Technical issues in implementing Micro Simulation

Jean-Christophe Dumont, March 2000

Different types of micro simulation models

The top down model

In this perspective micro simulation is nothing more than expending the number of household and having several representative households within each socio-economic group identified. The aggregated factors' income are distributed to the agents through a standard mapping process.

Concretely:
1) use the survey data to determine the factor endowment of each household
2) if possible, specify the transfer matrix between households
3) calibrate the demand function for 'each individual' just as in a standard model

The bottom up model

In this case the microeconomic specifications constitute the foundations of the model. Households' behaviour are estimated with the survey data and are supposed to be ground on some micro economic theoretical foundations. Aggregation is only done in fine.

More importantly the estimation work also identifies 'fixed effects' for households and individuals in the sample which describe the influence of non-observed characteristics, and possibly transitory factors on occupational choices and income. Thus the unexplained heterogeneity of the population is explicitly taken into account. These fixed effects also permits identifying 'potential' earnings of individuals in activities different from those observed in the survey and therefore makes possible to simulate the distribution effects of occupational changes caused by some kind of macro-economic shocks.

For instance, in Cogneau (1999) the labour supply responds to the expected wage with respect to the reservation wage. The latter is not directly observe but can be compute by a standard Tobit model (or a two steps Heckman procedure).

If consumption expenditures are available, one may also think of estimating demand functions at the individual household level and again identifying heterogeneity terms. (but obviously that does not means that demand function characteristics will be different across all households)

A dynamic micro simulation

Just to mention them dynamic micro simulation can help to represent change in demographic structure. Traditional method use some kind of transition matrix but here as

we can use stochastic techniques to produce new set of households and of individuals. The results then are no more determinists.

Choosing the good level of desegregation

It seems that the best level of desegregation of the household would be the best. However several points should be taken into account when choosing this patern:

1) the software capacities (cf. Indian survey)
2) the quality of the initial data
3) the potential problems of matching

Matching household data with macro data (Robilliard-Robinson 1999)

Reconciling household survey data and national accounts data is a well known problem. Computing macro aggregates from the household survey data by multiplying household production income, consumption and/or savings by the household sample weights and summing virtually never matches published national accounts data.

Many reasons are offered to explain this mismatch.
- On the household survey side, there may be sampling errors due to inadequate survey design and/or measurement errors because it is difficult to get accurate responses from the households concerning economic variables.
- On the national accounts side, while supply-side information on output and income for some sectors is based on high quality survey or census data for agriculture and industry, information for subsistence farmers and informal sectors producers is harder to obtain and usually of lower quality.

The three following solution can be considered:
- use the survey data to have shares and then apply RAS techniques (or similar methods). However taking into account the large number of observation (difficult to check) and the fact that some initial value may very low, such a procedure is certainly not efficient.
- produce the dual programing problem, which is staightforward when the constraint are all linear, and solve it using a standard NLP algorithm. Because there are far fewer shadow prices in the dual than endogenous variables in the primal that could appears as a solution (but not for CGE).
- reweight the household

In the latter case the estimation problem can be restated as follows:
Estimate a set of sampling probablities (household survey weights) that are close to a known prior and that satisfy various known moment constraints

Consider a sample of K households with prior survey probabilities pk which results in a vector xk of observed characteristics for each household such as household size, total household income, income by source, consumption and so forth. In addition from other sources, we have information about aggregations or weighted averages of some of the

household information. The estimation procedure is to minimize the cross-entropy measure of the distance between the new estimated probabilities and the prior.

$$Min \sum p_k \ln\left(\frac{p_k}{\overline{p}_k}\right)$$

Subject to moment consistency constraints and the adding normalisation constraint

$$\sum p_k f(x_k) = y$$
$$\sum p_k = 1$$

where *y* is an observed set of data that is required to be consistent with the distribution of probabilities or sample frequencies. The funcion *f(.)* represents a general aggregator of within household variables.

Some advice:
- check the gini coefficient afterwards
- adjust first your individual data by a lump sum coefficient to the aggregate income.

An example applied to Madagascar