MICROECONOMETRICS

*by David Brownstone**

1. The Microeconometric Problem

Microeconometrics is the application of econometric techniques to microeconomic data at the household or firm level. Of course this is not really a new field since economists and econometricians have been using micro data for decades. However, in the last decade there has been a marked shift in the research of many econometricians towards techniques and applications requiring micro data. The main reason for this shift has been the inability of macroeconometric models developed in the 60s to satisfactorily predict economic events in the last decade. The use of micro data has also been facilitated by the increasing availability of micro data sources and the decreasing cost of computers, large enough to effectively handle the large data sets necessary for empirical microeconometric work.

The growth of microeconometrics has also been caused by the large increase in government policies designed to directly affect microeconomic decisions and the distribution of wealth and income. In Sweden good examples of these policies are found in labor market policies and regional investment policies. Any model for evaluating these policies must be disaggregated at least to the smallest level treated differently by the policy. Since many policies interact with the tax and transfer systems this smallest level is frequently the individual or firm. Although most microeconometric work has been directed at detailed analysis of specific policies it is theoretically possible to generate aggregate predictions from these models. Since the aggregate data used to calibrate macroeconometric models are just summaries of the underlying micro data sources, it should be possible to generate improved macro predictions using microeconometric models. This is the basic idea behind the microsimulation modeling pioneered by Orcutt and implemented in the MOSES model (see Eliasson, 1980) for Sweden. Realization of these improved aggregate predictions requires much more research on large model specification as well as lower computing costs.

One frequent criticism of macroeconometric models is that they are based on ad hoc specifications and are generally inconsistent with any coherent theory of economic behavior. The primary difficulty, for those attempting to remedy this problem, is that the strong assumptions needed, to generate

---

*IUI and Stockholm School of Economics. The author acknowledges useful comments and references provided by Eliasson, G., and Klevmarken, A., but retains sole responsibility for remaining errors. The references given in this paper are to the most recent general papers on a given topic. These papers all contain complete historical references.

aggregate relationships consistent with standard economic theory, are clearly incorrect and lead to very restrictive aggregate specifications. If macroeconometric models are just going to be used for generating aggregate predictions, then consistency with any theory is of course not required. However, the practical impossibility of relating these models to theoretical constructs has served to alienate theorists and therefore increase the gulf between economic theory and practice.

In microeconometric models, the basic model specification assumptions are made at the individual level. For example, the specification of some models of consumer behavior begins with an explicit representation of the individual's behavioral assumptions into testable assumptions about the empirical model parameters. Although this procedure has not always led to better predictive performance, it does allow for fruitful cooperation between theorists and applied workers. This cooperation is evident in the current work on labor supply where empirical models are being generated from theoretical models of workers' job search behavior. The empirical results have been useful for identifying implausible assumptions as well as generating quantitative predictions.

Microeconometric models, which have been used to simulate the macroeconomy (like IUI's MOSES model), have also been criticized for making large numbers of ad hoc assumptions. These assumptions are usually necessary to identify parts of the model where no calibration data exist. One way of quantifying the impact of a priori assumptions is to adopt a Bayesian framework where the assumptions are incorporated into the prior distribution. The importance of these assumptions for a particular set of calibration data can then be measured as the contribution of the prior distribution to the posterior distribution calculated from the data. When viewed in this framework macroeconometric models clearly impose a large number of assumptions in addition to those discussed in econometrics textbooks. In order to insure that these models are structurally stable it is necessary to assume that a large number of aggregation conditions are satisfied. For the consumer sector these conditions generally require all consumers to have identical utility functions. When compared to these assumptions the admittedly ad hoc assumptions used in microeconometric simulation models do not seem so bad.

In spite of the claims made in the previous paragraph it is still true that microeconometricians have been guilty of making too many ad hoc assumptions. In a recent paper Heckman and Singer have shown that the results from microeconometric models of labor force dynamics are very sensitive to certain distributional assumptions used in these models. They also show that these assumptions are not necessary to identify these models and suggest new estimation techniques which do not require strong distributional assumptions.

