Klevmarken, N. Anders

Working Paper

Micro Simulation - A Tool for Economic Analysis

Working Paper, No. 2001:13

Provided in Cooperation with:
Department of Economics, Uppsala University

Suggested Citation: Klevmarken, N. Anders (2001) : Micro Simulation - A Tool for Economic Analysis, Working Paper, No. 2001:13, Uppsala University, Department of Economics, Uppsala

This Version is available at:
https://hdl.handle.net/10419/82890

Micro simulation – a tool for economic analysis

N. Anders Klevmarken*

Abstract

Micro simulation involves modeling the behavior of individuals and other decision units taking into account the effects of policy parameters such as tax rates, eligibility rules for benefits and subsidies and compensation rates in the social security system. The model is simulated to analyze the impact of policy changes not only on mean behavior but also on the entire distribution of target variables. Micro simulation models have thus, for instance, been used to analyze how changes in the income taxes influence the tails of the income distribution (the incidence of poverty).

Micro simulation complements a more traditional economic analysis both of which have pros and cons. Micro simulation is demanding in terms of modeling effort, data requirements and computer capacity. The issues of statistical inference related to micro simulation are in principle no different from those in econometric modeling generally. In practice the large scale and complex structure of a typical micro simulation model and the shortage of good micro data raise inference issues of particular relevance for micro simulation such as the choice of estimation criteria, calibration to benchmarks and model validation. Some of these issues are discussed in this paper.

Keywords: Micro simulation, ceteris paribus assumption, large scale modeling, calibration, constrained estimation, validation, method of simulated moments.

JEL Classification: B4, C5

Lecture presented at the International School on Mathematical and Statistical Applications in Economics, January 15–19, 2001, Västerås, Sweden

* Department of Economics, Uppsala University, P.O.Box 513, SE-75120 Uppsala, Sweden. Email anders.klevmarken@nek.uu.se

1. The *ceteris paribus* assumption in economic analysis

In the empirical verification of economic models economists almost always have to rely on observational studies while controlled experiments are difficult to implement, sometimes even considered unethical and thus very rare. This has the important implication that confounding factors have to become controlled by careful modeling and the application of sound econometric methods.

In their theoretical work economists usually abstract from these confounding factors and concentrate on the mechanism of key interest – the so called *ceteris paribus* assumption. To convey a new idea and to demonstrate the key implication of this idea this is an efficient and useful approach. But in empirical testing and in using the new theory for policy recommendations this substitute in mind for a controlled experiment is in general not applicable. Only under rather special circumstances there is an econometric correspondence to the *ceteris paribus* assumption, i.e. no confounding factors other than random noise.

Even a first year student of Econometrics knows that in regression analysis omitted variables will in general result in biased and inconsistent estimates of the partial effects of the included variables. In fact this result generalizes to almost all misspecified models. Policy recommendations based on models using an erroneous *ceteris paribus* assumption will thus in general simply be wrong.

When is it then possible to model potentially confounding factors as residual white noise and when is it necessary to explicitly include them as part of the economic model? What is important enough not to be seen as a random residual? There is no straight answer to these questions. This is the art of model building and of using "rhetoric" to convince fellow economists and policy makers of the benefits of a particular model[1]. The only advice the econometrician can give is to put the model to the test of data using our

[1] See McCloskey(1983) and Sims(1996).

battery of diagnostic tests. There is, however, no foolproof procedure to follow to the "true" model. For a given finite sample many models will pass the diagnostic tests and with very large samples almost every model would become rejected. The existence of a true data generating process is a useful assumption made by the econometrician to be able to discuss properties of econometric methods, but it is not really needed in economic analysis. What is important is to find a model that is useful for a particular purpose. In the search for such a model one might be willing to neglect less important deviations between model and data and accept certain properties of the model even if diagnostic tests would reject them. The economist must, however, be able to show that these properties are of no consequence or at least "less important" for his purpose. Some analysts take this kind of argument as an excuse to fit very simple structures to data and to neglect the diagnostic testing. This is not well advised. Diagnostic tests that signal likely specification errors are always warnings to be taken seriously and we usually prefer an economic model, which fits data better than a model that does not fit as well.

A modern economy is a complex interaction of many economic agents and one would think that a realistic model in general would have to be almost as complex as the reality it is thought to mirror. Our experiences of large-scale modeling are, however, not entirely good. The large macro models of the 1960s and 1970s did not keep what they promised. The computable general equilibrium models have their advocates, but this line of research have also been heavily criticized, see for instance Hansen & Heckman (1996) and Sims (1996). It is hard to convince fellow economists and policy makers about the merits of a large-scale model, the functioning of which they have difficulties in understanding. They don't trust a black box!

Macroeconomics is about understanding the relations between the aggregates of the national accounts and macro economic modeling involves attempts to formalize these relations. In doing so microeconomic arguments are frequently used but dressed in the language of the average economic man. The link between micro and macro is, however, weak. Usually we cannot derive the macro relations from micro entities. In the 1950s

and 1960s this was a concern for economists. There was a literature on the "aggregation problem", see for instance, Theil(1954), Fischer(1969) and Lutjohann(1974), which now appears almost forgotten. Under the influence of time-series modeling, macro modeling has become a more or less independent branch of economics. In their critique of research using computable general equilibrium models Hansen & Heckman (1996) p. 100 note:

> It is simply not true that there is a large shelf of micro estimates already constructed for different economic environments that can be plugged without modification into a new macro model. In many cases estimators that are valid in one economic environment are not well suited for another. Given the less-than idyllic state of affairs, it seems foolish to look to micro data as the primary source for many macro parameters required to do simulation analysis.

In researching a problem area it is frequently fruitful to use different approaches, but still, will we ever be able to understand the movements in the macro aggregates unless we are able to derive them from micro entities? We do recognize that there is heterogeneity in the behavior of micro units and we know from the early aggregation literature that simply adding micro relations can derive no stable macro relations. Aggregation is thus no simple adding up but involves the interaction of micro units on markets and under institutional constraints. To set up a research program which permits not only random but also nonrandom heterogeneity in the behavior of economic agents, that allow them to interact in various ways and that also allow the explicit introduction of institutional constraints and policy parameters, a new framework is needed. It is possible that the micro simulation technique could provide it.

2. What is micro simulation?

Micro simulation is a technique that uses the capacity of modern computers to make micro units act and interact in such a way that it is possible to aggregate to the level of interest. A micro simulation model can be seen as a set of rules, which operates on a sample of micro units such as individuals, households and firms. Each micro unit is defined and characterized by a set of properties (variables) and as the model is

simulated these properties are updated for each and every micro unit. The model might simply be a set of deterministic rules such as the income tax rules of a country operating on a sample of tax payers, and used to compute the distribution of after tax income, the aggregate income tax revenue or other fiscal entities of interest. But the model could also include behavioral assumptions usually formulated as stochastic models. Examples are fertility models, models for household formation and dissolution, labor supply and mobility.

