthetic micro dataset. The census tables are compiled by small area units called cho district (CD in this study). One CD accommodates approximately 595 people and 231 households on average for Kyoto. We selected five census tables at the CD level. They comprise, at the household level, the number of households by household size (7) and the number of household members aged 60 and over (3); at the individual level, the population by sex (2) and age group (8), the population by industry (10), and the population by occupation (4). The values in brackets denote the number of categories.

The PTS is used for seed samples for generating a synthetic micro dataset. The PTS contains information both on trips such as origins, destinations and times, and on demographic and socio-economic variables such as age, sex, household size, occupation and industry. The reasons why we select the PTS are that (1) the categories of these variables correspond to those of the census, (2) among public micro datasets, it is tabulated at detailed geographical unit called “PT zone”, contains approximately 70 CDs on average, and (3) its sample size is relatively large: it is 28,828 in the city of Kyoto.

2.2 Generating a synthetic micro dataset

Simulated Annealing (SA) algorithm, one of the combinatorial optimisation algorithms, attempts to find the best combination of the PTS samples by repeating a sample replacement until a synthetic micro dataset agrees with the five constraint tables explained in Section 2.1. The agreements between a synthetic micro dataset and the constraints are measured by Total Absolute Error (TAE) scores. The SA algorithm is applied five times for each of 4,499 CDs and a combination of samples showing the lowest TAE was chosen as the best one. Furthermore, we increase iteration times and check if the further reductions of TAE are possible for CDs that are twice as high as the mean TAE and TAE per household.

The results are summarised in Table 1. The mean TAE score is 20.22 and the CDs with a TAE lower than 10 and 20 comprise 36% and 72% of all the CDs respectively. These results indicate that only a few counts differ from the constraints on a cell-by-cell basis when the number of constraint tables is considered.

Table 1. Fit on constraints.

Statistics	
Average TAE	20.22
TAE per household	0.25
Number of CDs with TAE < 10	1,637(36%)
Number of CDs with TAE < 20	3,256(72%)

Values in brackes show cumulative percentages.
Areas with no population are excluded.

2.3 Adding earnings of individuals

The baseline earnings of individuals are derived from earnings by sex, age and industry from one of the BSWS tables. They are added to a synthetic micro dataset based on the attributes of the samples. Thereafter, using a similar method to standardised mortality ratio, the adjusted earnings weights for the class of position and production workers are calculated taking into account compositional effects of age. The baseline earnings are multiplied by these weights to compute the annual mean earnings.

3. Results

3.1 Validation

We conduct a cross comparison between estimated equivalent earnings and observed equivalent incomes of the Japanese General Social Survey (JGSS) 2000-2003 presented in Nakaya and Hanibuchi (2009). Equivalent earnings here are defined as the total earnings of all household members divided by the square root of the household size. The results are tabulated by the MOSAIC group (Figure 2). Although the estimated equivalent earnings by the MOSAIC group are 10-30% higher than the equivalent incomes of the JGSS, the values between them are agreed reasonably well (the Pearson correlation coefficient is 0.628, $p < 0.05$). The deviances are slightly larger in group B (Graduate newcomers), E (Middle Japan) and D (Older communities) but they seem to be within the accepted range in terms of the overall distribution.

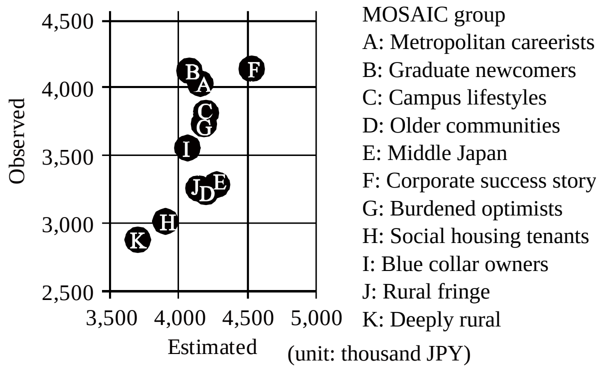


Figure 2. A comparison between estimated and observed earnings.

Furthermore, estimated distributions of average earnings by sex, age class and industry present realistic patterns (Figure 3): the large discrepancies in earnings levels between sex or age classes and the higher earnings in “Finance & insurance” and “Electricity, gas, heat supply & water” are found.

