



Current Development in Microsimulation and Assessment of Uncertainty in JUTTA Model

Meng Zhou
Valmiera Workshop



Overview

- What and why
- Static approach and Dynamic approach
- Detailed dynamic microsimulation
- Application (JUTTA model)
- Conclusion



Introduction

- Why?

Evaluating the effects of the proposed interventions or policies before they are implemented in the real world.

- What?

Using database contains records of individual units rather than the aggregate information to go through a set rules until reaching the goals.



Static microsimulation

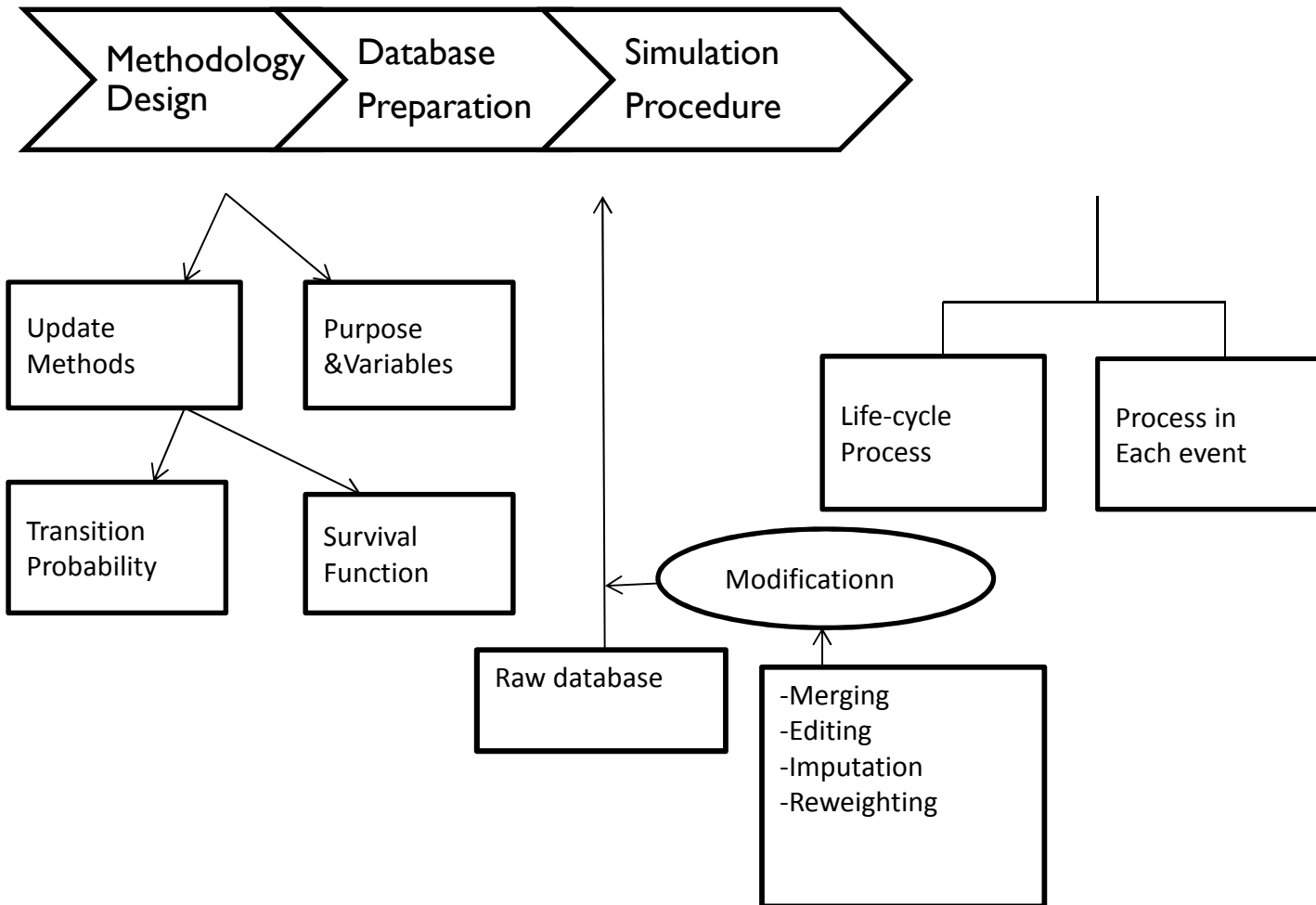
- It uses static aging techniques, for example changing certain variables on the original microdata file, to create the file with the demographic and economic characteristics expected in the future year.
- Notice: In static microsimulation, the change of people's characteristics do not depend on the probability , because it doesn't take the individual behavior into account . So once certain rules are made, they must be obeyed 100% without variation.