In micro simulation modeling there is no need to make assumptions about the average economic man. Although unpractical, we can in principle model every man. It is no simple task to model the behavior of single consumers and firms, but it is an advantage to model the decisions of those who actually make them and not the make believe decisions of some aggregate. It stimulates the researcher to pay attention to the institutional circumstances that constrain the behavior of consumers and firms. It also in a straightforward way suggests what data should be collected and from whom. Similarly, in a micro simulation model it is possible to include the true policy parameters and the rules which govern their use, such as the tax rate scales, eligibility rules, tax thresholds, etc. One is not confined to using average tax rates applied to everyone. This makes micro simulation especially useful for policy analysis.

The development of micro simulation can be traced to two different sources. One is Guy Orcutt's idea about mimicking natural experiments also in economics and his development of the behavioral dynamic micro simulation model DYNASIM[2] that later was further developed by Steven Caldwell into the CORSIM model[3]. Another source is the increased interest among policy makers for distributional studies. Changes in the tax systems of many Western economies has developed a need for a tool to analyze who will win and who will loose from changes in the tax and benefit systems. As a result many governments now have so called tax-benefit models. In Sweden Statistics Sweden and

[2] See Orcutt(1957), Orcut et.al.(1961, 1976)
[3] See Caldwell(1993)

the Ministry of Finance have the FASIT model[4] and at the European level EUROMOD is an ambitious attempt to build a tax-benefit model for all of EU[5]. These models usually do not include behavioral relations but only all the details of the tax and benefit rules. These rules are then applied to a sample of individuals for which one knows all gross incomes and everything else needed to compute taxes and benefits. For every individual in the sample one is thus able to compute the sum of all (income) taxes due and the disposable income for each household. The output becomes, for instance, the distribution of the disposable income. It is then possible to change the tax rates or anything else in the tax code and run the model once again and compare to the previous result. In this way one can analyze who will gain and who will loose from a tax change. The simulation model will, however, only give the first-order effect of a tax change, because household composition, work hours and incomes are assumed unchanged and not influenced by the taxes. This is both a strength and a weakness of the tax-benefit models. It is a strength because it is easy to understand what the model does and no controversial assumptions are needed. There is also no difficult inference problem. All the analyst needs to do is to draw an inference from the random sample of taxpayers to the population of taxpayers, which is something we know from sampling theory. The weakness is of course that we do not know the relative size of any adjustments of behavior to the tax and benefit changes. The first-order effect might be a bad approximation. As a result attempts have been made to enlarge the tax-benefit models with behavioral models to capture these adjustments. In this way the tax-benefit models approach Orcutt's DYNASIM and it successors. In Sweden there are a few dynamic micro simulation models. The oldest model first developed in the early 1970s is the model used by the National Insurance Board (RFV) to simulate the future of the Swedish public pension system (Eriksen 1973, Klevmarken 1973). Another model developed in the 1980s is MICROHUS, see Appendix, and a third is an adaptation to Sweden of the U.S. model CORSIM, a model called SVERIGE (System for Visualising Economic and Regional Influences in Governing the Environment) that is not yet

[4] http://www.scb.se/befovalfard/inkomster/iof/ioffasit.asp
[5] http://www.econ.cam.ac.uk/dae/mu/emod3.htm

completed[6]. At the Ministry of Finance there is also ongoing work on a dynamic micro simulation model SESIM[7]. It has, for instance, been used to evaluate the recently decided changes in benefits for students[8].

To give the reader an idea of the structure of a large-scale microsimulation model a survey of the MICROHUS model can be found in an Appendix.

3. Pros and cons of microsimulation

Most of these models are designed to focus on distributional issues in particular on the income distribution. This is a characteristic feature of micro simulation models. They are useful tools for analyses of distributions not only mean relations. This is something we should bear in mind when we discuss statistical inference in micro simulation models.

Tax rules and rules that determine who is eligible for various benefits are usually highly nonlinear and sometimes have discontinuous jumps. Micro simulation models have the advantage of relatively easily accommodating such functional forms. One is thus not confined to functions with nice properties.

There are also disadvantages and problems with micro simulation models. One is that the size and complexity of a typical model makes it hard to understand its properties intuitively. This is one reason why micro simulation has not become fully accepted by the Economics profession. Given the main tradition of working with small, stylized models and the relative failure of the large macro models of the 1960s and 1970s many economists are now skeptical about the usefulness of large models. In order to change this, micro simulation modeling has to rely on good economic theory and use sound econometric inference methods, but economist also have to learn what scientists in other disciplines already know, namely how to examine the properties of large simulation models.

---

[6] http://www.smc.kiruna.se/
[7] http://www.sesim.org/
[8] For a review of dynamic microsimulation models also in other countries see Klevmarken(1997).

Contributing to the skeptics of the Economics profession is also the view that the science of Economics has not yet given us knowledge such that it is meaningful to build large micro simulation models for policy analysis and policy advice. For instance, in their assessment of the needs for data, research and models the Panel of Retirement Income Modeling of the U.S. National Research Council concluded (Citro & Hanushek, 1997, p. 163):

"To respond to immediate policy needs, agencies should use limited, special-purpose models with the best available data and research findings to answer specific policy questions. Although such models may not provide very accurate estimates, the alternative of developing complex new individual-level microsimulation or employer models in advance of needed improvements in data and research knowledge has little prospect of producing better results and will likely represent, in the immediate future, a misuse of scarce resources."

This was a recommendation to government agencies as policy makers concerned with retirement behavior. It should not be interpreted as general recommendation against microsimulation. On the contrary they also suggested (p. 153):

"The relevant federal agencies should consider the development of a new integrated individual-level microsimulation model for retirement-income-related policy analysis as an important long-term goal, but construction of such a model would be premature until advances are made in data, research knowledge, and computational methods."

As pointed out by the panel one of the major problems in microsimulation work is the shortage of good micro data. Although the supply of micro data has increased very much in the last 20-30 years it is still hardly possible to find one data source or one sample which will contribute all the information needed for a typical microsimulation model. In fact many model builders have found it necessary to use guestimates of model parameters and then try to calibrate the model against known benchmarks. Calibration is nothing but an attempt to tune the unknown parameters such that the model is able to simulate reasonably well the distributions of key variables. In this respect there is a

similarity between microsimulation modeling and general equilibrium modeling. Both rely too often on the calibration technique. Hansen & Heckman (1996) criticized this approach because they found too little emphasis on assessing the quality of the resulting estimates. In fact the properties of the estimates are usually unknown and *mutatis mutandis* the same is true for the simulated entities. The calibration techniques also tend to hide a more serious problem, namely that typically calibration involves only one year's data or a single average or total. Because this reliance of a single or just a few points of benchmark data they do not always identify a unique set of values for the model parameters.