2. Distinguishing Features

The development of microeconometric techniques during the last decade has been heavily influenced by two features of most micro data sources: large sample size and the presence of discrete and qualitative data. Most applied microeconometric studies use samples with 500 to 5 000 observations. While these sample sizes are much smaller than those used in demography and many physical sciences, they are much larger than those used in macroeconometric work. These large sample sizes allow the application of asymptotic statistical theory to justify the efficient estimation of non-linear models by likelihood maximization techniques. The large sample sizes also allow the use of robust statistical estimation procedures. These robust procedures do not require specific assumptions about the distribution of the error terms. It has been shown that (see Basset and Koenker, 1982) incorrect distributional assumptions can cause serious biases in the parameter estimates. Unfortunately, good robust estimation procedures are only available for the standard linear model. More work is needed to develop usable robust procedures for the more complicated models used in microeconometric studies.

The presence of discrete and qualitative data in micro data sources has motivated a large amount of research on appropriate models and estimation techniques. (For a good description of this work, see Manski and McFadden, 1981.) In many microeconometric studies the endogenous variables to be predicted are qualitative. Examples of these variables include the decision to enter (or exit) the labor force and the choice of transportation mode for commuting to work. During the early 70s McFadden and others developed a new class of econometric models, called qualitative choice models, to handle these discrete endogenous variables. These models are extensions of models originally developed in biometrics by Thurstone (1927) and others. Qualitative choice models specify the probability that the individual will choose a particular discrete alternative given observations on the relevant exogenous factors. This probability, called the choice probability, is a continuous function of the exogenous variables and unknown parameters, so the econometric techniques previously developed for continuous models can be directly applied.

Although these choice probabilities are continuous they must lie between 0 and 1, and the sum of the choice probabilities over all the discrete alternatives must equal one for each individual. These restrictions do not allow the use of simple estimators like linear regression techniques. Therefore qualitative choice models must be estimated using non-linear, iterative computer algorithms. Reliable and relatively fast algorithms have been developed for simple qualitative choice models, but more work is needed to develop usable algorithms for more complex and realistic models. Until the next generation of "supercomputers" become available, the computational costs of using qualitative choice models will be much higher than those associated with other econometric models.

One common method for specifying a qualitative choice model is to assume that the utility derived from choosing discrete alternative i is given by:

$$V_i = U_i (Z_i, B) + e_i$$

where U is the representative or observable utility function of the exogenous attributes, Z, and parameters B. The "error term", e, represents the unmeasured utility. With these assumptions, the choice probability is given by:

$$P_i = \text{Probability that } V_i > V_j \text{ for all possible j}$$

Given a specific distributional assumption about how the e's vary across the sample it is possible to estimate the parameters, B, and therefore estimate the representative utility function, U. A qualitative choice model for firm behavior (i.e. technology choice) can be developed in a similar fashion where U and V are production or cost functions. Any qualitative choice model derived in this fashion is consistent with standard microeconomic theory. This fact allows the use of general welfare measures derived from the estimated parameter values (see Rosen and Small, 1981). Also many common simplifying assumptions about the properties of utility (or cost) functions can be translated into hypotheses about the parameters B. Testing these hypotheses using empirical estimates of B provides a simple way of testing theorists' assumptions.

If it were possible to directly observe the utilities V in the previous paragraph then the unknown parameters B could be estimated using standard econometric techniques for continuous endogenous variables. The difficulties here are caused by the fact that the V's are not directly observed; only indicators of their magnitudes are observed. Using this interpretation of qualitative choice models, Heckman (1978) has proposed a general class of econometric models which includes the standard simultaneous equation system as well as qualitative choice models. Heckman's system also includes other models, where the endogenous variables are only partially observed (truncated or Tobit models), as well as models where some variables are not observed at all (latent variable models).