Dynamic microsimulation

- It ages each individual from one year to the next by deciding whether or not this person will go through the certain events by probability, such as giving birth, retiring, dropping out of school , getting a job, becoming unemployed. This probability is called transition probability, it plays the most important role in the updating procedure.
- The above procedure will be repeated as many times as the user wants to achive the final simulation year.
- Going through one event may effect the person's characteristics in the subsequent years. (e.g. dropping out of school may effect the salary the person get in future).
- It may create the synthetic database since by adding extra new borns' information.

Overall looking at the dynamic microsimulation





Update methods

Main stream: transition probability.

Time space: one year.

Another method: survival functions.(Cox model, Fertility, DYNAMOD2, Australia)

Transition probability approach example: Employment status(Logistic regression)

Let $Y_i=1$ denote unemployment of person i and $Y_i=0$ denote employment, π_i denote the probability that the individual is unemployed during the year.

For binary variable employment status, we have Bernoulli distribution:

$$Y_i \sim \text{bernoulli}(\pi_i)$$

$$\Pr[Y_i = 1] = \pi_i$$

$$\Pr[Y_i = 0] = 1 - \pi_i$$

Example continues

This event could be typically accomplished by a logit regression model. The logit model is given as:

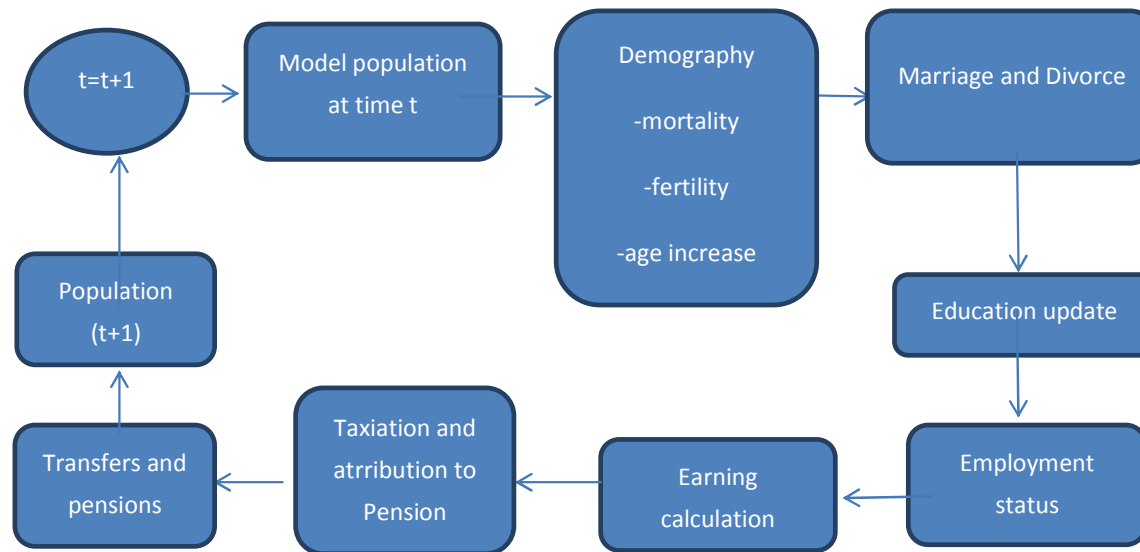
$$\pi_i = [1 + \exp(-X_i\beta)]^{-1}$$

Where X_i is a vector of individual or household characteristics like gender, age, working history. After calculating the estimated parameter $\hat{\beta}$, we could get the estimator $\hat{\pi}_i$.