Even if the parameter estimates are not calibrated guestimates but true estimates, the absence of comprehensive data has typically induced people to a piece meal estimation procedure. Each submodel is estimated from its own data set and there is no model-wide estimation criterion. If the model has an hierarchical or a recursive structure and if the stochastic structure imposes independence or lack of correlation between model blocks or sub-models, then a piece meal approach can be justified, but in general it cannot.

By way of an example consider the following simple two-equation model:

$$\begin{aligned} y_{1t} &= \beta_1 x_t + \varepsilon_{1t}; \\ & \qquad E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma_1^2 & \text{if } i=j=1. \\ \sigma_2^2 & \text{if } i=j=2. \\ 0 & \text{if } i \neq j. \end{cases} \\ y_{2t} &= \beta_2 y_{1t} + \varepsilon_{2t}; \end{aligned} \tag{1}$$

This is a recursive model and it is well-known that OLS applied to each equation separately will give consistent estimates of β_1 and β_2. The estimate of β_1 gives the BLUP $\hat{y}_t = \hat{\beta}_1 x_t$ while predictions of y_2 outside the sample range are $\hat{\beta}_2 \hat{y}_1$. However, this suggests the following model-wide criterion,

$$\frac{1}{s_1^2}\sum_t (y_{1t} - \hat{y}_{1t})^2 + \frac{1}{s_2^2}\sum_t (y_{2t} - \hat{b}_2 \hat{y}_{1t})^2; \tag{2}$$

Minimizing this criterion with respect to $\hat{\beta}_1$ and $\hat{\beta}_2$ yields the OLS estimator for β_1 but the following estimator for β_2,

$$\hat{b}_2 = \frac{\sum_t y_{2t} x_t}{\sum_t y_{1t} x_t}; \qquad (3)$$

In this case both the "piece meal" OLS estimator of β_2 and the "system-wide" instrumental variable estimator (3) are consistent but the OLS estimator does not minimize the prediction errors as defined by (2). In fact, under the additional assumption of normal errors the estimator (3) is a maximum likelihood estimator and thus asymptotically efficient.[9]

If we would add the assumption that e_1 and e_2 are correlated the recursive property of the model is lost and OLS is no longer a consistent estimator of β_2. The estimator (3) is, however, still consistent and under the assumption of normality a ML estimator. In this example we would thus prefer the "system-wide" estimator (3) whether the model is recursive or not.

4. Choice of estimation criterion and estimator[10]

Let us now turn to the choice of model-wide estimation criterion. The least-squares criteria commonly used assume that we seek parameters estimates such that the *mean* predictions give the smallest possible prediction errors, eq. (2) is an example. However, in micro-simulation we are not only interested in mean predictions, but we want to simulate well the whole distribution of the target variables. This difference in focus between micro simulation and a more conventional econometric analysis might suggest a different estimation criterion.

[9] The estimator (3) is a ML estimator because there is no additional x-regressor in the second relation. The reduced form becomes a SURE system with the same explanatory variable in both equations. In general the ML estimator will depend on the structure of the covariance matrix of the errors.

If the stochastic properties of the microsimulation model were fully specified including families of distribution functions, then the maximum likelihood method would use all the information in model and data to obtain efficient estimates. In practice, however, ML estimation will in general not become feasible. There are several reasons for this: a) For many submodels economic theory does not suggest any parametric family of distributions and we are usually unwilling to make strong assumptions which are not firmly based in theory or in previous research. Instead we might prefer to represent distributions of "residuals" by resampling from empirical distributions. b) In some cases, for instance when the distributions of income and wealth are simulated, we have to work with strongly skewed and highly non-normal distributions. c) A typical microsimulation model includes a rather complex mixture of different submodels, functional forms and dependence assumptions, such that it might become difficult to set up a likelihood function.

Given the general purpose of microsimulation the estimation criterion should not only penalize deviations from the mean but also deviations in terms of higher order moments. A natural candidate estimation principle then becomes the Generalized Method of Moments. The complexity and nonlinearity of a microsimulation model, however, cause difficulties in evaluating the moment conditions. A potential solution to this problem is to use the fact that the model is built to simulate and thus replace GMM by the Simulated Method of Moments.

A discussion of simulation-based estimation does not only lead to new estimators but also highlight the need to change the conventional estimation criteria to one, which is compatible with the simulation context. Assume the following simple model,

$$y_t = g(x_t, e_t, q); \qquad (4)$$

[10] For a discussion of inference problems in microsimulation see Klevmarken(1998).

x_t is an exogenous variable, e_t a random variable with known p.d.f. and θ an unknown parameter.

$$E(y_t \mid x_t) = E(g(x_t, \varepsilon_t, \theta_0) \mid x_t) = k(x_t, \theta_0); \quad (5)$$

We assume that $k(x_t, \theta)$ does not have a closed form.

The basic idea of estimating θ is to obtain a distribution of simulated y-values, y_t^s, with properties which as closely as possible agree with those of the p.d.f. of y_t. It would appear to be a natural approach to choose $\hat{\theta}$ such that it minimizes

$$\sum_{t=1}^{n} (y_t - y_t^s(\hat{q}))^2; \quad (6)$$

However, as shown in Gouriéroux & Monfort(1996) p. 20 this "path calibrated" estimator is not necessarily consistent. To see this consider an example, which differs a little from the one, used in Gouriéroux & Monfort(1996). Assume the following simple model,

$$y = bx + se; \ where\ e \sim IID(0,1) \quad (7)$$

We seek parameter estimates $\tilde{\beta}$ and $\tilde{\sigma}$ such that the model can be simulated,

$$y^s = \tilde{b}x + \tilde{s}e^s; \quad (8)$$

where e^s are draws independently of ε but from the same (known) distribution. Inserting (8) into (6) and solving the first order conditions gives the following estimators,

$$\tilde{b} = (\sum xy - \tilde{s}\sum xe^s) / \sum x^2; \quad (9)$$

$$\tilde{s} = \frac{\sum (bx + se)x\sum xe^s - \sum (bx + se)e^s \sum x^2}{(\sum xe^s)^2 - \sum (e^s)^2 \sum x^2}; \quad (10)$$

From the assumptions made and the additional assumption that $(1/n)Sx^2$ converges to a finite limit when n tends towards infinity, it follows that

$$\operatorname*{p\,lim}_{n\to\infty} \tilde{s} = 0; \text{ and } \operatorname*{p\,lim}_{n\to\infty} \tilde{b} = b; \tag{11}$$

Using this criterion we thus get an inconsistent estimate of σ but a consistent estimate of β. Essentially this estimator tells us to ignore the random drawings of ε^s when we simulate, i.e. only to use mean predictions. As already noted, such a procedure does not agree with the objective of micro-simulation. In this particular model the estimate of β is consistent, but if there was a functional relation between β and σ then the slope would also become inconsistently estimated. It is perhaps possible to generalize this result and suggest that if there is any functional relation between the parameters, which determine the mean path and those determining the dispersion around this path in a micro simulation model, then one cannot use a path-calibrated estimator.

