Department of Economics
Uppsala University
P.O.Box 513
S-751 20 Uppsala
Sweden
Fax:+46 18 471 14 78

Working paper 1998:
October 1998

STATISTICAL INFERENCE IN MICRO SIMULATION MODELS: INCORPORATING EXTERNAL INFORMATION

N. Anders Klevmarken

Statistical inference in micro simulation models: Incorporating external information

N. Anders Klevmarken*

Abstract

In practical applications of micro simulation models very little is usually known about the properties of the simulated values. This paper argues that we need to apply the same rigorous standards for inference in micro simulation work as in scientific work generally. If not, then micro simulation models will loose in credibility. The paper first discusses how the structure of the model will determine inference and then follow sections on estimation and validation. Differences between inference in static and dynamic models are noted and then the paper focuses on the estimation of behavioral parameters. There are three themes: calibration viewed as estimation subject to external constraints, piece meal vs. system-wide estimation, and simulation based estimation.

Keywords: micro simulation, alignment, calibration, system-wide estimation, simulation-based estimation

JEL Classification: C51, C52

This report was prepared for the Conference on Methodological Issues in Official Statistics, Stockholm, October 12-13th 1998

---

* Department of Economics, Uppsala University, P.O. Box 513, S-75120 UPPSALA, Sweden. Fax +46 18 471 1478, email anders.klevmarken@nek.uu.se

1. Introduction

Inference in micro simulation models (MSM) is in principle no different from statistical inference generally, but in current practice the inference aspects have been neglected. One has been satisfied if the model runs and approximately tracks observed data. The large size of a typical MSM and the difficulties to get coherent data has made many researchers and practitioners accept ad hoc methods.

There are, however, practical problems with inference in MSM related to the large number of relations and conditions, the frequent use of nonstandard functional forms often including discontinuities, and the fact that data typically are obtained from many different sources.

Micro simulation aims at statements about the distribution of some endogenous variables (for instance, the distribution of incomes) defined on a population (for instance, the population of Swedes in a particular year), given certain policy assumptions (for instance assumptions about tax rates) and initial conditions. These initial conditions are usually given by a sample of individuals on which the MSM operates. In the simulation sample values are changed or updated, and the new sample values are used to estimate properties of the distribution of interest (for instance a total, a mean or a Gini coefficient).

A proper inference usually involves several random experiments. One is drawing the sample of initial conditions, another is the random experiment or process assumed to generate population data, and a third is the generation of random numbers in the simulation experiment. The choice of methods is also determined by the mode of inference, whether there is an inference to a finite population or a “super population”.

Because the random experiments involved and the mode of inference in static micro simulation in general is different from that of dynamic micro simulation it is useful first to discuss inference in static models and then turn to dynamic models. Then follows a section on the estimation of behavioral models. Although primarily based on the “super population” thinking of dynamic models much of that to be said about incorporating external information and simulation estimation also applies to static models. The paper ends with a brief section on validation and with a few concluding remarks.

2. Inference in static MSM

The simplest case of a static model is one without behavioral response relations. It only includes a set of deterministic rules, for instance tax and benefit rules translated into computer code. The FASIT model of Statistics Sweden is an example. Given a sample of pretax incomes it computes taxes, benefits and disposable incomes for each individual in the sample. In this case there is no model-based inference but only an inference from the sample of initial conditions (pretax incomes) to the population from which this sample was drawn. In this case an inference to the finite population is meaningful and usually also desired.

If the sample of initial conditions is a probability sample this would seem to be a standard application of sampling theory. But usually the sample was drawn from a population dated a few years ago while an inference is desired to a population, which is present today, and in general these two populations differ. This problem is usually handled by reweighting. The sample weights are adjusted such that a standard inference will reproduce the observed distribution of certain variables in the present population. One might, for instance, know that the age distribution and the distribution of schooling have changed and then seek to adjust the sampling weights accordingly. A technical approach to achieve this is calibration, see Merz(1993, 1994) and Lindström(1997). The idea is to obtain new weights, which are so close to the old ones as possible, but make the simulated values aggregate to know totals. Closeness is defined by some measure of distance. The choice of distance measure is rather arbitrary, but is has been shown that certain distance functions give estimators which are well-known in the sampling literature (Deville, J.C. And Särndal, C.E. ,1992, and Lundström, 1997). Although calibration estimators aggregate to known totals this is no guarantee one obtains an inference to the desired population. The problem might remain if the knowledge of totals does not include the key variables of interest, or if not only the center of location but also the dispersion of key variables have changed. Without a thorough analysis of the causes to population changes any reweighting become ad hoc. Formulating a model, which captures causal relations, on the other hand leads into a dynamic MSM.

