Application of microsimulation models in the Hungarian Central Statistical Office

Zsolt Kovari[1]

In this article I would like to present a static microsimulation model developed in the Hungarian Central Statistical Office (HCSO), which was used primarily for household's incomes study. This model was intended for a first step of a greater research, thus we worked with the simplest tools. The items of household's incomes were aged, for this the data were gained from the Household Budget Survey (HBS). We were looking for an answer by the method of the microsimulation to the question how exact and reliable result can be produced from a base year for the following year by a static model based on the data of the HBS comparing to the real data. We would like to demonstrate further by some impact analysis whether the models is suitable for analytical presentation of the consequences of economic and political decisions.

The result proved that the method operated well concerning the most important income items. The gross income and its parts, the incomes from work and social incomes were successfully estimated with sufficient accuracy. At those cases where there are considerably differences between the estimated and measured values, however those are interesting from theoretical point of view, in reality they are insignificant small values affecting only few people. The impact analysis is considered also successfully, since its results were as expected, well interpretable and they were in conformity to the measured real values.

Keywords: modelling, microsimulation, income, tax simulation, family allowance

1. Introduction

Modelling is a frequently used and accepted method for the goal of which is to get to know real-life processes more precisely. A model is a simplified picture created by humans. In its creation we do not take into consideration the process of less importance with regard to the examined phenomenon. Theoretical models often appear in a mathematical form for the building of which we use abstract concepts. Simulation is practically a procedure for analysis of mathematical models whose purpose is to make you understand the behaviour of system modelled or to make different experiments with it. In the examination of the economic systems' simulation models are of utmost importance because they are too complex to examine them in another

[1] Zsolt Kovari, senior councillor, Hungarian Central Statistical Office (Budapest)
Author expresses his thanks to Timea Toth for their useful advice.

way and we cannot make experiments with them. Simulation is nearly always carried out using a computer and we draw conclusions on the behaviour of the systems on the basis of the examination.

Microsimulation models are based on so-called micro-units. These are the basic building blocks of the examined system. In case of economic situation these blocks are individuals, households, enterprises and other units. In the process of modelling the characteristics and behaviour are examined on the basis of the statistical data collected. The procedure was developed by G. Orcutt who published his famous article entitled "A new type of socio-economic system" in 1957. Thanks to the extensive research since then, nowadays microsimulation models are made for numerous purposes. These are classified according to many aspects. There are two types: (Molnár 2003)

- Data based models
- Agent based models

Data based models store the micro-units in a database. The database is made by data collection and contains detailed information on micro-units and their economic background. The other component of the model is the rule system which determines changes caused by the examined processes. (Bourguignon–Spadaro 2006) These changes can be events of some probability. Some characteristics of the items, however, can be modified by changing the way of calculation. In the agent-based models the micro-units are usually objects, where they have interrelation with each other and with their environment. Their behaviours are influenced by behavioural rules which are described by algorithms. This article concentrates on data based models.

The data based models are divided into two groups depending on whether we take into consideration the changes in the composition of micro-units in time or not. Thus:

- Static models
- Dynamic models

Changes in time are not taken into consideration in static models, the demographic composition of the micro-units can be left as invariant. Thus these models are used for studying short-term effects of economic or political arrangements. In a household a child can be born, a person may die, or the couple may get a divorced. Moreover new households may be established or ceased. An enterprise may go bankrupt, transform or open new local units. The dynamic models carry out these events with a defined probability and then modify the simulated database properly. Therefore these models are able to analyse the effects of the examined economic decisions in the long run.

Nowadays some microsimulation models appear, which have been developed for special studies, such as spatial models used for regional effects on eco-social political decisions. Microsimulation models were also developed for modelling the effects of crises. In 2008 Robilliard at al. used microsimulation to examine the effects

of poverty and inequality of the financial crisis which happened in Indonesia in 1997. (Robilliard et al. 2008) The effects of the sub-prime crisis which broke out in 2006 in the US and which resulted in the world-wide bank crisis in 2008 will probably be manifested in the data of the Household Budget Survey in 2008 and 2009. This article cannot analyse these questions because at the time of writing this material data on crisis were not readily available. It is an exciting challenge to collect and implement the methodological instruments for the analysis of the effects of the financial crisis.