An alternative approach is to use a "moment calibrated" estimator, which minimizes the distance between observed and simulated moments. This approach does not only permit calibration to moments in cross-sectional distributions, but also to transition frequencies and intertemporal correlations, which become important in dynamic models.
Let θ be a vector of size p and x_t a vector of size r. Furthermore let $K(y_t, x_t)$ be a vector function of size q, and

$$E(K(y_t, x_t) \mid x_t, q_0) = k(x_t, q_0); \tag{12}$$

K could, for instance be the identity function and the square of y_t. Also define a r x q matrix $Z_t = I_{q \bullet q} \otimes x_t$. From the exogeneity of x_t it follows that

$$E\left[Z_t(K(y_t, x_t) - k(x_t, q_0))\right] = 0; \tag{13}$$

Because there is no closed form of k(x_t, θ), we will define an unbiased simulator of k,

$$\tilde{k}(x_t, e^s, q) = \frac{1}{S}\sum_{s=1}^{S} K(g(x_t, e_t^s, q), x_t); \tag{14}$$

where e^s is a vector of S independent random errors e_t^s drawn from the p.d.f. of e_t.

A simulated GMM estimator is then obtained as,

$$\hat{q} = \arg\min_{q}\left(\sum_{t=1}^{n} Z_t\left[K(y_t, x_t) - \tilde{k}(x_t, e^s, q)\right]\right) \Omega\left(\sum_{t=1}^{n} Z_t\left[K(y_t, x_t) - \tilde{k}(x_t, e^s, q)\right]\right); \tag{15}$$

where Ω is a r x r symmetric positive semi-definite matrix. As shown Gouriéroux & Monfort(1996) this is a consistent estimator. The covariance matrix of the estimator has two components, one, which is the covariance matrix of the ordinary GMM estimator, and one, which depends on how well k is simulated. An optimal choice of Ω depends on the unknown distribution of y_t. A simulation estimator of the optimal Ω is given in Gouriéroux & Monfort(1996) p. 32.

Two observations are in place:
The number of moment conditions (13) invoked must be no less than the number of unknown parameters, otherwise the model becomes unidentified.

The quadratic expression in (15) can be minimized using the usual gradient based methods if first and second order derivatives with respect to θ exist. If the model includes discontinuities in θ one would have to rely on methods not using gradients.

Micro simulation models that include tax and benefit legislation typically have discontinuities in variables, which may or may not imply discontinuities with respect to behavioral parameters.

5. Model validation

An important part of any model building effort is testing and validation. Validation involves two major issues. First the choice of criterion and validation measure, and second the derivation of the stochastic properties of this measure taking all sources of uncertainty into account. The choice of criterion for validation is of course closely related to that for estimation. As already mentioned, we are not only interested in good mean predictions, but also in good representations of cross-sectional distributions and of transitions between states. When an event occurs becomes important in any dynamic microsimulation exercise. A micro-simulation model is likely to have a number of simplifying assumptions about lack of correlation and independence, both between individuals and over time. For this reason one might expect too much random noise in the simulations and too quickly decaying correlations compared to real data. In addition to model wide criteria one might thus be interested in criteria that focus on these particular properties. Work is needed to develop such measures with known properties.

For a model not to big and complex in structure it might be feasible to derive an analytic expression for the variance-covariance matrix of the simulations, which takes all sources of uncertainty into account: random sampling, estimation and simulation errors. For an example see Pudney & Sutherland(1996). In general micro-simulation models are so complex that analytical solutions are unlikely. Given the parameter estimates the simulation uncertainty can be evaluated if simulations are replicated with new random number generator seeds for each replication. There is a trade off between the number of replications needed and the sample size. The bigger sample the fewer replications.

To evaluate the uncertainty which arises through the parameter estimates one approach is to approximate the distribution of the estimates with a multivariate normal distribution

with mean vector an covariance matrix equal to that of the estimated parameters. By repeated draws from this normal distribution and new model simulations for each draw of parameter values an estimate of the variability in the simulation due to uncertainty about the true parameter values can be obtained.

To avoid the normal approximation one might use sample re-use methods. For instance, by boot strapping one can obtain a set of replicated estimates of the model parameters. Each replication can be used in one or more simulation runs, and the variance of these simulations will capture both the variability in parameter estimates and the variability due to simulation (model) errors. If the boot strap samples are used not only to estimate the parameters but also as replicated bases (initial conditions) for the simulations, then one would also be able to capture the random sampling errors.

Much of the total error in simulated values will come from the choice of a particular model structure. Sensitivity analysis is an approach to assess the importance of this source of error. As pointed out in Citro & Hanusheck(1997) p. 155 “sensitivity analysis is a diagnostic tool for ascertaining which parts of an overall model could have the largest impact on results and therefore are the most important to scrutinize for potential errors that could be reduced or eliminated”. If simple measures of the impact on key variables from marginal changes in parameters and exogenous entities could be computed they would potentially become very useful.

Most models will almost always show deviations between simulated and observed values. If these deviations are within the bounds suggested by the stochastic properties of the simulation exercise, then one might like to constrain the model to simulate these known benchmarks with certainty. The model is aligned or calibrated to the benchmarks. If they were not used when the parameters of the model were estimated, this is a way to include new information. Alignment can thus be seen as a form of constrained estimation. Policy analysts also have another reason to force the model to simulate benchmarks. They think that the whole simulation exercise becomes more credible if the

model reproduces what most people recognize as statistical facts[11]. The idea seems to be that if the model is aligned to known benchmarks it will also do a better job in simulating other variables. This may be true, but it will not be true in general! If the benchmarks are tested against the model and rejected, then the model should be revised rather than aligned.

It should be possible to include constraints of the kind discussed above in the simulation-based approach. Suppose K is the identity function in y_t so the moment condition becomes,

$$E(y_t - E(g(x_t, e_t, q_0)) = 0; \tag{16}$$

The empirical correspondence to the expression to the left of the equality sign is

$$\bar{y} - \frac{1}{n}\sum_{t=1}^{n} \tilde{k}(x_t, e^s, q_0); \tag{17}$$

Suppose now that we know the finite sample mean $\bar{Y}$. How could we use this information? If we also knew the x_t values for all individuals in the finite sample, we could substitute $\bar{y}$ in (17) for $\bar{Y}$ and extend the summation in the second term of (17) to N, and thus get an empirical correspondence to (16) for the whole finite population. In practice this is of course not possible. One only knows the x-observations of the sample, but with known selection probabilities p_t they can be used to compute the following estimate,

$$\bar{Y} - \frac{1}{N}\sum_{t=1}^{n} \frac{1}{p_t} \tilde{k}(x_t, e^s, q_0); \tag{18}$$

[11] For similar reasons policy analysts sometimes want to align to ”official” demographic projections and macro economic forecasts of labor force participation rates, unemployment rates, etc.