Static MSM:s can also include behavioral relations, for instance, labor supply as a function of the budget set (incomes, taxes and benefits). A static model (in the usual economic sense) has no time dimension, but in practice a micro simulation analyst wants to say something about a population in real time. It follows from the tax and benefit rules that a static tax-benefit model without behavioral relations gives the immediate, first-order effects of tax and benefit changes, but if behavioral relations are included there is an issue about their interpretation. Does a labor supply relation, for instance, give the behavioral response which materialize within a year, or does it give the total accumulated effect until some steady state is reached? Most economists probably think of static models in the latter sense. But, this raises new issues. To test and estimate such a model one needs a sample of individuals who have all reached a steady state. Is the adjustment process so quick that a random cross-section of individuals is suitable for inference?

Let us assume that this is the case. An inference would then have to account both for the random uncertainty, which arises because the model is simulated on a sample of initial conditions, and the uncertainty which is generated by the estimated behavioral model. The latter will include two components. The first arises because the unknown parameters are estimated. The properties of these estimates depend on the properties of the model, how data were obtained and on what estimation method was used. The second component arises because invoking a random number generator simulates the estimated model. The properties of the random numbers may depend on estimated parameters.

Assume the following model

$$y_i = g(x_i, e_i, q, p) \qquad (1)$$

i symbolizes a single individual or observation, π is a vector of policy parameters, θ a vector of unknown behavioral parameters and ε a vector of random errors from a known distribution. x is a vector of known values on exogenous variables, while y is a vector of simulated values. This model applies to everyone in a finite population,

$$Y = \{y_1, \ldots\ldots y_N\};\ and\ \ X = \{x_1, \ldots\ldots x_N\}; \qquad (2)$$

X is thus a matrix of given, exogenous numbers, while Y is a random matrix dependent on the random variables ε through the function g.

One only observes a random sample,

$$y = \{y_1,.....y_n\}; and\ x = \{x_1.....x_n\}; \tag{3}$$

obtained by a sampling design P. Assume further that the whole purpose of the simulation experiment is to estimate some statistic for the distribution of Y. Let this statistic be μ(Y). One can, for instance, think of μ as a mean, a coefficient of variation or a Gini-coeffient. Because ε is a random variable μ is also random with properties conditional on X, θ and π. For future reference we will only write μ as an explicit function of x and θ. π is suppressed. μ(Y|X,θ) is estimated by $\hat{m}(y, x \mid e, q, P)$. It is assumed that

$$E_P(\hat{m}) = m; \tag{4}$$

In practice one has to use some estimator to estimate the unknown parameters θ. Lets denote the estimates $\hat{q}$, and look upon $\hat{q}$ as a vector of random variables obtained from a distribution the properties of which we can at least estimate. θ needs not necessarily be estimated from the same sample used for simulation but these estimates could have been obtained from external data. The result of the simulation can now be written,

$$\hat{y}_i = g(x_i, e_i, \hat{q}, p); \tag{5}$$

The estimator $\hat{m}$ is thus a stochastic function of ε, $\hat{q}$ and the sample drawn from the finite population.

The purpose of the simulation is to estimate,

$$E_e(m(Y \mid X, q)); \tag{6}$$

The error committed in a single simulation thus becomes,

$$
\begin{aligned}
&\hat{m} - E_e(m(Y \mid X, q)) = && \text{Total error} \\
&\qquad \hat{m} - m(Y \mid X, \hat{q}) + && \text{random sampling error} \\
&\qquad m(Y \mid X, \hat{q}) - E_e(m(Y \mid X, \hat{q})) + && \text{random simulation error} \\
&\qquad E_e(m(Y \mid X, \hat{q})) - E_{\hat{q}} E_e(m(Y \mid X, \hat{q})) + && \text{random estimation error} \\
&\qquad E_{\hat{q}} E_e(m(Y \mid X, \hat{q})) - E_e(m(Y \mid X, q)); && \text{systematic estimation error}
\end{aligned}
\tag{7}
$$

The last nonrandom bias component arises because the functions g and μ need not be linear. A micro-simulation model g is likely to be highly nonlinear. Even if the estimates $\hat{q}$ are unbiased, this property does not carry over in nonlinear transformations. By using a Taylor expansion of $m(\hat{q})$ at the true θ, and a large sample argument one might be able to show that this bias component tends to zero if $\hat{q}$ tends to θ.