It can be seen clearly that microsimulation is used very effectively in the process of preparation for decision makings. In the meantime it can be a tool of eco-political experiments, since reliable impact analyses are made with its assistance. The database models show which social groups may be winners or losers of a certain measure with different parameters. You can study the results of the tax rate modifications or imposing of tax on various income components.

2. The former microsimulation work in HCSO

In 1983 at HCSO the Household Statistical Microsimulator (Zafir 1987, 1988) development was started in cooperation with the University of Darmstadt. It was used for income distributions of tax variants. The application was ready for the practical use. (Szívós 1993) In the mid 90's HCSO and Cambridge University developed a simplified microsimulation model, which examined the effects of family supports on the households' incomes. This was useful for the 1996 income survey, but the expedient microsimulation work has not been continued. (Csicsmann-Papp 1998, Éltető-Havasi 1998) In 2006 HCSO in cooperation with Ecostat developed a static model, which was used for the study of tax and imposition of tax on family allowance. The purpose of this article is the presentation of this model.

3. Modelling of the change of the household incomes

The model developed in 2006 was founded on two targets. The one is to be producing some simple IT instruments which can be a starting point for new developments and put the productivity of a rule-based model to the test. The other one to have create a new developer team, which collect the relating special literature, gain developing experience and thus become capable to create more sophisticated models. We set up a static model in 2006, in which the analysis of incomes was emphasized.

3.1. The microsimulation database

The database of the model contained the data of 2004 Household Budget Survey. It was chosen because we would have liked to make the ageing for the next year, and

the 2005 data was available too (but not yet the 2006 data), therefore the calculated and real values could have been correspondent promptly.

The HCSO has been making Household Budget Survey for 50 years. The data were collected up to 1983 yearly, between 1983 and 1993 by every odd year, and after 1993 yearly again. Hungarian citizens living in Hungarian private household were covered by the population. A sample was taken by multiple-stage procedure from the population. The unit of sampling is the dwelling and the observation unit was the household. The sampling framework was the 2001 census. The data collection is continuous all year. Each month the one twelfth of the households in the sample kept a diary, in which making a note

- the expenditures of household's members
- the provisions bought, the quantity of clothing
- the benefits and donations received in kind
- the quantities consumed from own produced foodstuff or given presents

At the process of the questionnaire interrogation the census-takers record the data about

- the most important characteristic of the household's members (sex, age, educational level, economical activity, employment)
- the housing and equipments of the house
- the degree of supply of the consumer durables goods

At the end the households ensure data about

- personal incomes or other revenues
- non-personal (household) incomes
- expenditures on buildings, buying real estate, agricultural activity or significant consumer durable articles.

Consequently the Household Budget Survey contains detailed information about incomes and consuming expenditures of the members of the households in the sample. The data collected are made suitable by proper weighting procedure for representation the total population. After the data quality adjustment and imputation the database is stored in two SAS datasets, the personal and the household file. The two tables can be link by a unique household ID. This is a 15-digit text identifier. The 2004 database contains 23,513 individuals and 8,710 households. 62 variables based on individuals data and 92 variables based on household data were ageing or calculated. These variables can be classified to the next groups:

- Incomes from work (i.e. incomes from all-day employment, different refund of expenses, incomes from secondary employment or casual work, etc.)
- Social incomes (i.e. pensions, unemployment benefits or family allowances, etc.)
- Other incomes (i.e. income of interest payment, dividends, incomes from abroad, etc.)