The covariance matrix of the resulting estimate $\hat{\theta}$ should now have a third component, which reflects the sampling from the finite population.

6. Conclusions

Microsimulation has the potential of linking micro and macro and enhancing our understanding of fundamental macro relations. It also has the potential to answer questions related to heterogeneity in behavior and differences in outcome of economic and social policy. Microsimulation is particularly well suited for analysis of the *distribution* of well-being.

To achieve this goal we need a major long-term research program that attracts experts from different fields. Such a program should,

- systematically analyze the heterogeneity in micro behavior,
- model the interaction of people, firms and institutions in various markets,
- study and model the detailed influence of true policy parameters on individuals and markets,
- collect adequate data,
- develop and use sound inference methods for estimation, testing and simulation such that the properties of simulation results are known.

In principle inference in microsimulation models is no different from inference in other applications of Economics, but there are practical difficulties due to the large scale and complex structure of a typical simulation model. This and a general caution for too strong assumptions about the stochastic properties of a model suggested that simulation-based estimation and sample re-use methods might be a good approach to the inference problems. These methods are, however, very demanding in terms of computations, and it remains to see if we currently have the computing power needed.

Finally, is such an ambitious research program at all feasible? It would certainly contrast with the individualistic and small-scale research that is most common in Economics. But if it is possible in other disciplines, why not in Economics?

References

Caldwell, S. (1993), "Content, validation and uses of CORSIM 2.0, a dynamic microanalytic model of the United States" paper presented at the IARIW conference on Micro-simulation and Public Policy, Canberra, Australia 1993.

Caldwell, S. (1996), "Health, Wealth, Pensions and Life Paths: The CORSIM dynamic Microsimulation Model", Ch 22 in A. Harding (ed.) *Microsimulation and Public Policy*, North-Holland 1996.

Citro, C.F. and E.A. Hanushek (eds.) (1997), *Assessing Policies for Retirement Income. Needs for Data, Research, and Models,* National Research Council, national Academy Press, Washington, D.C.

Eriksen, T., (1973), *En prognosmodell för den allmänna tilläggspensioneringen*, Riksförsäkringsverket, Stockholm

Fisher, F. (1969), "The Exixtence of Aggregate Production Functions", *Econometrica*, 37,553-577

Gouriéroux, C. And A. Monfort (1996) , *Simulation-Based Econometric Methods*, Oxford University Press.

Hansen, L.P. and J.J. Heckman (1996), "The Empirical foundations of Calibration", *Journal of Economic Perspectives*, 10,1,87-104

Klevmarken, N.A., (1973), En ny modell för ATP-systemet, *Statistisk Tidskrift*, (*Statistical Review*) 1973:5, 403-443

Klevmarken, N.A. (1997), Behavioral Modeling in Micro Simulation Models. A Survey. Working Paper 1997:31, Department of Economics, Uppsala University.

Klevmarken, N.A.(1998), Statistical inference in Micro Simulation Models: Incorporating External Information, Working Paper 1998:20, Department of Economics, Uppsala University, forthcoming in *Mathematics and Computers in Simulation.*

Lutjohann, H. (1974), *Linear Aggregation in Linear Regression*, Stockholm University, Stockholm ISBN91-7222-049-X (diss.)

McCloskey, D. (1983), "The Rhetoric of Economics", *Journal of Economic Literature*, 21,481-517

Orcutt, G.H. (1957) "A new type of socio-economic system", *Review of Economics and Statistics,* 58,773-797

Orcutt, G.H., M. Greenberger, J. Korbel and A. Rivlin, (1961) *Microanalysis of Socioeconomic Systems: A Simulation Study,* Harper and Row, New York

Orcutt, G.H., S. Caldwell and R. Wertheimer, (1976), *Policy Explorations Through Microanalytic Simulation,* The Urban Institute, Washington D.C.

Pudney, S. And Sutherland, H.(1996), Statistical Reliability in Microsimulation Models With Econometrically-Estimated Behavioural Responses, Chapt. 21 in A. Harding (ed.), *Microsimulation and Public Policy,* North-Holland Elsevier.

Sims, C.A. (1996), "Macroeconomics and Methodology", *Journal of Economic Perspectives*, 10,1,105-120

Theil, H. (1954) *Linear Aggregation of Economic Relations*, North-Holland, Amsterdam

Appendix

MICROHUS

A dynamic micro simulation model with behavioral response applied to Swedish households

Introduction

The project to build a micro simulation model for the Swedish household sector was motivated by the need to have a tool which could be used in analyzing not only the direct effects of tax and transfer changes on the distribution of income and wealth but also the indirect effects, i.e. people´s behavioral adjustments to policy changes. Another motive was to use this model as an "umbrella" under which we could successively collect results obtained from analysis of the HUS-project data base (Klevmarken & Olovsson (1993)). These data, a random sample of Swedish households, have given us the sample of individuals and households on which the simulation model operates. They have also, with few exceptions, been the empirical base for the estimation of behavioral relations. A third motive to start the MICROHUS project was to learn how to build a large micro simulation model with behavioral response. The endeavor to build a model which includes both the tax and transfer systems and behavioral relations is a never ending one, as there are always improvements that can be made. In order to get something which could run before all research funds had been used up, we have occasionally found it necessary to sacrifice the ambition of front line research for less sophisticated and quicker solutions. In this sense the model, now christened MICROHUS, should be considered work in progress. It might be useful to "calibrate" peoples expectations by pointing out the major difference between a behavioral model and the type of "accounting" models which are used to simulate the first-order effects of tax policy changes. The latter type of model typically includes all or almost all details of the tax and transfer rules and, depending only on the accuracy of the data used and coding detail, it can reproduce the distribution of income with almost no error (c.f., for instance, the model FASIT developed by Statistics Sweden). In contrast, any attempt to model behavioral relations will necessarily imply sizable specification errors and, as a result, it might not be necessary to include the tax and transfer system with the same degree of detail as in the accounting models. What we hope to accomplish is to get some indication of the relative size of the adjustment effects. If they are small we can be satisfied with accounting models, but if the adjustment effects are not negligible, then more effort has to go into modeling behavior.

The general structure of MICROHUS

MICROHUS is structured in model "blocks" each of which consists of one or more modules. Table 1 lists all blocks and modules currently included in the model.

Preparations is a block which adjusts the age distribution of the 1984 HUS-sample to conform with that of the Swedish population. It compensates for a mildly selective nonresponse in the 1984 HUS-wave. This block also imputes weekly work hours for those employed in 1984 who did not respond to the question about work hours.