Disregarding the bias component we have three error components. They can be written,

$$
E_{\hat{q}e}(Var_P(\hat{m})) + E_{\hat{q}}(Var_e(m \mid \hat{q})) + Var_{\hat{q}}(E_e(m \mid \hat{q})); \tag{8}
$$

It is not obvious how they should be estimated. The problem is that we do not know and cannot simulate the x-values not included in our sample. An inference has to be conditional on the observed x in the sample. If the sample is large and drawn according to a simple design like simple random sampling, then this might not be a great problem, but in general we would have to rely on sampling theory providing us with consistent estimators.

Suppose, for instance, that the super-population statistic μ can be written as a sum of individual contributions,[1]

$$m = \sum_{i=1}^{N} u(y_i); \tag{9}$$

then a consistent Horvitz-Thompson estimator is

$$\hat{m} = \sum_{i=1}^{n} w_i u(y_i); \tag{10}$$

where the w:s are the sampling weights. Expressions for the sampling variance of this estimator can be obtained from the sampling literature and used to compute the first component of expression (8). The two remaining components can in general be estimated by replicated simulations over the domain of $\hat{q}$ and ε. Using a large sample argument and assuming that θ was estimated by maximum likelihood Pudney & Sutherland(1996) gave an explicit variance expression for a few statistics.

3. Inference in dynamic MSM.

In a dynamic MSM there is no constant population to which an inference can be drawn, because the model defines how the population changes both in size and in composition. Only an inference to the super-population defined by the model would seem meaningful. Let's write the model in the following way,

$$y_t = g(y_{t-1},\ldots y_{t-s} \mid y_0, e_t, q); \tag{11}$$

where y_0 is a vector of initial conditions, in practice set by the sample on which the simulations are done. Using the same notation as before, we are thus interested in estimating

[1] Many of the statistics commonly used to describe simulation results are functions of one or more sums of individual contributions.

$$E_{e_1,\ldots e_t,y_0}(m(y_t)) = \int \ldots \int \int mf(e_1)\ldots f(e_t)\, f(y_0)de_1\ldots de_t dy_0; \qquad (12)$$

where it is assumed that e_s, e_t *and* y_0 are independent for all s≠t. The distribution of y_0 will in practice become estimated by the corresponding empirical distribution function obtained through the sample of initial conditions. One could either condition on the sample of initial conditions, if t is large and the model has some ergodic properties the influence of y_0 becomes small, or one could use the bootstrap technique to evaluate the random influence from the choice of initial conditions sample. The distribution of e_t is assumed known.

In general it will not be possible to evaluate expression (12) analytically, but by replicated drawings from the distribution of e_t a number of replications of μ is obtained, and the mean of these μ-values is an unbiased estimate of expression (12). This procedure assumes that the parameters θ are known. In practice they are not and have to be replaced by some estimates $\hat{q}$. This implies that the simulated estimate of (12) is a random function of $\hat{q}$. If μ and g satisfy certain regularity conditions[2] and if $\hat{q}$ is a consistent estimate of θ, then the estimate of E(μ) is consistent too. But even if $\hat{q}$ is unbiased, the estimate of E(μ) is in general not unbiased, because μ and g are in general nonlinear functions. By replicating also over the domain of $\hat{q}$ we might thus like to estimate $E_{\hat{q}} E_{e_1,\ldots e_t,y_0}(m(y_t \mid \hat{q}))$. These replicated simulations will also give an estimate of the corresponding variance.

4. Estimation of parameters in behavioral models

4.1 Model alignment using external information – an estimation problem

Our limited capacity as model builders, the difficulties to get good comprehensive data from which the model parameters can be estimated, and the piece meal approach usually adopted in practice to estimate the model sub-model by sub-model, all contribute to deviations of simulated values and distributions from observed data. To make the model "stay on track" some model builders have aligned their models to external benchmark data. Population totals and means from official statistics or estimates from surveys not used to estimate the model are sometimes used as benchmarks. If a model is to gain credibility with users they often require that the model is able to reproduce the basic demographic structure of the population and predict well-known bench marks like for instance, the labor force participation rate, the unemployment rate, the mean and dispersion of disposable income, etc. For this reason model builders have forced their models to predict these numbers without error. In CORESIM, for instance, adjusting the simulated values (and not the parameter estimates) does this alignment.

Alignment is usually done by simple proportional adjustments, but there are also more sophisticated procedures. The ADJUST procedure developed by Merz(1993, 1994b) and originally designed for reweighting in static models (cf. above) might also be used for alignment. However, in this context the whole approach appears even more ad hoc than when it is used for reweighting. It does not consider the stochastic properties of the model at all.

