Reihe »Wirtschaftswissenschaft«, Band 11

Herausgeber und Autoren der Reihe wollen mit anwendungsorientierten Arbeiten, die theoretische Analyse *und* empirische Prüfung verbinden, beitragen zur Lösung der wachsenden Probleme, vor denen die modernen Industrienationen stehen.

Editors: *Dr. Dr. Johann Kurt Brunner* is Lecturer at the Institute of Economics, University of Linz.
Dr. Hans-Georg Petersen is Professor of Economics at the University of Gießen, and author of 9 books and numerous articles.

Johann Kurt Brunner, Hans-Georg Petersen
(Editors)

Simulation Models in Tax and Transfer Policy

Proceedings of an International Symposium

Campus Verlag
Frankfurt/New York

CHRISTIAN, Ch. and P. FRISCHMANN, 1988, Survivorship Bias in the Statistics of Income Individual Panel Tax Data. Unpublished manuscript. Arizona State University.
CLOTFELTER, Ch., 1980, Tax Incentives and Charitable Giving: Evidence from a Panel of Taxpayers. Journal of Public Economics, 319 - 340.
DANIEL, J., 1988, The Price and Income Elasticities of Charitable Giving: New Evidence from a Panel of Taxpayers. Unpublished manuscript. University of Minnesota.
FELDSTEIN, M., J. SLEMROD, and S. YITZHAKI, 1980, The Effects of Taxation on the Selling of Corporate Stock and the Realization of Capital Gains. Quarterly Journal of Economics 4, 771 - 791.
FELDSTEIN, M., J. SLEMROD, and S. YITZHAKI, 1984, The Effects of Taxation on the Selling of Corporate Stock and the Realization of Capital Gains: Reply. Quarterly Journal of Economics, 111 - 117.
HSAIO, Ch., 1983, Analysis of Panel Data (Cambridge).
MINARIK, J., 1981, Capital Gains, in: H.J. AARON and J.A. PECHMAN, eds., How Taxes Affect Economic Behavior (Washington, D.C.), 241 - 277.
MINARIK, J., 1984, The Effects of Taxation on the Selling of Corporate Stock and the Realization of Capital Gains: Comment. Quarterly Journal of Economics, 93 - 110.
U.S. DEPARTMENT OF THE TREASURY, 1985, Capital Gains Tax Reductions of 1978. Report to Congress.

Modeling Direct and Indirect Impacts of Tax and Transfer Programs on Household Behavior

ALICE NAKAMURA AND MASAO NAKAMURA

I. Introduction

From its inception microsimulation has been viewed as a promising methodology for studying the effects of proposed tax and transfer programs. The reasons for this are both substantive and methodological. The substantive reasons are rooted in the nature of tax and transfer programs and the related policy concerns. The rules for these programs are usually specified in terms of individual and family characteristics such as the level of earnings, number of dependent children, age and disability status. Yet the effects of tax and transfer programs that are of interest from a policy perspective are mostly macro in nature. For instance, there is often interest in the total revenues and costs of a program, and in general distributional effects. Aggregation problems are the key methodological challenge arising from this substantive context (see, for instance, ORCUTT, GREENBERGER, KORBEL and RIVLIN, 1961; ORCUTT, MERZ and QUINKE, 1986; PECHMAN, 1965 a and 1965 b; and HAVEMAN, HOLLENBECK, BETSON and HOLMER, 1980).

Economists have traditionally treated aggregation as the problem of aggregating relationships. For instance, economic theory might suggest a particular functional form, certain explanatory variables, and qualitative parameter restrictions for a relationship describing the consumption behavior of an individual. Provided that this relationship is simple enough, the corresponding macro relationship could be derived mathematically. The macro relationship could be estimated using aggregate data and used as the basis of macro policy analysis.

Until recently the only data available in many areas of economic inquiry were macro data. This was a compelling reason for pursuing aggregation through aggregation of micro relationships. The approach also has a number of well-known drawbacks. Aggregation of nonlinear relationships is mathematically challenging. Econometric problems of multicollinearity, autocorrelation and simultaneous feedbacks are often severe in a macro time series data context. Moreover macro empirical relationships can support only a narrow range of macro policy investigations confined to the level of aggregation for the macro relationship. Thus if the level of aggregation is the nation, it will not be possible to consider state aggregates or the aggregate behavior of particular demographic groups such as adults living alone or families with young children. Nor is it possible to consider distributional effects.