These general models have been used where there are related discrete and continuous endogenous variables. These models arise when qualitative and continuous choices are derived from demand for an underlying good. Examples include the decision to purchase a car and the number of miles to drive it, and the firm's decision to choose a particular discrete technology for a plant and how much to produce in the plant. In the first example the underlying good is transportation services, and in the second it is profit. Microeconomic theory places many restrictions on the parameters of the joint discrete-continuous choice model. These restrictions can either be tested empirically or used to greatly reduce the number of parameters which must be estimated (see Brownstone, 1980, and Duncan, 1980, for examples). These models are also potentially useful for policy analysis. For example, it would be possible to consistently estimate the effects of a gasoline price increase on the decision to purchase a car as well as miles driven. Furthermore, it would be possible to examine these effects for different groups (i.e. rich or poor) in the sample.

Once a microeconometric model has been specified and estimated, there is still the problem of how to use it for policy analysis. Most applied microeconometric work has used steady-state equilibrium models estimated with cross-section data. Policy analysis for these models is therefore limited to comparing equilibrium values across hypothetical changes in the exogenous variables. Since most of these models are non-linear due to the presence of qualitative choices it is not possible to analytically compute equilibrium values for these models. The most popular solution to this problem is to numerically simulate the effects of the exogenous changes for each individual in a random sample (which is usually identical to the estimation sample) and then sum up the individual effects to estimate the overall effect. It is easy to modify these techniques to account for the stratification used in most existing surveys.

The problems of policy analysis and prediction are much more difficult for *dynamic* microeconometric models. These models must be calibrated using panel (or longitudinal) data sets which generally contain information for a relatively large number of individuals over a small number of time periods. Most reasonable dynamic specifications depend on the initial values of the exogenous variables at the beginning of the process. Due to short time coverage, these initial conditions are rarely observed in current panel data, so they must be treated as unobserved latent variables. If assumptions are made about the distribution of these initial values over the population then it is generally possible to estimate the other parameters in the model, but frequently the resulting estimates (and therefore predictions) are very sensitive to the choice of distribution for the initial values. This problem has been noticed by researchers at IUI's MOSES project. They have discovered that the predictions from MOSES are very sensitive to the initial values chosen. The only solution to this problem is to estimate (or otherwise specify) a specific distribution for the initial values and then numerically calculate expected predictions with respect to this distribution. Heckman (1981) has clearly elucidated the identification problems caused by unobserved initial conditions. MacCurdy (1982) and Chamberlain (1982) have also proposed promising techniques for estimating the distribution of initial values without imposing highly restrictive assumptions. In addition to the problems with initial conditions, Klevmarken (1980) has shown that there are also problems with the dynamic simulation algorithms used in existing microsimulation models.

3. Survey Data Problems

Since most microeconomic data sets are collected using survey interviews of households or firms, microeconometricians have been forced to develop techniques for dealing with the shortcomings of survey data. The largest and most common problem is missing data. In any survey there are always people who refuse to answer any questions or who give obviously incorrect answers. It is well known that if these non-respondents are different from the respondents, then predictions based on the survey may be incorrect (this effect is called sample selection bias). Therefore, before using any survey the analyst must first test to see if there is sample selection bias, and if so use techniques to remove the bias. These problems are not unique to microeconometrics (see the survey by Little, 1982), but due to the sensitive nature of the questions economists like to ask, non-response is frequently very high.

It is useful to note that the decision to answer an interviewer's questions can be modeled using qualitative choice models discussed earlier. This approach to sample selection problems, called the model-based approach, provides a general framework for testing and correcting for sample selection bias. It also provides a framework for using auxiliary information about non-respondents (like census records) to improve the accuracy of these model-based techniques.

Another problem with economic survey data is the possibility of lying by respondents. This is particularly likely if the respondents are cheating on their taxes and they think that the tax authorities will have access to the survey data. One approach to this problem is to develop qualitative choice models for the event that the respondent is lying, and then use this model to remove likely cheaters from the sample as well as correct for the resulting sample selection bias. The difficulty with this approach is that it is very difficult to obtain data for calibrating the cheating model. It may be possible to ask the same question in different forms in the course of a long interview. Hopefully the cheaters will forget the original incorrect figure they gave for the first question and give a different (probably also incorrect) figure for the second question. The other possibility is to link the survey data to tax records and look for inconsistencies.