A natural way to incorporate this kind of externally given information is to look upon the estimation problem as one of constrained estimation. Assume the micro-simulation model can be written in the following way.

$$Y_t = g(Y_{t-1}, X_t, \varepsilon_t, \theta) \tag{13}$$

where $Y_t = \{y_{ikt}\}_{n\times K}$ is a matrix of K current endogenous variables for n individuals. $X_t = \{x_{ilt}\}_{n\times L}$ a matrix of exogenous variables, $\varepsilon_t = \{\varepsilon_{imt}\}_{n\times M}$ a matrix of random errors with expectation zero and some variance-covariance matrix Ω, and θ a vector of P parameters. Assume also that a sample of Y and X-variables is available for n individuals in T time periods.

Given some estimate of θ say $\hat{\theta}$, it is possible to define the following predictions within the sample period,

[2] MSM often include discontinuities which could imply that these regularity conditions do not hold, but models are more likely to be continuos in the behavioral parameters θ than in variables, the values of which are determined by legislation and government rules.

$$\tilde{Y}_1 = g(Y_0, X_{1,} \tilde{\varepsilon}_1, \hat{\theta}) \tag{14a}$$

$$\tilde{Y}_2 = g(Y_1, X_{2,} \tilde{\varepsilon}_2, \hat{\theta}) \tag{14b}$$

$$\vdots$$

$$\tilde{Y}_T = g(Y_{T-1}, X_{T,} \tilde{\varepsilon}_T, \hat{\theta}) \tag{14c}$$

where $\tilde{\varepsilon}_t$ is a matrix of random numbers drawn from a random number generator or an empirical distribution function.

Define $Y = \{Y_t\}_{n \cdot TxK}$ and $\tilde{Y} = \{\tilde{Y}_t\}_{nT \times K}$ and assume that there is a criterion function defined on the difference $Y - E(\tilde{Y})$, say $L(Y - E(\tilde{Y}))$, where E is the mathematical expectation over the distribution of ε_t conditional on Y_0, X_1, ...X_T and θ. Estimates $\hat{\theta}$ are in principle obtained by maximizing *L* with respect to θ.

Assume now that benchmark data are available in the form of population totals $\bar{y}_{kt}$ for some year t within the sampling period. Define $\bar{Y} = \{\bar{y}_{kt}\}_{1 \times K}$. If the sample of individuals was drawn by simple random sampling the model estimate of this total is

$$\bar{\tilde{Y}}_t = \frac{N}{n} J' g(Y_{t-1}, X_t, \tilde{\varepsilon}_t, \hat{\theta}) \tag{15}$$

where J is a n dimensional vector of ones. If *n* is large enough to make the effects of random variation through $\tilde{\varepsilon}_t$ small, one could maximize the estimation criterion *L* subject to the constraint $\bar{Y}_t = \bar{\tilde{Y}}_t$.

If the external information applies to a date outside the sample period $\bar{\tilde{Y}}_t$ has to be computed slightly differently,

$$\bar{\tilde{Y}}_t = \frac{N}{n} J' g\left(\tilde{Y}_{t-1}, X_t \tilde{\varepsilon}_t, \hat{\theta}\right) \tag{16}$$

In the special case of a linear model this estimation problem reduces to a well-known constrained estimation problem found in textbooks. To simplify, assume there is only one endogenous variable and that the estimation criterion is the usual least-squares criterion. Then the model becomes

$$\begin{aligned} &Y_t = \{Y_{t-1}, X_t\}\theta + \varepsilon_t; \\ &E(\varepsilon_t | Y_{t-1}, X_t) = 0 \end{aligned} \tag{17}$$

Let $Z_t = \{Y_{t-1}, X_t\}$, $Z = \{Z_t\}$, and $Y = \{Y_t\}$. Assume a total $\bar{y}_{t_0}$, is known for period $t_o \leq T$. The constraint then becomes

$$\frac{N}{n} J' Z_{t_0} \theta = \bar{y}_{t_0} \tag{18}$$

Minimizing the sum of squared residuals subject to this constraint gives the usual constrained least-squares estimator,

$$\hat{\theta}^* = \hat{\theta} + (Z'Z)^{-1} R' \left[R (Z'Z)^{-1} R' \right]^{-1} \left(\bar{y}_{t_0} - R\hat{\theta} \right) \tag{19}$$

where $R = \frac{N}{n} J' Z_{t_o}$ and $\hat{\theta}$ the unconstrained least-squares estimator. In this case there is thus a simple adjustment of the least-squares estimator which can be used also when the external information became available after the model was estimated. If the model is nonlinear there is in general no such simple adjustment factor. Depending on the model structure the whole model might have to be re-estimated when new external information becomes available. It is straightforward to derive the variance-covariance matrix of the estimator in eq. (19), but for a nonlinear estimator it is not. One would either have to resort to large sample theory or, if simulations are not too time consuming use sample re-use methods like bootstrapping and jack-knifing.