Microsimulation is an alternative approach to aggregation. The core of a household microsimulation model is a population of individuals which constitutes a sample representation of some actual population. Behavioral relationships specified and estimated at the micro level can be applied sequentially to the individuals in the simulation population. The predicted individual values of the relevant dependent variable(s), the **micro outputs**, can then be aggregated in whatever ways are desired for policy analysis. (See ORCUTT, 1968; ORCUTT, WATTS and EDWARDS, 1968; and EDWARDS and ORCUTT, 1969 for further discussions on the microsimulation approach to aggregation.) The drawbacks to the microsimulation approach include computational complexity and expense. In addition, there is often a desire to ascertain that economic and demographic quantities generated by a microsimulation model are consistant with (published) macro or aggregate quantities. This problem has not been solved in general. (See, for example, MERZ, 1986 and NAKAMURA and NAKAMURA, 1978 and 1985 d.) Nevertheless interest in the microsimulation approach for analyzing the effects of tax and transfer programs has continued to grow, as evidenced by the Giessen workshop and the range of papers being presented in this workshop.

In Section II we begin by considering the range of tax and transfer programs that might be incorporated into a microsimulation model. The potential impacts of various types of tax and transfer programs is the subject of Section III. Consideration is also given to the potential policy importance of various types of tax and transfer program effects. Implementation problems are the subject of Section IV, with a distinction being drawn between the problems of implementing tax and transfer programs in static versus dynamic microsimulation models. The exercise of building a microsimulation model brings into sharp focus the problem of choosing among competing behavioral models as represented by alternative sets of estimated relationships. In Section V we briefly review accepted model selection strategies, and indicate how output space analyses carried out using microsimulation may be an important addition to this list. Summary observations and conclusions are presented in Section VI.

II. Types of Tax and Transfer-Programs

Tax and transfer programs can be grouped into those that directly alter the earned or unearned disposable incomes of individuals and families, those which affect buying

power by changing the prices of purchased goods and services, and transfers in-kind. Each of these three types of programs can potentially be incorporated in a microsimulation framework.

Work behavior in a year (worked or did not work), annual hours of work, the average hourly wage rate, and annual earnings are among the individual characteristics which are updated in many household microsimulation models. Thus tax and transfer programs which alter disposable earned income can be, and frequently are, implemented in microsimulation models (see, for example, ORCUTT, CALDWELL and WERTHEIMER, 1976, Chs. 9 and 10). It is only necessary that other important personal characteristics on which treatment under the program depends are also carried as characteristics in the microsimulation model.

In this regard, an important distinction can be drawn between current versus cumulative or historical classifying characteristics. Age and earned income in the current year are examples of current characteristics. The labor supply and earnings histories required to calculate Social Security benefit levels are examples of cumulative classifying characteristics, since these histories are cumulated over the working lifetimes of individuals. In the same spirit, the number of children an individual has might be treated as a cumulative classifying characteristic. However, unlike an individual's labor supply or earnings history, the number of children an individual has is typically available as a current charactertistic in most cross-sectional and longitudinal micro data sets.

Tax and transfer programs affecting disposable income from sources other than current labor supply are more difficult to incorporate. This is because incorporation of these programs requires that the appropriate components of nonlabor income be available in the microsimulation model as individual characteristics. There is little agreement on how to model nonlabor income flows.

Programs that affect the buying power for all individuals in the same manner, such as a nation-wide sales or excise tax, can be incorporated (together, with an appropriate deflator) in determining the year-to-year real disposable income levels for individuals. However, many of the tax and transfer programs that affect price levels are state-specific, and information on the locations of individuals is not carried in most microsimulation models. Incorporation of state-specific or region-specific programs is further complicated by the difficulty of obtaining deflators which appropriately reflect locational price differences due to factors other than tax and transfer programs.

Nor is it common for programs involving transfers in-kind to be incorporated in microsimulation models. Incorporation of programs of this sort is feasible. Ideally, two conditions must be met. First, the personal characteristics used in administratively determining the in-kind transfer levels must be carried as attributes of the individuals in the simulation population. Second, there must be some basis for determining how much individuals would spend on the goods or services being transferred in the absence of the transfer program. This way there is some basis for determining both increases in the standard of living due to program-induced increases in the levels of consumption for the transferred goods or services, and increases in the standard of living due to increases in the income available to purchase other goods and services. To date, however, when transfers in-kind (such as food stamps or Medicaid) have been incorporated in microsimulation models, the second of these conditions has been ignored (see ORCUTT, CALDWELL and WERTHEIMER, 1976, pp. 207 - 208 and WERTHEIMER, ZEDLEWSKI, ANDERSON and MOORE, 1986, for example). This brings us face to face with the issue that there are different degrees of complexity in the extent to which a tax or transfer program may be integrated within a microsimulation model. The simplest way is just to incorporate the program rules. Our discussion so far has implicitly taken this as the standard. In this case, the program effects the microsimulation model will reflect are the **direct effects** assuming complete compliance with and administration of the program exactly as it is drafted and assuming no behavioral changes in response to the program. Thus, for example, changes in consumer expenditure due to direct program effects on disposable income can be captured, but any changes in expenditure due to program-induced changes in labor supply (and hence indirectly in earnings) will be missed. Casual observation and economic and other social science theories suggest there may be a number of important indirect effects of tax and transfer programs. In the next section we review some of the main categories of hypothesized indirect effects.