The 1984 HUS-sample only included individuals below the age of 75 and for this reason we had to exclude the oldest cohorts from MICROHUS. In the first module of the **Demographics** block every sample member becomes one year older and when a household head becomes 75 the whole household is removed from the sample and put in a file which is later used in the block "Inheritance". Every remaining individual is then exposed to the risk of dying using conventional life tables. Every individual for which death is simulated is deleted from the population and the block "Inheritance" is called to determine how the wealth of the deceased should be allocated among his or her heirs.

In the block "Demographics" there are also modules which simulate the birth of new children, the formation of consensual unions and marriages, and separations. All these modules change the composition of existing households, form new ones and delete old. Fertility is modeled by a Box-Cox type quadratic duration model, which has the hourly wage rate and disposable income as arguments in addition to several socio-economic variables. This model was estimated on HUS-data. Continuous time duration models are also used to model the formation of consensual unions and dissolution of unions. The latter model builds on work by Jan Hoem et.al. and the model was estimated on data from the Swedish fertility survey. The matching of spouses is based on a canonical correlation analysis of the characteristics of spouses. Matching is done randomly by age, schooling and labor market experience.

In the demographic block economic incentives influence the decisions to form a union and find a partner through the schooling and years of experience variables. The more schooling a woman has the higher probability to start a consensual union and to marry. Experience also increases the probability to form a union. The probability to cohabit is an inversely u-shaped function of years of work experience with a peak at about 15 years, while the probability to marry is a continuously increasing function of experience. The probability to have a child also depends on these variables and on the woman´s wage rate and the disposable income of the household. Females with relatively long schooling postpone their first child but tend to have their next child (children) rather quickly. Females with just a few years of experience have a smaller probability to get a child than women with both less and more experience. The wage rate has a negative effect on fertility while there is no consistently positive or negative income effect.

The block **Geographic Mobility** simulates household moves between the big cities and the rest of the country using a constant transition matrix. Households, which are simulated to move, are randomly allocated to municipalities. The probability of moving to a particular municipality is proportional to the number of respondents from that municipality in 1984.

The **Housing** block simulates changes in tenure and consequent changes in market values, tax assessed values and in the expenditures for housing. The probability to move to a new house or apartment is modeled by a probit relation, while tenure choice conditional on a decision to move is simply simulated by a constant transition matrix. The decision to move is assumed dependent on economic incentives. The probit function includes among its explanatory variables labor income, net wealth and the housing costs of the household as well as the marginal tax rate of the household head. Separate models are used for couples and singles. The wealthier a household is and the higher the marginal tax rate is the less likely the household is to move. High labor incomes increase the probability to move for couples but decrease the probability for singles.

The number of observed changes of tenure in the HUS-data was not large enough to permit the estimation of a model, which made tenure choice a function of economic incentives. For this reason we have used a constant transformation matrix.

For each household simulated to move the cost of housing (rent for tenants), interest payments on mortgages and the market- and tax assessed values on owner occupied homes are assigned by a minimum distance method, i e. the values of these variables are copied from an existing household which as closely as possible resemble the receiving household as to size, number of children, age and sex of the head, population density of the housing area, and the earnings and wealth of the household.

The **Labor Market** block simulates entries into and exits from the labor market, hours worked, sickness absence, unemployment and wage rates. This block also includes modules, which simulate the length of schooling and the choice of occupation.

In MICROHUS every individual is assumed to decide about his/her future education at the age of 15. A model for the total length of theoretical education and vocational training was estimated from HUS-data. Due to the major changes in the educational system in post war Sweden this model was estimated for three groups of birth cohorts i.e. people born 1935-44, 1945-54 and 1955-, and also by gender. In simulating the behavior of future cohorts the estimates for the youngest cohorts are used. The length of schooling depends on the father´s and the mother´s schooling, the father´s occupation and whether the simulated individual grew up in a big city, abroad or somewhere else. A general result is that parents' background was of less importance for the young cohorts compared to the older cohorts. The model also includes measures of national unemployment and the relative wage rate of white collar workers compared to blue collar workers in an attempt to capture economic incentives on the schooling decision. However, none of these variables turned out significant. The model assumes that all educational activities take place immediately after compulsory schooling is finished. It thus cannot handle repeated spells of continuing education.

Transitions into the labor force for those who have just finished their schooling are currently simulated only by a constant transition probability (0.9). All individuals 18-64 years old who do not belong to this group and do not participate in the labor force, or are long-term unemployed, on parental leave or on early retirement have a risk to join the labor force and take up work. The model used to simulate this is a Weibull proportional hazard model. The set of explanatory variables are a rich set of interacting demographic variables and the income after tax if not working, if working half-time and if working full-time. The unemployment rate by county is used in an attempt to capture the demand side. People who are simulated to join the labor force are then transferred to a module, which simulates hours of work.

Those who enter the labor market are assigned a wage rate by a conventional earnings function.

Each individual who joins the labor force is also assigned one of four occupations: blue-collar worker, white-collar worker, entrepreneur, manager or executive. The assignment is done by constant transition probabilities, which depend on the father´s occupation and schooling. The model does not currently allow for changes in occupation.

The model, which simulates annual hours of market work is of the Hausman type with random heterogeneity and measurement errors. In addition to the marginal wage rate and virtual income, the following explanatory variables were used: age, number of children less than seven years old, number of household members, if the household owns their home, and if they live in a big city. The model was estimated separately for males and females living in unions. The spouse´s labor income was included in the virtual income, but in no other way the model captures any jointness in the decisions of two spouses. The model was also estimated for singles, but in this case data did not permit a model for each gender. Estimates were obtained from the 1984 HUS cross section of employed and self-employed, using income data from 1983[12]. The budget sets computed were determined by the 1983 income tax system, while income dependent transfers and subsidies did not influence them. The Slutsky conditions were enforced. They turned out to be binding only for females.

The following elasticities were obtained, when evaluated at 2 100 hours, a marginal wage rate of 21 SEK and a virtual income of 56 000 SEK:

[12] We thus use a static model which does not assume any memory nor any forward looking behavior. As a result it might produce simulations with unrealistically high mobility in hours of work. It would become devastating if we attempted to simulate longitudinal life histories, but it might be less of a problem if we are only interested in each cross-sectional distribution of hours.

	Singles	Couples	
	Males and females	Males	Females
wage rate	0.143	-0.013	-0.006
income	-0.581	-0.181	-0.149
compensated	0.601	0.130	0.112

According to these estimates the economic incentives to change a person´s supply of hours are small for couples but a little larger for singles.

When the model is simulated there is no constraint, which prevents a nonpositive outcome. It is interpreted as a transition out of the labor force.