It could be computationally easier to align the predictions to external information rather than to reestimate all parameters. If so one might thus prefer to do that, in particular if one is less interested in the parameter estimates as such but more in the predictions they produce. $\hat{\theta}^*$ in eq. (19) is a BLUE among those estimators which satisfy the external constraint. Predictions obtained with this estimator are BLUP. Given a matrix $Z_\tau = \{\tilde{Y}_{\tau-1}, X_\tau\}$ of initial conditions the predictions become,

$$\tilde{Y}_\tau = Z_\tau\hat{\theta}^* = Z_\tau\hat{\theta} + Z_\tau(Z'Z)^{-1}R'\left[R(Z'Z)^{-1}R'\right]^{-1}\left(\bar{y}_{t_0} - R\hat{\theta}\right). \qquad (20)$$

The last term of this expression gives the necessary alignment of the prediction. One may note that not even in this simple linear case it is a proportional adjustment. In our notation and for the case of K=0 and only one constraint an alignment proportional to the prediction error $\left(\bar{y}_{t_0} - R\hat{\theta}\right)$ can be written

$$1 + (R\hat{\theta})^{-1}\left(\bar{y}_{t_0} - R\hat{\theta}\right); \qquad (21)$$

which differs from the alignment factors obtained from eq. (20),

$$I + Z_\tau(Z'Z)^{-1}R'\left[R(Z'Z)^{-1}R'\right]^{-1}(\bar{y}_{t_0} - R\theta)\left[(Z_\tau\hat{\theta})'(Z_\tau\hat{\theta})\right]^{-1}(Z_\tau\hat{\theta})'; \quad (22)$$

where $\left[(Z_\tau\hat{\theta})'(Z_\tau\hat{\theta})\right]^{-1}(Z_\tau\hat{\theta})'$ is a generalized inverse of $(Z_\tau\hat{\theta})$. In this case each individual gets its own alignment factor. One may also note that in a model with more than one endogenous variable a constraint which applies to one variable will in general not only imply an alignment of that particular variable but also of all other variables. Furthermore, in nonlinear models there will in general not exist as simple alignment factors as in the linear case.

More or less explicitly the discussion above was based on the assumption that the sample used in the simulations had a size sufficiently large to justify the treatment of external data as exact constraints. If this is not the case one might not like the simulated total (mean) to equal the external total (mean) exactly but allow for the built in stochastic variation in the model. If the external data are estimates rather than population parameters then that is another reason not to enforce an exact equality. A natural approach to incorporate uncertain external information is that of mixed estimation, a technique, which is well developed for linear models in many textbooks, but less developed for nonlinear models.

4.2 Model-wide or piece meal estimation

Given the complexity and mixture of model types and functional forms in a large MSM its parameters are usually estimated in a piecemeal way, sub-model by sub-model. This is sometimes necessary because one does not always have access to one large sample including all variables, but have to use several samples collected from different sources. None-the-less the piece meal approach may be inappropriate. It depends on the model structure. If the model has an hierarchical or a recursive structure and if the stochastic structure impose independence or lack of correlation between model blocks or sub-models, then a piece meal approach can be justified (cf. The discussion in Klevmarken, 1997).

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

$$y_{1t} = \beta_1 x_t + \varepsilon_{1t}; \qquad y_{2t} = \beta_2 y_{1t} + \varepsilon_{2t}; \qquad E(\varepsilon_i \varepsilon_j) = \begin{cases} \sigma_1^2 & \text{if } i=j=1. \\ \sigma_2^2 & \text{if } i=j=2. \\ 0 & \text{if } i \neq j. \end{cases} \tag{23}$$

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 $\hat{y}_1 = \hat{b}_1 x_t$ while predictions of y_2 outside the sample range are $\hat{b}_2 \hat{y}_1$. However, this suggests the following model-wide criterion,

$$\frac{1}{s_1^2}\sum_t (y_{1t} - \hat{y}_{1t})^2 + \frac{1}{s_2^2}\sum_t (y_{2t} - \hat{b}_2 \hat{y}_{1t})^2; \tag{24}$$

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

$$\hat{b}_2 = \frac{\sum_t y_{2t} x_t}{\sum_t y_{1t} x_t}; \qquad (25)$$

In this case both the ”piece meal” OLS estimator of β_2 and the ”system-wide” instrumental variable estimator (25) are consistent but the OLS estimator does not minimize the prediction errors as defined by (24). In fact, under the additional assumption of normal errors the estimator (25) is a maximum likelihood estimator and thus asymptotically efficient.[3]

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 b_2. The estimator (25) is, however, still consistent and under the assumption of normality a ML estimator. In this example we would thus prefer the “system-wide” estimator (25) whether the model is recursive or not.