III. Possible Indirect Effects of Tax and Transfer Programs

We will discuss hypothesized indirect effects under the subheadings of second order effects, higher order personal effects, and macro feedback effects.

3.1. Second Order Effects

Economic theory suggests there will probably be changes in labor supply in response to tax or transfer plan induced changes in disposable income. These are what we will refer to as second order effects. In increasing order of complexity, we will briefly discuss three different varieties of second order effects.

3.1.1. One-Period Income And Substitution Effects

Economists view the potential effects of tax and transfer programs on labor supply in terms of income and substitution effects. Conceptually, income effects are due to changes in the level of income holding the wage rate (viewed as the price of leisure, or the opportunity cost of time spent working) fixed. If leisure is what economists term a normal good, then individuals will want to consume more leisure (and hence work fewer hours) as their income levels rise. Hence the income effects on labor supply are hypothesized to be negative. Some transfer programs, and certain types of taxes such as inheritance taxes, essentially alter the after-tax income levels of individuals without affecting their after-tax wage rates.

Income-compensated substitution effects are defined to be the labor supply responses to changes in the relative price of leisure with income levels being hypothetically adjusted so that utility is unchanged. These compensated substitution effects are hypothesized to be positive. It is clear that all tax and transfer programs which affect disposable income levels through after-tax wage changes should have both income and substitution effects. The net impacts of the income and compensated substitution effects are sometimes referred to as uncompensated or gross substitution effects.

The signs and magnitudes of uncompensated substitution effects due to after-tax wage changes cannot be predicted theoretically; rather the nature of these effects must be explored empirically. The consensus which has emerged for prime aged men is that these uncompensated effects are small and possibly negative. Thus prime aged men would be expected to work slightly more (less) if their earnings were taxed more heavily (lightly). A number of researchers have reported estimates of uncompensated substitution effects for married women that are positive and large in magnitude. If this were the case, the labor supply of married women might be very sensitive to changes in tax and transfer programs. However, recent empirical studies have yielded results for married women which are similar to those reported for men (see NAKAMURA, NAKAMURA and CULLEN, 1979; NAKAMURA and NAKAMURA, 1981, 1983, 1985 a and 1985 b; ROBINSON and TOMES, 1985; SMITH and STELCNER, 1988; and STELCNER and BRESLAW, 1985). If this is the case, one-period income and substitution effects of tax and transfer programs may be of limited interest from a policy perspective.

It should be borne in mind that empirical studies of income and substitution effects have measured these in terms of impacts on hours of work for working individuals. Yet the most dramatic changes in actual labor supply have been associated with changes in employment rates rather than hours of work. It is difficult to relate changes in the propensity to work to wage changes since we do not usually have data on the wage offers received by those who do not work. A second weakness of empirical research on this topic has to do with the way in which wage rates are incorporated into the estimated models. Wage rates are incorporated so that the magnitude of the per unit wage effect is constrained to be the same regardless of the direction, magnitude or reason for the observed wage changes. Given that volitional changes in labor supply are often possible only with job changes, and hence with unobservable search and transactions costs, it seems reasonable to expect that there would be threshold response levels. The failure to find evidence of quantitatively important wage-related changes in labor supply may be due to these and other difficulties.

3.1.2. One-Period Cross Substitution Effects

When only own-wage substitution effects on labor supply are considered, this is equivalent to ignoring the family contexts in which individuals live, except perhaps to the extent that the earnings of other family members (such as the husband) are treated as part of "other income" when modeling labor supply behavior (say, of wives). Theoretically, if the wage rate of another employed family member changes, there should be an income compensated cross-substitution effect as well as an income effect on the designated individual's labor supply. This cross-substitution effect is hypothesized to be positive or negative depending on whether the family members in question view their leisure as complements or as substitutes, respectively. Regardless of the sign, theory implies that the cross-substitution effects for each of the employed members of a family must be equal (see ASHENFELTER and HECKMAN, 1974, for example).