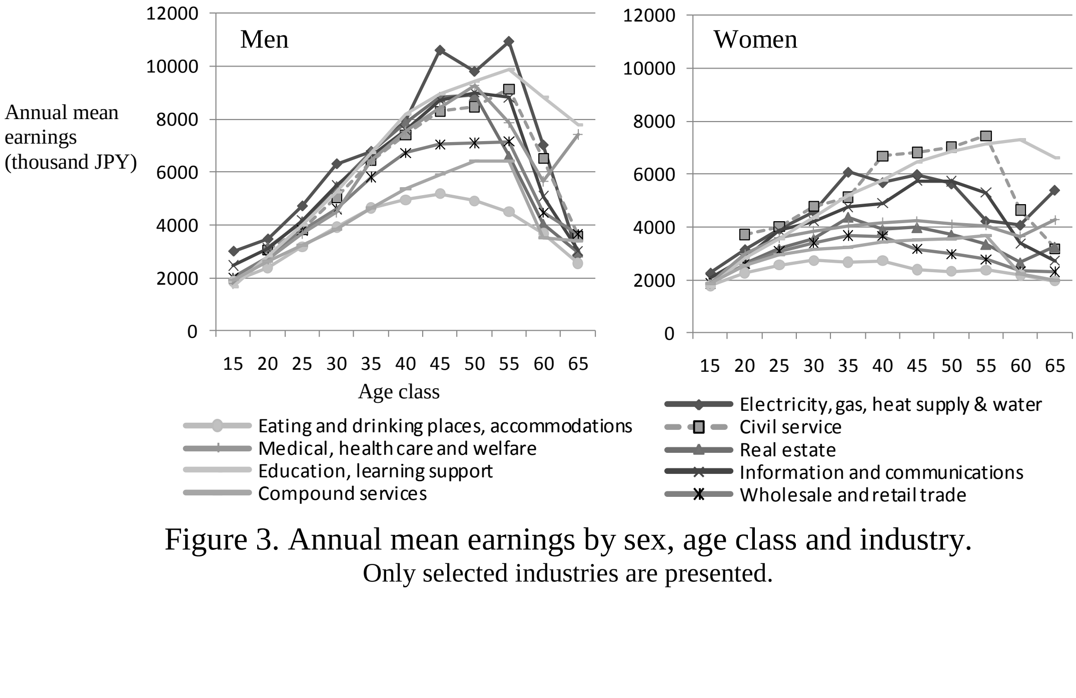
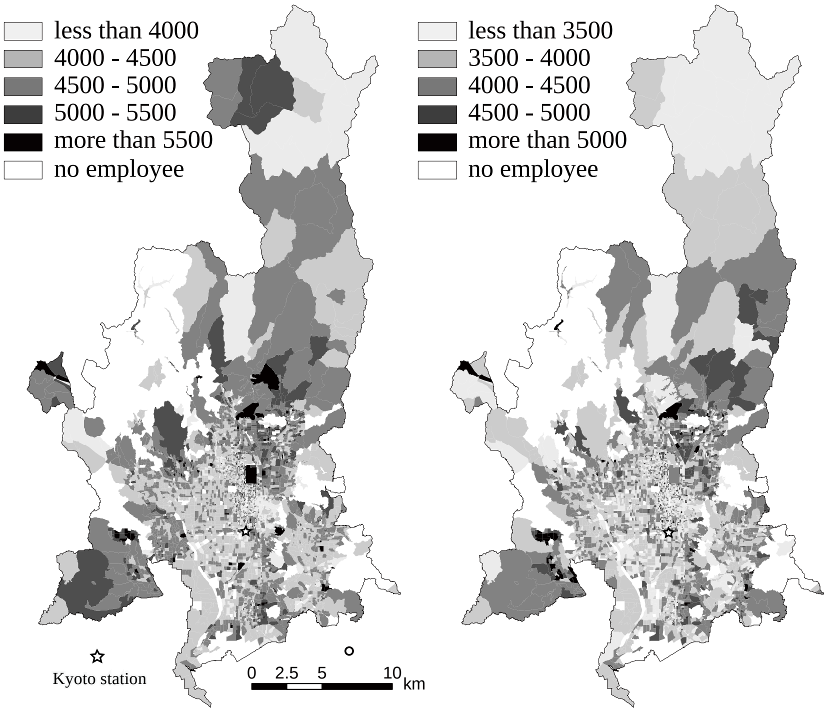


Figure 3. Annual mean earnings by sex, age class and industry.
Only selected industries are presented.

3.2 Mapping earnings measurements

The merit of using a spatial microsimulation approach is that the results are provided in the form of a micro dataset that is flexibly re-tabulated by any number of variables. One interesting topic not attempted previously is to compare spatial inequalities at the small area level based on different earning measurements: mean earnings, equivalent earnings and poverty ratio. Figure 4 displays spatial distributions of the mean earnings and equivalent earnings as an example. Overall both maps shows earnings are higher in north where the proportions of professional workers and managers are high but the map of equivalent earnings shows lower earnings in western suburbs where manly young families with children live.