Even if survey respondents do not intentionally give misleading or incorrect answers it is frequently not clear that they could give accurate answers even if they wanted to. One common example of this problem occurs when surveying firms to determine their cost or production functions.

Standard accounting practices do not produce figures which correspond to economists' notions of cost or production functions. Therefore it may be very difficult for the firm to give an accurate estimate of these quantities. Furthermore, if the firm undertakes the effort to calculate its marginal costs, then this new information might change the firm's actions. There is no general solution to these problems, but if the firm does not have estimates of its cost and demand functions then it is hard to understand how they are conforming to the neoclassical profit-maximizing assumptions.

The actual collection of sample survey data is by far the most expensive part of applied microeconometric work. Since economists are interested in policies which only affect a small number of people (like unemployment benefits) it is frequently necessary to use a stratified sampling procedure to insure that enough people are sampled in each category. Unfortunately this stratification is frequently based on variables which are endogenous to the models being considered. It is well known that this type of endogenous stratification makes it impossible to use estimation techniques developed for random sampling schemes. Recent work by Cosslett (1981) has shown that it is possible to derive consistent estimates from these endogenously stratified samples by making some very simple changes to the random sample estimation procedures. In some contexts, Cosslett's results can be used to design more cost effective sampling plans. For example, if one is interested in commuters' choice of transportation mode then it is much easier to take separate samples from each mode than to take a random sample of the population large enough to insure sufficient observations in each mode. These types of sampling schemes are also frequently used in epidemiology.

Microeconometricians are concerned with a number of other survey problems. Techniques for measuring time use and consumer expenditures are one area of active research (see Klevmarken's paper in this volume). One difficult problem is the measurement of small but infrequent purchases by consumers. Deaton and Irish (1982) have recently proposed a model which accounts for possible under-reporting of these expenditures. The main feature of their model is a qualitative choice model to predict whether the respondents report all of their purchases.

4. Microeconometric Research at IUI

IUI has been actively engaged in microeconomic data collection and analysis for some time. Current projects utilizing advanced microeconometric techniques include the labor supply work by A. Björklund and B. Holmlund and the HUS project (see page 104 in this volume). During the next year the author and A. Klevmarken will be doing methodological research in a number of different areas for the HUS project. Although this research is aimed at developing techniques for analyzing the HUS project data, these techniques will be useful for other researchers at IUI and elsewhere. IUI's MOSES model is also an example of a dynamic microeconometric model. Because of the disparate data sources used to calibrate MOSES, the techniques discussed in this article cannot be directly applied. However, the general principles and techniques discussed here may be useful for improving the calibration and predictive accuracy of the MOSES model.

The HUS project pilot study shows that one of the largest problems with the HUS sample is going to be high non-response and the possibility of sample selection bias. Current techniques for dealing with these problems require strong assumptions about the household's decision to participate in the survey. In most applications these assumptions cannot be tested since there is generally no data available for non-respondents. The HUS project is fortunate to have HINK register data for all members of the sample, including non-respondents. These HINK data include tax and basic census information for all years beginning in 1978. Certainly these data can be used to test the modeling assumptions made by current sample selection models. In addition it should be possible to develop new methods which use the supplementary information to improve the accuracy of the bias corrections. If these new techniques are successful, then these results could be used by other researchers, to justify obtaining similar supplementary information for other surveys where sample selection bias is a problem.

The author is also working on developing robust estimation techniques for qualitative choice models. Current qualitative choice models require strong assumptions about the distribution of the "error terms" across the sample. These assumptions are very difficult to test and in some cases the resulting estimates appear to be quite sensitive to the specific assumptions made. A related problem is the high computation costs associated with estimating more realistic qualitative choice models. The author is one of the main developers of the QUAIL computer package, which is the most popular computer package for estimating qualitative choice models. He has recently completed a small study with K. Small (1982) showing that it is computationally feasible to efficiently estimate moderate sized Nested Logit models. These models are important members of a group of qualitative choice models proposed by McFadden (see Manski and McFadden, 1981) which require less restrictive distributional assumptions than current popular models. Other researchers are working on these problems in the U.S. and England, and it is likely that any new results could be directly applied to analysis of HUS project data and other microeconometric projects at IUI.