Empirical estimates of cross-substitution effects are fairly rare. In most of these empirical studies, the theoretically implied equality of the cross-substitution effects for employed husbands and wives is rejected. The small number of published estimates of cross-substitution effects for husbands are mostly insignificantly different from zero, while most of the estimates for wives are significantly positive. If this is an adequate reflection of reality, one-period cross substitution effects of tax and transfer programs are of little policy interest for prime-age men but may be important for married women.

3.1.3. Inter-Temporal Effects

All of the tax and transfer effects considered so far can be represented within the context of static, one-period models of labor supply behavior. As PENCAVEL (1986, pp. 44 - 45) explains, however:

> An important development in research in labor supply over the past ten years has been the specification and estimation of life-cycle, multi-period, models according to which consumption and labor supply decisions in each period are made with regard to prices and wage rates in all periods. Utility is defined over lifetime consumption and lifetime hours of work and similarly the budget constraint incorporates incomes and expenditures in different periods plus the opportunity to reallocate incomes and expenditures across periods by borrowing and lending.

Within the context of a life-cycle model, the introduction of tax or transfer programs affecting individuals' disposable wage rates and incomes in one period could hypothetically affect their labor supply in the future as well as in the current period.

Empirical attempts to estimate life-cycle labor supply models fall into two broad categories. In the first category are models which can be viewed as extentions of habit persistence and stock adjustment models. Models of this sort have been estimated using individual panel data by HOTZ, KYDLAND and SEDLACEK (1988) and JOHNSON and PENCAVEL (1984). Studies in the second category are based on what are alternatively called lambda constant or FRISCH demand and supply equations (see HECKMAN 1974 and 1976. See also MACURDY, 1981 and 1983; and BROWNING, DEATON and IRISH, 1985). The lambda in these models is the Lagrange multiplier attached to the lifetime budget constraint, and is interpreted at optimality as the marginal utility of initial wealth. (For further explanation concerning these two approaches, see PENCAVEL, 1986, pp. 44 - 51.)

PENCAVEL (1986, pp. 93 - 94) summarizes what he views as the outcome of empirical research based on life-cycle labor supply models:

> Empirical research at the microeconomic level on male life-cycle labor supply is barely a few years old so it is premature to offer a confident evaluation of its performance. Some provisional judgments can be made, however ... At this stage of the research, the focus of the life-cycle research has been upon the labor supply responses to evolutionary movements in wages. The evidence to date indicates that these labor supply responses for prime-age men are very inelastic with respect to life-cycle changes in wages. Similarly, across male workers, the labor supply responses to differences in entire wage profiles appear to be small ...

If PENCAVEL's tentative assessment of life-cycle labor supply effects is correct, the inter-temporal effects of tax and transfer programs may not be an important policy concern. It seems entirely possible, however, that the models and estimation techniques used to date have simply failed to capture existing intertemporal substitution effects.

3.2. Higher Order Personal Effects

Limited empirical evidence and evolving bodies of theory in both economics and sociology suggest tax and transfer programs may have substantial impacts on aspects of behavior such as family formation and dissolution, fertility, child rearing practices, and the accumulation of human capital. Also there may be alterations in buying and saving behavior which go beyond adjustments to program-induced changes in labor supply, and hence in earned income. Effects such as these cannot be captured in a microsimulation model unless it incorporates relationships for the behavior which is thought to be affected, and these relationships in turn incorporate program-affected variables such as income net of tax and transfer payments. Limited attempts have been made to do this in the areas of marriage, divorce, fertility and savings behavior.

One of the main obstacles to proceeding further in this direction is the lack of convincing empirical evidence on which to base the required behavioral relationships. These higher order effects need to be prioritized, and research needs to be directed toward improving our understanding of the more important of these effects. Certainly some of the most pressing social and political controversies concerning proposed tax and transfer programs have had to do with family stability and savings behavior.

3.3. Macro Feedback Effects

Another focus of public debate concerning tax and transfer programs are possible macroeconomic feedbacks. It is argued, for instance, that programs such as Social Security and unemployment insurance which raise the cost of labor to employers may cause general changes in the economy (such as the substitution of capital for labor) which in turn will affect individual employment opportunities. Programs of these sorts may affect both the inflation and unemployment rates, and inflation and unemployment affect the real incomes of individuals and families. As another example, cuts in personal income tax rates are usually thought to spur consumer spending, and the macro effects of this are thought to lead to more job opportunities and higher earnings for individuals and families.

In trying to incorporate macro feedbacks into a microsimulation framework, the model builder faces two main problems. First, the empirical evidence concerning many of these hypothesized effects is inadequate. Second of all, there are difficult choices to be made with respect to the representation of macro phenomena in a micro simulation context (see NAKAMURA and NAKAMURA, 1985 a and 1985 d.)