The first step was the ageing of the incomes of individuals, and then they were summarized for household level. In the third part the household level incomes

(which could not be assigned to persons) were ageing. Household level incomes are for example the amount from insurance companies as life-, pension- or accident- insurance, received life annuity, amount of loan.

3.2. Ageing rules

The model rules were sorted to the following groups:

- Calculation according to regulations (I.e. pensions, different allowances, tax types, tax privileges, social security fees, net incomes are produced on the base of legislation).
- Calculation on the base of a statistical indicator (There were some income items, which were calculated in virtue of inflation or harmonized consumer price index).
- Calculation on the base of some ratio (We figured average growth rate on some data of previous year, to generate the counted values and we hoped that these would be not change significantly in the near future).
- Aggregation (The gross income, the tax base, the household total income, or the total income components were calculated on the summary of the estimated data. In some cases the law had to be taken into account, like the tax regulation asses the content of the base tax.)
- Calculation in virtue of statistical methodological rules (The results for the analysis, like the margins of the income categories, means or dispersion indices has been calculated on the base of usual methodological provisions.)

There are some obvious calculations, which were not sorted to the above mentioned groups, like the production of per capita net income.

When the model was built up we knew the legislation environment and the values of the used statistical pointers were also known. We can't rely on this in the general case, but it is not a contradiction. Naturally a model has some assumptions relating to the certain attributes of the environment for instance the legislation environment. One of the advantages of microsimulation is that we can choose these attributes and rules freely. A hypothetical law can be built in the model because it just means that certain legislation environment will be described. The necessary statistical pointers can be estimated or are assumed some values. Actually we examine the relation between the households and the values. So we always presume the legislation environment and the needful statistical data are known.

4. Computerized implementation

We had to create a proper computer program to carry out the ageing and producing the suitable outputs. Since the starting database stored in SAS datasets, it was a natural choice to elaborate a SAS program to perform the calculations. Those experiences were taken into account, which accumulated on the building of previous sys-

tems, and we used the guides which were found in the special literature or known from general programming methodological notes.

The key parameters were stored in separate parameter tables in order to give or change them easily. We used totally 6 parameter tables. These kept the data for personal income calculations, for defining the family tax privileges or family allowances, moreover further index numbers used in ageing.

The program was broken down to so-called micro-modules. The micro-module – similarly to the modules well-known in program development – is responsible for a certain function. This solution ensured that the modification of the model could be executed with a change or rewriting some micro-modules or parameters. Numerous different models can be formed from suitable system of micro-modules and parameter tables so the construction is extremely flexible.

The final version of our program contained 32 micro-modules, which can be sorted to the following groups:

- Preparatory modules (7 ones has been made. These are responsible for making the copy of the original database, or loading the parameter tables. Accordingly their activity is limited to technical-like tasks.)
- Ageing performing modules (22 ones has been made. These calculate the new values of the income items and produce the aggregated data. Their number is less, then the variables, since the same micro-module operates the variables ageing by the same way.)
- Producing output modules (3 ones has been made. These summarize the individuals and household incomes with and without weights, and create tables, which present different income groups of the households.)

Later other modules have been made too, which produced different publication tables, but these are not built organically into the program. For comparison to the 2005 real data there has to make utility programs, which produce the result from the Household Budget Survey 2005 after the ageing calculation.

5. Results of ageing

The most important result was the comparison between the aged and real income data values of the households. The table below shows the data of the most important income groups:

*Table 1. T*he most important income groups

	Grouping of the income (2004, aged)			
	N	**Total (million Ft)**	**Average**	**Standard deviation**
			(thousand Ft)	
Gross income	**3,775,955**	**10,067,190**	**2,666**	**2,205**
Income from work	3,149,455	7,301,745	2,318	2,392
Social incomes	3,157,745	2,613,833	828	594
Other incomes	3,031,969	151,613	50	326
	Receipts not accounted as income			
Cash from savings	1,181,917	268,780	227	494
Received loans, credits	192,164	114,712	597	1,301
Net income	**3,775,955**	**7,913,930**	**2,096**	**1,370**
	Grouping of the income (2005, real)			
	N	**Total (million Ft)**	**Average**	**Standard deviation**
			(thousand Ft)	
Gross income	3,837,087	9,984,986	2,602	2,045
Income from work	3,187,109	7,083,581	2,223	2,231
Social incomes	3,248,725	2,719,856	837	595
Other incomes	3,007,325	181,549	60	421
	Receipts not accounted as income			
Cash from savings	873,304	166,691	191	365
Received loans, credits	60,709	38,046	627	1,787
Net income	**3,837,087**	**7,986,438**	**2,081**	**1,352**

Source: Own creation

Notes: N means the number of households which have the income from the given type.

It can be seen in the table above that the gross and net incomes were estimated well by the model. The two largest components, the income from work and social incomes, have a good correspondence to their real values. Only in the receipts not accounted as an income has some differences to the fact values. The model significantly overestimated the number of raiser credits and loans, and the money amount from this category. But these are only just a small part of the gross income and touch very few people.

The reason of the differences can be found as the assumption rules are partly realized. The cash from savings was not modified in the model because it was guessed they would not be changed significantly. The government party preparing to the 2006 election did not wanted to deal economic problems, thus the general economic situation has been unchanged. This caused that fewer people must have fewer

cash from previous savings. The estimated value seems to be better, if we take this as cash from savings per capita. While the total amount in 2005 was 62% of the predicted data, the cash from savings per capita estimated 227,000 Fts, however, the real value was 191.000 Fts, as 84% of the estimated in the model.

We estimated the loans and savings on the base of the amount of loans by monetary financial institutions. The largest problem was that the numbers of credit receivers were modified drastically. It can be seen better again if we compare the credit per capita to the aged data. The total received loans in 2005 were 33% of the estimated, the credit per capita was 627,000 Fts, and the aged data was 597,000 Fts. This means only 5% difference. Entirely the receipts not accounted as an income item in 2005 was 53% of the model estimation, but this is not significant concerning the gross income, thus the discrepancy is not considered serious deficiency of the program. The table below shows that the gross income how distributes between the largest income groups.

Table 2. Distribution of the gross income

Gross income share in percentage	
2004 aged	
Gross income	100,0
Income from work	72,5
Social incomes	26,0
Other incomes	1,5
Net income	78,6
2005 real	
Gross income	100,0
Income from work	70,9
Social incomes	27,2
Other incomes	1,8
Net income	80,0

Source: A háztartások jövedelemalakulásának elemzése mikroszimulációs módszerrel (2007)

It's fully visible that the category of other incomes is under 2% both. This insignificant item was underestimated by the model, which is understandable, since the reasons recognizable the changes of laws and changes in questionnaire techniques.

Summarizing the consequences it can be say that the model's assumptions were operated well, since the larger importance income items were estimated by fewer than 5% error. Similarly we received favourable result to the major indicators of the income distribution and the income deciles.

6. Impact analysis

The impact analysis was mentioned among the utilization possibilities of microsimulation. For this demonstration we carried out two impact analysis. The one is related to taxation, in which we calculated the private income tax on the basis of accepted income tax rate of 2005 from the ageing data. This can be kept as an ageing procedure, since the planned tax legislation modifications are know at the previous year autumn, thus it was easy to built into a model. On the other hand it illustrate well, how simple to specify the tax tables, and to study the effects of the planned modifications. The ageing produced immediately the gross and net income, from which it was easily to calculate net income per capita. In addition we examined the shaping of the income deciles, the net income distribution of the different family compositions concerning the sex, or number of children, and we carried out regional examinations too. The incomes were grouping according to the economic activity of the head of household, and the number of active and inactive earners in the household.

The most important part of the modification of the 2005 year tax was that the 3-rated tax became to 2-rated by deleting the middle rate.