Next step is using Monte Carlo method, comparing $\hat{\pi}_i$ and the random number u_i , $u_i \sim U(0, 1)$, give $Y_i=1$ if $\hat{\pi}_i > u_i$, otherwise, $Y_i=0$.

Simulation procedure

The simulation goes through the events every year and the results from year n will be used as the input information to estimate the results in year $n+1$. This procedure is repeated till the target year arrived.





Application

Background of JUTTA model:

It is the Finnish static model developed by KELA with the purpose of simulating the tax and benefits. It has ten sub-models and one whole model. The whole model is the summary of the sub-models. The sample size is 10989 households and around 30000 individuals, and they are from Statistics Finland.

Some examples of the Submodels:

- SAIRVAK: sickness insurance benefits
- TTURVA: unemployment benefits
- TOIMTUKI: means-tested income support.
- OPINTUKI: study grants

Assessment of JUTTA model

- Accuracy measurement:

Two forms:

absolute difference percentage:

The difference percentage is depending on the absolute error between the real value and estimated value simulated by JUTTA model. The absolute errors have been classified into 5 intervals $[0, 1)$, $[1, 10)$, $[10, 100)$, $[100, 1000)$, $[1000, \infty)$.

Relative difference percentage:

The difference percentage is depending on both absolute error and real data, because it is calculated as: $\text{abs}(\text{real value} - \text{simulated value}) / \text{real value}$. Then the relative difference percentage has been classified into five intervals which are $[0\%, 0.1\%)$, $[0.1\%, 1\%)$, $[1\%, 10\%)$, $[10\%, \infty)$.

Steps:

1. Classification into different intervals.
2. Calculating the percentage in each interval by n_i/N .



Assessment of JUTTA model

Results:

All the models perform well, only except that TOIMTUKI (support benefit), with large percentages in the largest intervals $[1000, \infty)$ and $[10\%, \infty)$. So next, we will improve this model by using statistical methods.



Statistical methods applied to the TOIMTUKI

Two models are set:

1. Linear regression
2. Two stage least squares(2SLS)(Logistic regression and linear regression).



Linear regression model

Algorithm:

- Giving the estimated value zero in case the real data is zero.
- Fit the regression with the rest of the individual, and estimate the values.
- Calculate the absolute and relative accuracy.

Result:

See Table one. Both absolute difference percentage and relative difference percentage have been improved to some extent.

2SLS method

Algorithm:

1. Estimate the binary variable-status, giving value 0 and 1, meaning not getting benefit and getting benefit respectively. This step is done by logistic regression using:

$$\pi_i = \exp(X\beta) / (1 + \exp(X\beta))$$

2. Then generating the random number from uniform distribution, compare the random number with the probability. If the random value is larger than the probability, giving status value 0, if not, giving value 1.

3. Fit the regression model when in case status=1, otherwise, give the estimated value 0. However, the estimated value by regression could be negative, where in reality, it should be nonnegative value, so we change the negative value to 0.



2SLS method

Result:

Please see table three, from this table we see that there is some improvement in the absolute difference percentage, but not in the relative difference percentage.



Conclusion

JUTTA model as a static microsimulation model performs quite well in all sub-models, only except for the support benefit model. The Linear Regression and 2SLS models both make some progress in its accuracy. In future, we hope we could find more statistically significant variables that will improve its accuracy.



Thank you for your attention!