IV. Implementation of Tax and Transfer Programs in Microsimulation Models

The problems of capturing the effects of various types of tax and transfer programs have already been briefly discussed in Sections II and III. The direct effects of a program are the changes which would result from its application exactly as written, with no induced behavioral changes. In order to capture these direct effects, all variables which are directely affected by or which play a significant role in the administration of a

tax or transfer program must be carried in the microsimulation model. Consider a tax on earned income which is calculated based on the level of earned income and a table of deductions for dependents. Dependent deductions, in turn, vary with the number and ages of children in the family and, if applicable, the level of earned income of an employed spouse. The direct effects of such a program can be captured in a microsimulation model in which the earned income, number and ages of children, and employment and earnings status of a spouse if present are carried as characteristics of each individual in the simulation population.

The question of whether second order and other indirect effects of a tax or transfer program can be reflected in a microsimulation model depends on whether the model embodies the appropriate behavioral relationships and explanatory variables. Consider the problem, for instance, of capturing second order behavioral effects on labor supply. These effects can be captured if there are relationships in the model for determining employment status and hours of work, and if these relationships embed the tax rate in some form (such as in an after-tax wage variable) (see NAKAMURA and NAKAMURA, 1981 and 1983). As another example, a higher order effect on fertility could be captured if the model contains a behavioral relationship for, say, having a baby in the current period which embeds the tax rate in some form. For instance, for a married couple, fertility might depend positively on the after-tax earnings of the husband and negatively on the after-tax wage rate of the wife.

There are certain basic characteristics of microsimulation models which facilitate (or hamper) the incorporation of various sorts of tax and transfer program effects. The static/dynamic distinction is the most basic of these distinctions. Both static and dynamic microsimulation models incorporate an initial population for some base year. For instance, this initial population might be a sample representation of the U.S. population for a year for which Census Public Use Sample data can be used to determine most of the characteristics of the individuals in the initial population. The difference between static and dynamic microsimulation models lies in the way in which the composition and characteristics of the initial population are changed over a specified simulation period.

In **static models**, the initial population is advanced over time by reweighting the individuals in it to reflect expected changes over the designated simulation period in the population of interest. As a consequence of this reweighting, individual identities are not preserved over time. Thus static models are not suitable for analyzing tax and transfer programs which entail individual payment or benefit histories, or for analyzing longitudinal effects (including usage) of programs such as welfare support.

Even when the objective is simply to capture the current direct effects of programs which do not rely on any historical characteristics of individuals, there are problems with the static approach. Any analysis of direct effects will depend crucially on the distribution of earned income and other characteristics in the simulation population over the time period of policy interest. In this regard, ELLIASON (1986, p. 99) observes that the "distributional characteristics of the micro database" are "decisive for the results". Yet there is no clear theoretical or empirical basis for the reweighting processes which determine these distributional characteristics in a static microsimulation model. Of particular importance, static models do not embody any sort of a behavioral model for the determination of labor supply and earned income. Thus there is no satisfactory way of capturing second order, higher order, macro, or economic efficiency or dead weight loss effects of tax and transfer programs that involve labor supply behavioral responses. (For definitions of and discussion of the potential importance of dead weight loss effects see, for instance, DIEWERT, 1981 and 1984; HAUSMAN, 1981 and KEELEY, 1981.)

Despite these and other recognized problems, most of the microsimulation analyses of tax and transfer programs that have been carried out have utilized static models (see, for example, BEEBOUT, 1986 and MCCLUNG, 1986 for U.S. applications; and LIEFMEYER, 1986 for European applications). This may be largely because static microsimulation models have been viewed as less costly to build and maintain than dynamic ones. Some also may view static models as a more credible basis for policy analyses precisely because these models explicitly embody so little in the way of behavioral relationships. This aspect of static microsimulation models is emphasized sometimes by referring to them using terms such as "tax calculators". The implication is that the models entail purely mechanical applications of the proposed programs, and hence no behavioral relationships that might be subject to question. Note, however, that the absence of specified behavior is also a "behavioral assumption".

Miscrosimulation models such as the DYNASIM model developed by ORCUTT, CALDWELL and WERTHEIMER (1976) are **dynamic**. WERTHEIMER, ZEDLEWSKI, ANDERSON and MOORE (1986, p. 205) write:

> The crucial difference between DYNASIM and other (static) microsimulation models is its reliance on dynamic aging. The population is moved forward in time through the simulation of the aging process on the living population and through reweighting of the sample.... [The dynamic] model is well suited to simulations requiring the accumulation of longitudinal data. Simulations of social security and private pensions are two examples of such exercises.