In applied micro simulation work it might not always be necessary to insist on efficient estimators. Usually large micro-data sets are used for estimation and then also the variance of less efficient estimators might become acceptable. In the example above we could perhaps do with the piece meal OLS estimator if the sample is large, but only if the model is recursive.

Finally, a more general comment on the choice of estimation criterion is in place. The least-squares criteria commonly used assume that we seek parameters estimates such that the *mean* predictions give the smallest possible prediction errors, eq. (24) 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. We will return to this topic in section 4.3 .

4.3 Simulation-based estimation[4]

[3] The estimator (25) 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.
[4] This section relies to a large extent on Gourieroux & Monfort(1996)

The complexity and nonlinear character of MSM and the fact that they are designed to simulate suggest that simulation-based estimation is a feasible approach to obtain system-wide estimates. 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,

$$y_t = g(x_t, e_t, q); \qquad (26)$$

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

$$E(y_t \mid x_t) = E(g(x_t, e_t, q_0) \mid x_t) = k(x_t, q_0); \qquad (27)$$

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 $\hat{q}$ such that it minimizes

$$\sum_{t=1}^{n} (y_t - y_t^s(\hat{q}))^2; \qquad (28)$$

However, as shown by Gourieroux & 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 Gourieroux & Monfort(1996). Assume the following simple model,

$$y = bx + se; \; where \; e \sim IID(0,1) \qquad (29)$$

We seek parameter estimates $\tilde{b} \; and \; \tilde{s}$ such that the model can be simulated,

$$y^s = \tilde{b}x + \tilde{s}e^s; \qquad (30)$$

where e^s are draw independently from ε but from the same (known) distribution. The estimation criterion (28) now becomes,

$$L = \sum_{i=1}^{n} (y_i - bx_i - se_i^s)^2; \tag{31}$$

Solving the first order conditions gives the following estimators,

$$\tilde{b} = \frac{\sum xy - \tilde{s} \sum xe^s}{\sum x^2}; \tag{32}$$

$$\tilde{s} = \frac{\sum (bx + se)x \sum xe^s - \sum bx + se)e^s \sum x^2}{(\sum xe^s)^2 - \sum (e^s)^2 \sum x^2}; \tag{33}$$

From the assumptions made and the additional assumption that $\frac{1}{n}\sum x^2$ converges to a finite limit when n tends towards infinity, it follows that

$$\underset{n \to \infty}{p\lim}\ \tilde{s} = 0; \tag{34}$$

$$\underset{n \to \infty}{p\lim}\ \tilde{b} = b; \tag{35}$$

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 were 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, for instance

$$\hat{q} = \arg\min_q \left[\frac{1}{n}\sum_{t=1}^{n} y_t - \frac{1}{n}\sum_{t=1}^{n} y_t^s(\hat{q}) \right]^2; \tag{36}$$

which is consistent. To demonstrate this in a simple example let's use the same model as above but assume that $\beta=\sigma$ and that x is unity. (This is the model Gourieroux & Monfort(1996) used to demonstrate that path calibrated estimators are not consistent in general.) Minimization of the criterion,

$$L = \left[\bar{y} - \frac{1}{n}\sum (s + se^s) \right]^2 \tag{37}$$

gives the estimator,

$$\tilde{s} = \frac{\bar{y}}{1-\bar{e}^s}; \tag{38}$$

which is consistent.

More generally, 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

$$E(K(y_t, x_t) \mid x_t, q_0) = k(x_t, q_0); \tag{39}$$

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

$$E\left[Z_t(K(y_t, x_t) - k(x_t, q_0))\right] = 0; \tag{40}$$

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

$$\tilde{k}(x_t, e^S, q) = \frac{1}{S}\sum_{s=1}^{S} K(g(x_t, e_t^s, q), x_t); \tag{41}$$

where e^S is 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,

$$\hat{q} = \arg\min_{q}\left(\sum_{t=1}^{n} Z_t\left[K(y_t, x_t) - \tilde{k}(x_t, e^S, q)\right]\right)' \Omega \left(\sum_{t=1}^{n} Z_t\left[K(y_t, x_t) - \tilde{k}(x_t, e^S, q)\right]\right) \tag{42}$$

where Ω is a r x r symmetric positive semi-definite matrix. As shown in Gourieroux & 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. Gourieroux & Monfort(1996) p. 32 gives a simulation estimator of the optimal Ω.