Mean earnings Equivalent earnings

Figure 4. Spatial distributions of mean earnings and equivalent earnings (thousand JPY) in Kyoto City, Japan.

4. Using the results in a crime analysis

Shimada and Harada (1999) argue that a size of areal unit for crime analysis needs to represent a neighbourhood community. However, there are a few crime studies at such a small area level due to data availability. The results of a spatial microsimulation are tabulated by different areal units. In this case, we tabulate earning class, the average number of crimes per 1000 households for assault, burglary, motor vehicle theft and larceny by police jurisdiction district (123 districts in Kyoto City).

Figure 5 presents the results of crime analysis. The ratios equals to one minus poverty ratio are computed for a comparison. The figure shows how frequently crimes occur in a given jurisdiction classified by earning, in other words, vulnerabilities of neighbourhoods against crimes. Overall, relationships between earning class and the number of crimes appear to be negative. Therefore, poorer areas are more vulnerable against crimes although other factors regulating crime frequencies need to be considered. The results of earnings estimates provide indicators representing areal economic conditions. They can be used, for example, as independent variables for a multiple regression analysis on crime frequencies. In addition, re-aggregation of geographical units is possible to create suitable ones for analysis.

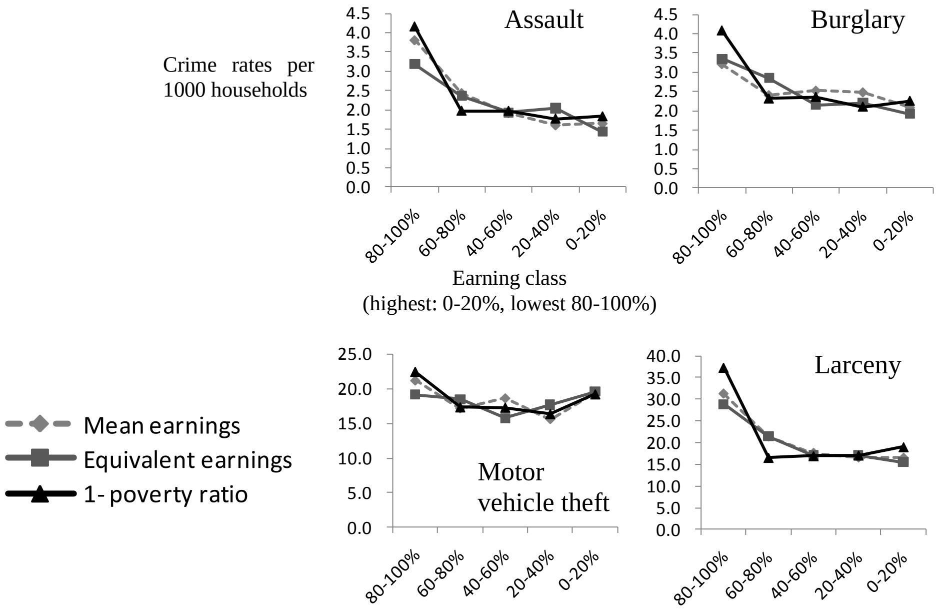


Figure 5. Relationships between earnings and crime rates.

5. Conclusions

In this study, we focus on generating a synthetic microdata with earnings estimates by using a spatial microsimulation approach. The results are summarised as follows:

(1) Using non-census micro dataset (the Person Trip Survey) and census tables at the CD level, we have created a synthetic micro dataset of individuals and households with relatively high accuracy.

(2) The results of estimated levels of earnings show realistic distributions by sex, age class and industry. Furthermore, the estimated earnings by the MOSAIC group are similar to the observed by the JGSS.

(3) A synthetic micro dataset is re-tabulated by household and the CD to calculate the mean and equivalent earnings as well as the poverty ratio.

(4) Our findings reveal that there are negative relationships between earnings and crimes such as assault, burglary and larceny. As a socio-economic application, we demonstrate that how we can use the outputs of spatial microsimulation to understand patterns of crimes and earnings at the small area level.

Acknowledgements

I thank the Kyoto Prefecture to kindly provide us the Person Trip Survey.

References

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

Green A E, 1998, The geography of earnings and incomes in the 1990s: An overview. *Environment and Planning C: Government and Policy*, 16 633-647

Nakaya T, Hanibuchi T, 2009, A note on data-linkage between a micro-dataset of a social survey and geodemographics: An application to examining geographical variations in self-rated health at a small areal level using the JGSS cumulative data 2000-2003. *JGSS Research Series* 6 23-36

Shimada T, Harada Y, 1999, The relationship between crime and socioeconomic factors in urban areas: Methodological problems and a solution by GIS. *Reports of the National Research Institute of Police Science: Research on Prevention of Crime and Delinquency*, 39 102-116