Even in a dynamic microsimulation context, however, the range of possible tax and transfer program effects which can be captured depends crucially on the richness of the model. For instance, second order labor supply effects or dead weight loss effects cannot be captured in dynamic models where reduced form functions are used to determine the probability of employment and hours of work (unless the tax rate enters as a variable in these reduced form functions). The empirical basis for modeling many of the hypothesized indirect effects of tax and transfer programs is weak. Nevertheless dynamic models offer an appropriate framework within which to capture these effects

if, and when, the empirical foundations for these modeling efforts are more adequately developed.

V. Choosing Among Competing Behavioral Hypotheses

Dynamic microsimulation brings model builders face-to-face with the problems of choosing among competing behavioral hypotheses: the problem of model choice, as it is termed in the econometric literature. Twenty years ago Guy ORCUTT (1968, p. 94) wrote:

> Economic research has become quantitative in nature: it does avail itself of the best that statisticians and econometricians have had to offer, it does make extensive use of mathematics and computers, and it has even begun to seriously use sample survey data. In other words economic research has adopted most of the trappings of modern science and technology, but it still fails to achieve adequate testing of basic economic hypotheses. Theories come and go but not on the basis of sound and convincing evidence.

Unfortunately ORCUTT's description of the state of the art of model choice is still applicable.

In the ideal textbook case, model choice is a manageable problem. Economic theory or other a priori knowledge provides the form and variables for a behavioral relationship. The model choice process (limited on the basis of a priori knowledge) then proceeds as follows:

1. Statistical methods are called upon only to provide the precise magnitudes of the coefficients of the model, to aid in the determination of the properties of the error term, as a check on logical errors in deriving the model from the underlying theory or possible problems with the data collection process, and to provide the basis for the construction of confidence intervals and for tests of significance for the coefficients of the model.
2. Considerable weight is placed on demonstrating that the estimated behavioral responses of interest are stronger than what might be expected by chance. Yet it is recognized that the statistical tests used for this purpose are sensitive to the assumed properties of the unobserved error term. Thus JOHNSTON (1984, p. 505) cautions that "it is essential to examine the properties of the disturbance term in order to assess the validity of the statistical tests being applied."
3. Economic theory is the final arbiter of when a satisfactory empirical model has been achieved, subject to the aforementioned specification error tests (some of which also presume that the systematic part of the model is correctly specified).

4. There is an unstated presumption that the underlying behavioral process can be fully represented in terms of a "true" model consisting of an equation (or possibly a system of equations) which is simple enough in its form, has few enough observable variables, and which has a sufficiently well-behaved error term that it can be consistently estimated using available econometric methods and available or obtainable data.

Suppose that the available theory and other a priori evidence leaves some limited doubt as to the choice of explanatory variables, or the functional form relating the dependent variable to the explanatory variables. Suppose, for instance, that two competing theories suggest two alternative specifications which differ in terms of the included explanatory variables, and suppose furthermore that both of these specifications yield satisfactory empirical models according to the criteria laid out above in our discussion of the simple textbook approach to model selection.

If the competing models can be expressed so that one is a special case of the other, then nested model selection statistical tests can be used to differentiate them (see, for instance, KMENTA, 1986, pp. 593 - 595 and JUDGE et al., 1985, pp. 855 - 880). These techniques result in one or the other of the alternative models being selected as the "true" model. If neither model can be formulated as a special case of the other, then nonnested statistical tests might be used (see, for instance, KMENTA, 1986, pp. 595 - 598 and JUDGE et al. 1985, pp. 881 - 884). Nonnested tests can potentially lead to the rejection of both models. In this case, either or both of the alternative models might be reformulated and reestimated, and the preferred nonnested test might be applied again to see if either model could be accepted as the "true" model.

Notice that this more involved model selection procedure subsumes all of the steps, and hence all of the properties, of the simplest case with the exception that economic theory is no longer the final arbiter of when a satisfactory empirical model has been achieved. However, economic theory is still used to restrict the model specification as much as possible prior to estimation, and a satisfactory empirical model must still be found to be in agreement with theoretical expectations concerning the signs and magnitudes of the estimated coefficients.

Unfortunately, however, in areas of empirical investigation such as fertility behavior, family formation, and even work behavior, available theory and other a priori information leave more than limited doubt as to the appropriate specification of the relationships to be estimated. Nor is it obvious that the isolation of a unique "true" model is a realistic or even a desirable goal in a microsimulation modeling context. From a modeling perspective, what is really needed are relationships which capture key features of the underlying behavior in an approximation sense. There is a long tradition in economic theory of viewing models as simplifying approximations of reality. The testing methodology which has been developed in econometrics, however, is not fully compatible with a view of models as approximations, except to the extent that the approximation aspects of a model can be subsumed in an error term which obeys a list of conditions that must be satisfied in order for estimates of the parameters of the model to be consistent. Moreover most of what have come to be regarded as the more sophisticated econometric model choice methodologies cannot be used for appraising and ranking alternative approximate models.