Two observations are in place:

- The number of moment conditions (40) invoked must be no less than the number of unknown parameters, otherwise the model becomes unidentified.

- The quadratic expression in (42) 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. MSM which include tax and benefit legislation typically have discontinuities in variables, which may or may not imply discontinuities with respect to behavioral parameters.

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

$$E(y_t - E(g(x_t, e_t, q_0)) = 0; \tag{43}$$

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

$$\bar{y} - \frac{1}{n}\sum_{t=1}^{n} \tilde{k}(x_t, e^S, q_0); \tag{44}$$

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 (44) for $\bar{Y}$ and extend the summation in the second term of (44) to N, and thus get an empirical correspondence to (43) for the whole finite population. In practice this is of course not possible. One only knows the x-observations of the sample, but with know selection probabilities p_t they can be used to compute the following estimate,

$$\bar{Y} - \frac{1}{N}\sum \frac{1}{p_t}\tilde{k}(x_t, e^S, q); \tag{45}$$

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

Let's now conclude this section with a simple illustration of the simulation approach to estimation. Assume a simple regression model in the form of an earnings function,

$$y_t = b' x_t + e_t; \tag{46a}$$

$$e_t \sim IN(0, s^2); \tag{46b}$$

where y_t is the log wage rate, and x_t a vector of three exogenous variables and a unit value for the intercept. The variables are years of labor market experience, its square and years of schooling. X-data are a sub-sample from the HUS surveys (Klevmarken & Olovsson, 1993). The sample size is 474. y-data were computed using x-data, the parameter values given in Table 1 below and random draws ε from a standard normal distribution. y- and x-data were then used to estimate β and σ^2 by the method of simulated moments (MSM) and OLS. For this model the latter estimator is an efficient ML-estimator and it serves as a standard of comparison.

The moment conditions used in the MSM-estimator were,

$$E\left[Z_t(y_t - E(y_t)) \mid Z_t\right] = 0; \tag{47a}$$

$$E\left[Z_t(y_t^2 - E(y_t^2)) \mid Z_t\right] = 0; \tag{47b}$$

and thus,

$$K_t(y_t, x_t) = \begin{Bmatrix} y_t \\ y_t^2 \end{Bmatrix}; \tag{48}$$

Ω was simply set to the identity matrix, which is not an optimal choice. The resulting estimates will thus not be efficient, but consistent. The simulator (41) was computed with S=300 random draws.

The whole experiment was replicated 100 times and the mean and standard deviation of the parameter estimates are displayed in the second and third columns of Table 1. The model was also estimated using the assumption that the mean of the dependent variable for the whole population was known. This additional information was included as a constraint on the estimated parameters. The mean of all predicted log wages in the sample was constrained to coincide with the population mean. As discussed above this approach is only justified if the sample is so large that random deviations between the two means can be neglected. The results from 100 replications of this experiment are displayed in column four of Table 1. (The OLS results in column five were not constrained. They were, however, computed from the same samples as the constrained estimates.)

Table 1. Means and standard deviations of 100 replicated estimates of an earnings function

True	MSM	OLS	MSM-constr.	OLS
	Means of parameter estimates			
1.00	1.000	0.999	0.996	1.005
0.04	0.040	0.040	0.040	0.040
-0.002	-0.002	-0.002	-0.002	-0.002
0.05	0.050	0.050	0.050	0.049
0.06	0.050	0.060	0.048	0.060
	Standard deviation of parameter estimates			
	0.072	0.054	0.064	0.052
	0.005	0.003	0.005	0.003
	13.28e-05	7.08e-05	12.84e-05	7.34e-05
	0.005	0.004	0.004	0.004
	0.049	0.004	0.038	0.004

Note: The first parameter is the intercept and the last parameter is σ^2.

Both methods appear to give unbiased estimates, with the possible exception of the MSM estimates of the residual variance. The second half of the table shows that the efficiency of the MSM estimates is much less than that of the OLS estimates. The difference in efficiency is especially large for the estimate of the residual variance. The precision of the simulation estimates can be improved in at least two ways. One is to use an optimal estimate of Ω, and another is to use a more efficient simulator. An interesting finding is that adding the information about the population mean only increases efficiency marginally. This simple illustration has demonstrated the simplicity and feasibility of the simulated method of moments. In a true application one would of course not use this approach to estimate a simple regression model. But in the more complex micro simulation models it should have a relative advantage.