Table 3

2004			**2005**	
0–800 000	0	18%	0	18%
800 001–1 500 000	144 000	26%		
1 500 001	326 000	38%	270 000	38%

Source: Hungarian Tax and Financial Control Administration

The tables shows that the new Personal Income Tax has no effect on the incomes under 800,000 Fts, thus there was no foreseen changed for lowest income groups. The income groups between 800,000 and 1,500,000 Fts the system leave more incomes and – as reading from the table – the earners more than 1,500,000 Fts would be richer by 56,000 Fts.

In reality other impacts exists, but summarizing the new tax system lifted the net income per capita in every income deciles. However, the results show that the new system is favourable to the rich. Other investigation results are in harmony with the idea that the households with smaller income did not feel the effect of tax changes. For instance an examination was carried out for the regional impact analysis on the changes of net incomes of the households. Relevant difference was not demonstrable, but the smallest growth was estimated in Northern-Hungary and Northern-Great-Plain (1.3%), which are sorted to the poorest regions of the country. The biggest growth (1.9%) can be declared to the richest part, Central-Hungary.

Similarly explanation is when the households with active earner head of household could be rely on 2.3 % growth, meanwhile the pensioner households

could get 1.3% net income growth. As more active earners were presenting in a household, as larger net income growth happened, this also lead back to the features of the tax system detailed already. Where there was no active earner, the ageing net income surpassed only by 0.1% of the 2005 data, namely significant change was not predicted. Where 3 or more active earners lived in the house, we waited 2.6% growth in net income. More detailed examination would need to reveal the link of the net income growth and the number of children in the household. In our estimation the biggest decrease can be experienced by the couple with one child or with more children in 2005 (2% and 1.9%). Here the interpretation can not refer only to the tax system, since these families can receive family allowances or tax privileges. Simultaneously the result illustrated that the new system does not penalize the family with children.

The other illustrative impact analysis was the family allowance taking into tax base. The original idea was to increase the family allowance by 30%, and calculated into the tax base with taxation on the base of 2005 year rules.

In order to do this we had to link the family allowance to a member of the household. The member must not be child-aged or pensioner and we chose the one who possessed the lowest income. Household Budget Survey considered family allowance household level income, and we also did this during the simulation process.

The result was examined by the similar theory as the taxation. The net incomes were experienced to have increased to a small degree. The biggest growth (2,3%) happened in the first income decile. It is understandable, as the incomes are the lowest there, thus the taxed family allowance could not lifted them. This is considered to be as a positive impact. The other results have similar interpretation. We predicted 2,6% net income growth to the households which had not earner members. Insignificant effect presented by the types of settlement groupings. Here 0,4% net income growth could have been predicted to the household in villages.

Greater differences would have been revealed if the family allowances could have been lifted by 30%. We calculated this amount as the incomes in forms of taxes back to the budget has not been changed. The gross-up of net wages by more than 30% would have needed budget approval, which was not realistic in that economic and political situation.

The government in 2009 submitted a proposal, which involved the family allowance to be taken to the tax base. however, this proposal shared the family allowance among the members of the household in other way, than our way. According to the planned orders the half of the family allowance has to be calculated to the tax base of the parents who bring up together the children. But this proposal was annulled by the Constitutional Court of Hungary, because it trespasses on the theory of the proportional sharing on taxation. Namely in case when the state does not make payable the family allowance to the parent, this means that there is a taxation of not acquired income.

In our version the only one member would have to be under taxation and – assumption a rational activity of human beings – we supposed that this person would be who has the smallest tax base. The theory is rather unfeasible, as according to the Constitutional Court the family allowance to be under taxation can infringe the theory of taxation by ability. Theoretically a case can occur, when a parent brings up more children, she/he has to pay more tax. Because of these decisions – in spite of this can be a manageable by microsimulation – the result executed by the examination carried out can not be usable. But this does not diminish the productivity of the model and the elaboration of the model was not redundant, since we worked several algorithm out which can be useful for similar tasks.