Procedures for choosing among competing models based on their predicted outputs may be more compatible with the objectives of microsimulation model builders. **Output space model evaluation methods** have been most widely used in the development of macroeconometric models. Predicted time series for key variables, that may result from the interaction of a number of equations in a macroeconometric model, are compared in various ways with the actual time series for these variables. A better fit, in some specified sense, is taken to be evidence of a better model. Thus these methods can be used to rank alternative macro models, or alternative versions of the same macro model.

Two sorts of disparaging observations are sometimes made with regard to the output space evaluation procedures applied to macroeconometric models. The first basically has to do with the paucity of the available data at the macro level. In a macro data environment, only a single value is observed in a given unit time period (usually a year or a quarter of a year) for the dependent variable of a model. Thus only time series comparisons between the actual and predicted values of the dependent variable are possible. All available observations, and most of the information contained in these observations, are usually used in estimating macro models. As a result, in-sample predictive evaluations of the model are of limited usefulness. However, out-of-sample model evaluation cannot be carried out until more data become available with the passage of time. Moreover, in most cases the out-of-sample observations on the variable of interest will be largely an extrapolation of the in-sample values, since most macro time series are highly autocorrelated. Thus even out-of-sample predictive comparisons will not provide a very rigorous basis for evaluating a macro model.

This first criticism is not as relevant in a micro data environment largely because of the abundance and richness of micro data in many areas. Often there are sufficient data that some portion of the data can be used for estimation and the rest can be reserved for immediate out-of-sample testing. The out-of-sample data are not usually autoregressively related to the in-sample data as is the case with macro time series data. Thus the out-of-sample tests that can be carried out are more meaningful. Also in a micro data environment it is often the case that not all of the information in the in-sample data is used in estimating the model. This is particularly likely to be so with panel data. In this situation, even in-sample predictive tests may provide useful insights into the strengths and weaknesses of an estimated model (see NAKAMURA and NAKAMURA, 1985 a, 1985 b and 1985 c).

The second criticism which is leveled against output space model evaluation methods is that they only appraise the abilities of models to capture the observed movements of the dependent variable of interest. More specifically, these methods are seen solely as indicators of the predictive abilities of models. This is an important concern since interest in dynamic microsimulation stems largely from the hope of capturing important behavioral responses.

It is true that output space evaluation methods were first developed in a macroeconometric forecasting environment where predictive ability is of paramount importance. It is also true that the model which provides the most accurate forecasts for some dependent variable in, say, a mean square error sense, will not necessarily provide the most accurate estimates of how the dependent variable will change in response to a specified change in some explanatory variable. This point is often made in the context of discussing the relative merits of a structural versus a reduced form representation of a variable. Yet there is a relationship between the predictive ability of a model and the ability of a model to properly capture responses to the included explantory variables. Moreover an estimated model, including the assumed and estimated properties of the error term, should be able to reproduce the observed distribution of the dependent variable conditional on the observed values for the explanatory variables. If it can not, then at the very least there is probably some difficulty with the specification of the error term, which may mean that the standard tests of significance are inappropriate.

A key advantage of output space model evaluation methods is that they yield distance measures for the **relative** goodness of alternative models, and these distance measures have meaning whether or not any one of the alternative models can be viewed as the "true" model. Also the validity of these comparisons does not rest as heavily on specification assumptions, and hence on our ability to detect departures from these assumptions via a priori reasoning or specification error tests.

One promising class of output space model evaluation methods are based on variants of PEARSON's chi-square statistic for goodness-of-fit. Chi-square statistics have traditionally been used to test whether an estimated model can be regarded as the "true" model in the sense that the observed distribution of the dependent variable can be regarded as a drawing from the population distribution defined by the estimated model. Chi-square type statistics can be used for diagnostic purposes in the spirit of conventional specification error tests (see ANDREWS 1988 a, 1988 b for example). Chi-square statistics can also be used, however, to rank alternative models in terms of the degree of congruence between the predicted and actual distributions of the dependent variable(s) of interest. The first application of this sort that we are aware of in the economics literature is due to HECKMAN (1981). HECKMAN's application of a chi-square type ranking method involves models for the simple binary choice of working versus not working in each successive year. NAKAMURA and NAKAMURA (1983) apply this method to the evaluation of models for hours of work involving the decision to work part-time or full-time, and the choice of hours of work for those found to work. Thus the predictions for hours of work result from a system of two-limit probit and regression relationships. NAKAMURA and NAKAMURA (1985 a, 1985 b) further extend this approach to the evaluation of multi-equation models for annual earned income. Both the cross-sectional distributions for the earned incomes of various types of individuals, and the distributions of individual earned income cumulated over a seven year period are utilized in the model evaluation process. HECKMAN and WALKER (1988) use a chi-square type of statistic, as well as other approaches, to choose among alternative models of fertility. For marketing applications that may be relevant to microsimulation model builders, see BAGOZZI and YI (1988).