5. Model validation

If the tax and benefit legislation has been translated into computer code with sufficient detail and care and the data are detailed and accurate enough there is no need to validate a conventional static tax-benefit model without behavioral adjustments, because there is nothing to validate. However, if the simulation model includes behavioral adjustments there is a validation problem. How would one go about validating a static model? Is it at all possible? The problem with the comparative statics of a static micro-simulation model is that it does not give predictions for any specific time point or time interval, and thus, it is hard to know to what the predictions should be compared. Suppose for instance, that a labor force participation equation is estimated from a cross-section at the end of a long period of unchanged tax and benefit systems and a stable labor market. Then a major tax reform takes place. Is it a good idea to validate the predictions from this model by comparing with observed participation rates from the first, second or third, etc year after the reform?

Validation of a dynamic and dated model does not suffer from the same problem. In this case predictions have a correspondence in the real world. Model validation should proceed along two different lines. One is conventional specification testing of each single sub-model in the model building phase, the other is the testing of model simulations from the entire model against external data. That is, data not used in the estimation and simulation of the model. In validating the model one would like to take account of the fact that the simulations are subject to stochastic errors. These errors originate from two sources. One is the stochastic model structure. Invoking random number generators generates events. The other source is the set of parameter estimates. We do not know the true parameters only error prone estimates.

For a model not to big and complex in structure it might be feasible to derive an analytical expression for the variance-covariance matrix of the simulations, which takes both sources into account, 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 uncertainty generated by the model as such can be evaluated if a simulation is 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 the distribution of the estimates can be approximated by a multivariate normal distribution with mean vector and 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 simulations due to the uncertainty about the true parameter values can be obtained.

To avoid the normal approximation one could consider using sample re-use methods. For instance, by jack-knifing or bootstrapping 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 that due to the random nature of the model.

Even if model simulations do not deviate more from new bench marks than is normal given the stochastic properties of the model, one might like to improve the precision of the parameter estimates by updating or calibrating them to this new bench mark information. How this can be done was discussed above. If the simulations deviate significantly from the new bench marks, that is an indication of a misspecified model. In this situation it would seem improper just to calibrate the parameters to the new benchmarks or align the model simulations to them. A re-specification of the model might be necessary.

Finally we should also note that validation need not only be done against benchmarks like means and totals. If frequency distributions or measures of dispersion and correlation are available they could also be used. As noted above a micro simulation model is likely to have a number of simplifying assumptions about independence of variables, which might cause variances and in particular correlations to decrease over time. Validation against observed correlations and dispersions will then prove useful.

Much of the total error in simulated values will come from the choice of a particular model structure or specification. Sensitivity analysis is an approach to assess the importance of this source of error. As pointed out in Citro & Hanushek(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.

5. Concluding remarks

The credibility of micro simulation models with the research community as well as with users will in the long run depend on the methods used for estimation, testing and validation. It is now time to go from the infancy stage of accepting and using a model if it just runs and produces simulations which are not too far off, and instead take the inference problems seriously. Some of the procedures, which are now used in the field, can probably be understood and defended in the light of inference theory, while others should be abandoned.

This paper has attempted to put some structure to a discussion of inference in MSM. The model structure will determine both how data could be collected and which methods are appropriate for estimation. In particular a piece meal approach is dependent on a hierarchical or recursive structure of the model.

It was also suggested that the alignment procedures now used in practical work could be seen as part of the estimation procedure. It is, however, important to test if the constraints imposed by alignment are accepted by data. If they are not that is a clear indication that something is wrong with the model, and it should be reformulated rather than forced "on track" by alignment.

Finally it was also suggested that the simulation approach to estimation could be useful in micro simulation work. These models are designed to simulate and they also frequently include nonlinear and complex relations, which suggests that simulation-based estimation has a relative advantage.

The most important result is perhaps that path-calibrated estimates in general are inconsistent and thus should be avoided. A better alternative is moment-calibrated estimates.

References

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.

Deville, J.C. and Särndal, C.E. (1992), "Calibration estimators in survey sampling", *Journal of American Statistical Association* 87, 376-382

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

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

Klevmarken, N.A. and P. Olovsson (1993), *Household Market and Nonmarket Activities (HUS). Procedures and Codes*, IUI/Almquist & Wiksell International, Stockholm

Lindström, K. (1997), "Utvärdering av framskrivningstekniken i FASIT-modellen" (Evaluation of the calibration technique used in the FASIT model), memo Statistics Sweden

Lunström, S. (1997) *Calibration as a standard method for treatment of nonresponse*, Department of Statistics, Stockholm University (diss.)

Merz, J. (1993) "ADJUST – A program package for adjustment of micro data by minimum information loss principle. Program Manual" FFB-Documentation No 1, University of Luneburg

Merz, J., (1994), "Microdata adjustment using the minimum information loss principle", Discussion Paper no 10, Forschungsinstitut Freie Berufe, Universität Luneburg

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