Another way out of the labor force is to retire. Everyone is assumed to take old age pension at the age of 65. Early retirement is simulated using age and time specific transition probabilities obtained from the Swedish National Insurance Board. An individual can choose between full-time or part-time early retirement. This choice is simulated with age specific but constant transition probabilities from the same source. There are thus no economic incentives influencing the retirement decision. People may also be temporarily absent from the labor market, either because they are sick or care for their children or because they are unemployed.

When a child is born the mother is assumed to use the maximum number of months with paid leave. No months of parental leave are allocated to the father.

A Tobit model simulates short spells of absence (< 11 weeks) due to sickness. The marginal income loss due to sickness, after compensation from the sickness insurance but before tax, has a small negative effect on the duration of absence. The more a person loses the shorter he is likely to stay away from work. Females have longer spells of sickness than males and young and old people longer than middle aged. Families with children and families who live in the big cities also have longer spells. The Tobit model does not simulate well very long spells of sickness. To get a good representation also of spells exceeding 10 weeks the Tobit model is supplemented by random drawings from a rectangular distribution.

The model, which simulates months of unemployment, is also a Tobit model. The general level of unemployment is exogenously given by the national unemployment rate. Individual differences are explained by differences in gender, experience, schooling, seniority and age. Well-educated, middle aged individuals and females have shorter spells of unemployment.

The annual work hours simulated by the labor supply model are reduced proportionally to the simulated number of weeks sick and months unemployed. The corresponding compensation from the national sickness insurance is computed as well as the unemployment compensation.

The Labor Market block finally also includes a module, which updates the wage rates of people in the labor force. The model is a combination of an autoregressive structure and a conventional Mincer type earnings function. Separate models are used for males and females.[13]

MICROHUS distinguishes between a few **assets and liabilities**: Real estate (owner occupied house or condominium), mortgages, financial assets, other loans, and other assets. The latter category includes consumer durable. For taxation purposes the model assumes that household wealth is divided equally between two married spouses while cohabiting but not married keep the assets they brought into the union as their personal property.

[13] The model was estimated from the 1984 and 1986 waves of the HUS panel. The log wage rate is thus explained by its lagged value with a lag of two years. The estimated autoregression parameter for females was estimated to 0.637. The predictions from the model however implied an unreasonable decrease in the average female wage rate. For this reason this parameter was increased to 0.785. The estimate for males was 0.801.

In the current version of MICROHUS market values of real estate, financial assets and other assets are adjusted exogenously by average national rates obtained from official statistics. Mortgages and loans are assumed amortized by 3 per cent each year.

For households which change housing or experience a marriage or a separation there are consequential changes in assets, while other households are assumed not to have any changes in assets.

A household, which moves from one house/apartment to another, will get a new real estate market value and new mortgages in the Housing block. Any residual balance between new and old market values and mortgages is picked up as a change in financial assets. Similarly, a household, which moves from a rented apartment to an owner occupied house, will have a decrease in financial assets equal to the difference between the market value of the house and the mortgage.

If two spouses separate each spouse will bring his/her share (50%) of the household wealth to the new households. This implies that a former wife, who keeps the couple's house, will pay to her former husband his share in the house.

If two individuals marry their assets will be added and then divided into equal shares.

Inheritance is another source of wealth changes. The transfer of assets between generations and between deceased and surviving husbands and wives is controlled by the block Inheritance. A separate file keeps track of all children-parents and husband-wife relations. Bequest to more distant heirs are not considered. If a married person dies and the total wealth exceeds a certain relatively low threshold the spouse is assumed to inherit 50 per cent (the legal share) and any children 50 per cent. Otherwise the spouse inherits everything. When a parent dies and there is no living spouse, the children who are alive will share the wealth left by their parent in equal parts. In order not to run the total wealth of the simulation population towards zero, the wealth of a dead person who has no spouse or children is also redistributed among sample households.

The block **Incomes and taxes** collects information about labor supply, wages, wealth and demographic matters to compute labor income, transfer payments and other nonlabor income components which give taxable income. Income and wealth taxes are then computed. The tax scales, tax bases and taxation rules for each year in the period 1984-1992 are applied. Transfer payments are also computed using the rules for each year. This block finally computes disposable income.

The treatment of income components is summarized below:

<u>Labor income and equivalent:</u>

1. Wages and salaries as the product of annual hours worked and a wage rate, both from the labor market block.

2. Sickness allowances are calculated using information about number of weeks sick and earnings from the labor market block and the compensation rules of the public sickness insurance.

3. Parental leave cash benefits are allocated to the mother in households with two spouses and to the household head in households with a single adult. The amount of benefits is calculated from the compensation rules based on the mother's (household head's) earnings. It is assumed that the maximum number of months with full benefits are used in one sequence.

4. Old age pensions are given to everyone who has reached the age of 65. The benefits are basic pension, pension supplement and supplementary pension (ATP). Pension scores, which determine the size of the supplementary pension, are calculated on the basis of the labor income generated in the labor block. For individuals who could have worked before 1984, the first year of simulation, pension scores were

assigned for the period 1960-1983 by a minimum distance method using micro data from the National Social Insurance Board (RFV).

5. Early retirement pensions include disability pension and early partial pension. For those who have been simulated to get these benefits their size is calculated using the rules of the social security system and the ATP pension scores as described above.

6. Note that pensions which are paid as a result of contracts between unions and employers and private pension policies are not included in MICROHUS.

7. There are two types of unemployment compensation in Sweden: Benefits from an unemployment benefit association (unemployment insurance) and cash labor market assistance (KAS). It is not possible to draw benefits from both programs at the same time. MICROHUS allocates unemployed randomly to one of these benefits. The allocation probabilities were obtained from national statistics. Unemployed who get unemployment insurance are assumed to get a certain percentage of the labor income they had before they became unemployed. Those who are allocated to KAS get the maximum daily rate of a particular year.

8. Other labor incomes, from a taxation point of view, are free use of a car, free board, housing or other perquisites, travelling and subsistence allowances and other kinds of labor compensation. These incomes are not computed in MICROHUS. As a consequence no deductions for expenses are made in excess of the standard deduction everyone can make. Periodical allowances are also assumed to be zero.

Nonlabor income

1. Income of capital is computed as a percentage of the household´s value of financial assets (excluding owner occupied homes and other real estate) less loans other than mortgages. This percentage is proportional to the discount rate and the factor of proportionality is determined randomly each year by a procedure based on observed rates of returns in 1984 and 1986.

2. No capital gains or losses are currently included and taxed in MICROHUS.

3. Housing allowances for nonretired households depend on income, the housing standard, the heating costs and on the number of children. MICROHUS computes these benefits for each year 1984-1992. Municipal supplements are, however, not included.

4. Housing allowances for households with pensioners (KBT) depend on the composition of the household, its incomes and on the housing costs. The benefits depend in addition on municipality, which is taken into account by MICROHUS.

5. Child allowances only depend on the number of children and how old they are. The benefits are computed by MICROHUS and allocated to the mother.