REFERENCES

Basset, G.J., and Koenker, R., 1982, "Tests of Hypotheses and 1–1 Estimation", Paper delivered at the 1982 European Meetings of the Econometric Society in Dublin, Ireland.

Brownstone, D., 1980, "A Joint Discrete/Continuous Choice Model for Consumer Durables", Unpublished Ph.d. dissertation, Dept. of Economics, University of California, Berkely.

Chamberlain, G., 1982, "Panel Data", Working Paper No. 8209, Social Science Research Institute, University of Wisconsin, Madison.

Cosslett, S. R., 1981, "Efficient Estimation of Discrete-Choice Models" in Manski, C.F., and McFadden, D., (eds.), 1981, *Structural Analysis of Discrete Data with Econometric Applications.* MIT Press, Cambridge.

Deaton, A., and Irish, M., 1982, "A Statistical Model for Zero Expenditures in Household Budget Surveys", Paper delivered at the 1982 European Meetings of the Econometric Society in Dublin, Ireland.

Duncan, D., 1980, "Formulation and Estimation of the Mixed Continuous/Discrete Dependent Variable Model in Classical Production Theory", *Econometrica* 48, pp. 839–852.

Eliasson, G., 1980, "MOSES: A Micro Simulation Model of a National Economy" in Bergmann, B., Eliasson, G., and Orcutt, G., (eds.), *Micro-Simulation: Models, Methods and Applications.* IUI Conference Reports 1980:1.

Heckman, J., 1978, "Dummy Endogenous Variables in a Simultaneous Equation System", *Econometrica* 46, pp. 931–959.

Heckman, J., 1981, "The Incidental Parameters Problem and the Problem of Initial Conditions", in Manski, C.F., and McFadden, D., (eds.), 1981, *Structural Analysis of Discrete Data with Econometric Applications.* MIT Press, Cambridge.

Heckman, J., and Singer, B., 1982, "The Identification Problem in Econometric Models for Duration Data", Discussion Paper 82–6, Economics Research Center/NORC, Chicago.

Klevmarken, A., 1980, "On Estimation and Other Problems of Statistical Inference in the Micro-Simulation Approach", in Bergmann, B., Eliasson, G., and Orcutt, G., (eds.) *Micro-Simulation: Models, Methods and Applications.* IUI Conference Reports 1980:1.

Little, R.J.A., 1982, "Models for Non-Response in Sample Surveys", *Journal of the American Statistical Association* 77, pp. 237–250.

MacCurdy, T.E., 1982, "The Use of Time Series Processes to Model the Error Structure of Earnings in a Longitudinal Data Analysis", *Journal of Econometrics* 18, pp. 83–114.

Manski, C.F., and McFadden, D., (eds.), 1981, *Structural Analysis of Discrete Data with Econometric Applications.* MIT Press, Cambridge.

Orcutt, G., 1980, "Hypothesis Formation, Testing and Estimation for Microanalytic Modeling" in Bergmann, B., Eliasson, G., and Orcutt, G., (eds.) *Micro-Simulation: Models, Methods and Applications.* IUI Conference Reports 1980:1.

Rosen, H., and Small, K., 1981, "Applied Welfare Economics with Discrete Choice Models", *Econometrica* 49, pp. 105–130.

Small, K., and Brownstone, D., 1982, "Efficient Estimation of Nested Logit Models: An Application to Trip Timing", Econometric Research Program, Research Memorandum 296, Princeton University, Princeton, New Jersey.

Thurstone, L., 1927, "A Law of Comparative Judgement", *Psychological Review* 34, pp. 273–286.