7. Consequences

The microsimulation model developed in 2006 aimed to realize a tool with simple methods in order to carry out income ageing and impact analysis. It was a trial to compare the real and calculated data, and to find that the model rules and the starting database capable for microsimulation examinations. In this article I summarized some results, which proved that the model in its framework operated well. It was important to write it down, because it focuses the methodology of microsimulation and the fact that microsimulation can be happened by simple tools, raising interesting questions. For instance those cases, when the rule based model has also differences between the real and predicted data. Their explanation may lead to the same simple reasons, however, it can reveal a new idea to the model, and it encourages to a better understand of a part.

We planned the model showed to age in more dimensions. It would be important from economical or statistical point of view to build in demographical modules, or model of the activity rate or build in consumer data. It would make it capable to build sufficient and sensitive tax models. By informatics point of view it would be great to solve the easily handling of the parameters, or the automation of data supply.

A new tax simulator is described in XXXVII issue of recurrent publication which has been developed by private persons recently (Belyó et al. 2009). The designers produced a complex model oriented microsimulation system used open source components that can be used for simulation of personal income tax, corporation tax and value added tax. The authors made simulated personal income tax data from tax return data 2007. Corporation data was estimated from fact data 2008. The ageing data of VAT was calculated by consumption patterns of households which were constructed based on Household Budget Survey 2007. The authors could use information on 25,573 firms, 8,547 households and 22,305 individual.

The calculation of personnel income tax and corporation tax carried out by the similar ageing rules as of the model of HCSO[2]. The developers created a modern IT background and designed an effective framework system to help the modelling. The application basis on Web service technologies is a serious step towards a general microsimulation system which does not need collaboration of IT expert or computer man for the building of the model. The main importance of this part of the work, however, that it points to the necessity of a framework system by which the preparatory work can be carried out for the simulation and can be set of the model parameters.

References

Molnár, I. 2003: *A mikroszimuláció alkalmazása.* Ecostat, Budapest.

Bourguignon, F. – Spadaro, A. 2006: *Microsimulation as a Tool for Evaluating Redistribution Policies.* ECINEQ Working Paper 2006-20

Robilliard, A. – Bourguignon, F. – Robinsor, S. 2008: *Crisis and Income Distribution: a micro-macro model for Indonesia* The impact of macroeconomic policies on poverty and income distribution: macro-micro evaluation techniques and tools (Editors: François Bourguignon, Maurizio Bussolo Luiz A. Pereira Da Silva) Palgrave-Macmillan Publishers 93-118p.

Zafir, M. 1988: A háztartási mikroszimuláció felhasználása a statisztikai munkában. *Statisztikai Szemle*, 66, 4, 367-372. p.

Zafír, M. 1987: A *háztartási mikroszimuláció.* KSH, Budapest.

Szívós, P. 1993: A mikroszimulációs rendszer és alkalmazása. *Statisztikai Szemle*, 71, 4, 316-325. p.

Csicsman, J. – Papp, P. 1998: A családtámogatási rendszerek hatásvizsgálata mikroszimulációval. *Statisztikai Szemle*, 76, 3, 238-249. p.

Éltető, Ö. – Havasi, É. 1998: Mikroszimulációs kísérlet a családtámogatások hatásvizsgálatára. *Statisztikai Szemle*, 76, 4, 324-340. p.

Cserháti, I. – Dobszayné – Havasi, É. – Keresztély, T. – Kővári, Zs. – Szép, K. – Takács,. T – Tallér, A. – Tamási, B. – Varga, Zs. 2007: *Háztartások jövedelemalakulásának elemzése mikroszimulációs modellel.* Ecostat, Budapest.

Bátori, D. – Belyó, P. – Molnár, I. – Sinka, I. – Szathmári, B. – Tóth, G. Cs. 2009: *Adó-szimulátor.* Ecostat, Budapest.

[2] The model of HCSO only focused on the personal income tax.