6. Alimonies are currently not included in MICROHUS.

Imputed rents and tax deductions

1. Imputed rent on owner occupied houses less deductions for interest payments are calculated based on information from the housing block. The result is "deficit in one or two family houses".

2. "General deductions" (allmänna avdrag) are estimated based on observed behavior in the 1984 and 1986 HUS-surveys. For the income years 1983 and 1985 the distributions of the ratios between "general deductions" and "total income" were computed separately for men and women. From these distributions MICROHUS randomly selects a ratio which is used to compute each individual deduction. The 1983

distributions were used for 1984 and the 1985 distributions thereafter. In 1991 the tax system changed which had implications for the general deductions. Deductions from income from employment were by and large limited to deductions for payments to private pension policies.

3. Imputed rent on real estate including owner occupied homes, the so called "garanti summa" was added to the income which had to be declared for local tax assessment until 1986. It was 1.5% of the taxassessed value.

4. "Extra deductions" motivated, for instance, by costs arising from disability or long term sickness are assumed zero.

Taxes

1. Local and national income taxes are computed. In computing local taxes the rate for each municipality and year is used.

2. The wealth tax is computed too. In Sweden there is joint taxation of wealth. After the joint tax has been computed it is allocated to the wealth holders of the household proportionally to each individual´s taxable wealth.

The behavioral models used in most modules have been estimated in constant prices, i.e. monetary variables are in constant prices, while the tax system has tax bases, thresholds etc. in current prices. Monetary values thus have to be converted from fixed prices to current prices when the simulation enters the tax modules and back again to fixed prices when the simulation leaves the tax modules. The conversion is done by CPI, which is given exogenously.

In Sweden about 50 per cent of all children below school age use **public childcare.** Parents pay a price for these services, which on average only covers 10-15 per cent of the running costs of childcare centers. The structure of these charges varies from one municipality to another. In some there is a flat rate of family income with a lower and upper threshold, in other municipalities the rates charged are income dependent. In MICROHUS all households with children below the age of 7 are exposed to the risk of using public childcare. The model used is an ordered probit model for the number of months in public childcare. The variables used to explain the transition probabilities are: the number of children less than seven years old in the household, full-time child care charge, a measure of the demand situation which is the share of kids below seven using public child care in the municipality, years of schooling of the spouse with most schooling, if custodian is single, and a measure of family income. For all households, which are simulated to use public childcare, a probit model is used to determine if the children spend half-time or full-time at the childcare center. For the years 1984-1986 and 1992 we had access to the fee schedules of most municipalities. For missing municipalities we have used the fee schedule recommended by the Association of Swedish Municipalities (Kommunförbundet). For 1987-1989 the fees of 1986 were used and for 1990-1991 and after 1992 the 1992 fees were used. The marginal effects on the budget set of a household, which arises from these fee schedules have not yet been integrated into the labor supply modules.

MICROHUS finally also includes a few modules which assign identification numbers and household numbers and store lagged values from previous years. There are also modules, which help in extracting output data from MICROHUS as the simulations proceed.

Each simulation consists of a number of runs and each run is a full sequence calling all or a selected number of modules and simulating the changes in one year. The simulation is done recursively, i.e. each module is called sequentially (approximately in the order of presentation above) and a later module cannot influence the outcome of an earlier module until in the next run. This implies that all demographic changes are determined before changes in housing and labor supply, and geographical moves are simulated before changes in housing which in turn precede changes in labor supply. Joint decisions about, for instance, housing and job are thus

ruled out. This is primarily done for reasons of simplicity. A model of joint housing and job decisions would become rather complex.

The model is programmed in fortran. The program has about 23000 lines of code. It runs, for instance on a pentium notebook with a 100 mhz processor. A full simulation for the years 1984-2000 with a starting population of about 4500 individuals takes about 30 minutes on such a small machine.

Applications

The model has been used in an evaluation of the 1990/91 tax reform, see Klevmarken & Olovsson(1996), and in a study of the impact of demographic changes on the income distribution, Klevmarken(1994).

References

Klevmarken, N.A., 1994, The impact of demographic changes on the income distribution: Experiments in microsimulation. Paper presented at the 8th annual conference with the European Society for Population Economics, June 2-4, in Tilburg, Holland.

Klevmarken, N.A. and P. Olovsson, 1989, Hushållens ekonomiska levnadsförhållanden (HUS). Teknisk beskrivning och kodbok, Department of Economics, Gothenburg University

Klevmarken, N.A. and P. Olovsson, 1993, *Household Market and Nonmarket Activities. Procedures and Codes 1984-1991*, The Industrial Institute for Economic and Social Research (IUI), Almqvist & Wicksell International, Stockholm

Klevmarken, N.A. and P. Olovsson, (1996) "Direct and behavioural effects of income tax changes – Simulations with the Swedish model MICROHUS", chapter 10 in A. Harding (ed.), *Microsimulation and Public Policy*, Contributions to Economic analysis, North-Holland, Amsterdam

Table 1 ***The general structure of MICROHUS, A microsimulation model of the Swedish household sector***

Blocks and modules:

PREPARATIONS
- EXPANS
- DISTWEEK

DEMOGRAPHICS
- DEMOAGE
- DEMODEA
- DEMOSEP
- DEMOLEA
- DEMOCON
- DEMOSPO
- DEMOMAR
- DEMOFER

INHERITANCE
- HEIRS
- HEIRSCOMP
- INHERIT
- SOLEIN
- DIVIDE
- NETTO

- BALANS
(see block WEALTH)

GEOGRAPHIC MOBILITY
- GEOURB
- GEOCOM

HOUSING
- HOUSENEW
- HOUSETEN

LABOR MARKET
- NEWPENS
- EDULEN
- OCCUPATION
- LABPARS
- LABPAR
- LABOINIT 2
- LABOSUPP
- LABOWAGE
- LABOABSE
- LABOUNEM

- KINKAR

- LABOINSUR

WEALTH
- BALANS

INCOMES AND TAXES
- CASHIN
- TAXATION

- FORPENG
- SJUPENG
- ALDPENS
- FORPENS
- DELPENS
- LABOINSU
- GRUND
- TILLA
- MARGA
- FORMA
- SKRATT

- TAXNEW
  - STATAX
  - FORMTAX
  - EXREDUK
  - GRUNDAX 1
  - GRUNDAX 2
  - BEGAL
  - REDUK
- TRANSBOBI
- REVIHINK
- TRANKBT
  - EXPEHOME
  - EXPEHEAT
- BARNBID (This module is called immediately before LABOSUPP)

CHILDCARE
- CHMONTHS
- CHTIME

PRICE LEVELS
- DEFLATE
- INFLATE

ADMINISTRATION
- PREPAR
- STAT
- STATENK
- LAGUPDATE
- REGEX
- REGEZ
- EXTRACT