Output space analysis as carried out in NAKAMURA and NAKAMURA (1983, 1985 a, 1985 b) is implemented using microsimulation over a historical period for which the actual values of the dependent variable(s) of interest can be observed. Thus the microsimulation methodology is being harnessed to help cope with the fundamental model choice problems encountered in developing dynamic microsimulation models (see NAKAMURA, NAKAMURA, and DULEEP, forthcoming).

VI. Conclusions

Dynamic microsimulation is a promising methodology for analyzing tax and transfer program effects. Fulfillment of this promise rests on two conditions. First of all, the operational processes used to advance the simulation population over time must be well justified in both a behavioral theory and an empirical sense. Second of all, these models must incorporate the behavioral relationships and explanatory variables required to implement and represent the key direct and indirect effects of the tax and transfer programs of interest. Thus fulfillment of the high hopes many hold for dynamic microsimulation depends on empirical efforts to improve our understanding of the relationships for births, marriages, divorce, labor supply and so forth which are the behavioral essence of a dynamic microsimulation model.

Model choice is an important aspect of these needed empirical efforts. There must be some way, for instance, of choosing among alternative models of the labor supply of married women which embody drastically different responses to changes in the after-tax wage rate. Given the objectives of microsimulation model builders, output space analysis may have important advantages over more conventional approaches to the model choice problem. In a micro data setting, output space analysis makes use of the microsimulation methodology, with simulated and actual realizations of the variables of interest being compared over a historical time period.

References

ANDREWS, D.W.K., 1988 a, Chi-Square Diagnostic Tests for Econometric Models: Introduction and Applications. Journal of Econometrics 37 (1), 135 - 156.

ANDREWS, D.W.K., 1988 b, Chi-Square Diagnostic Tests for Econometric Models: Theory. Econometrica 56 (6), 1419 - 1453.

ASHENFELTER, O. and J.J. HECKMAN, 1974, The Estimation of Income and Substitution Effects in a Model of Labor Supply. Econometrica 42, 73 - 85.

BAGOZZI, R.P. and Y. YI, 1988, On the Evaluation of Structural Equation Models. Journal of the Academy of Marketing Science 16, 74 - 94.

BEEBOUT, H., 1986, Evaluating Reagan Administration Social Program Changes: Two Applications of Math, in: ORCUTT, MERZ, and QUINKE, eds., 83 - 97.

BROWNING, M.J., A. DEATON, and M. IRISH, 1985, A Profitable Approach to Labor Supply and Commodity Demands Over the Life-Cycle. Econometrica 53, 503 - 543.

DIEWERT, W.E., 1981, The Measurement of Deadweight Loss Revisited. Econometrica 49, 1224 - 1225.

DIEWERT, W.E., 1984, The Measurement of Deadweight Loss in an Open Economy. Economica 51, 23 - 42.

EDWARDS, J.B. and G.H. ORCUTT, 1969, Should Aggregation Prior to Estimation Be The Rule? Review of Economics and Statistics 31, 409 - 420.

ELLIASON, G., 1986, Comment on the Paper by Beebout, in: ORCUTT, MERZ, and QUINKE, eds., 98 - 100.

HAUSMAN, J.A., 1981, Exact Consumer Surplus and Deadweight Loss. American Economic Review 71, 662 - 676.

HAVEMAN, R.H., K. HOLLENBECK, D.BETSON, and M. HOLMER, 1980, Microeconomic Simulation Model for Analyzing the Regional and Distributional Effects of Tax-Transfer Policy: An Analysis of the Program for Better Jobs and Income, in: R.H. HAVEMAN and K. HOLLENBECK, eds., Microeconomic Simulation Models for Public Policy Analysis, Vol. 2, (New York), 137 - 162.

HECKMAN, J.J., 1974, Life Cycle Consumption and Labor Supply: An Explanation of the Relationship Between Income and Consumption Over the Life Cycle. American Economic Review 64, 188 - 194.

HECKMAN, J.J., 1976, A Life-Cycle Model of Earnings, Learnings, and Consumption. Journal of Political Economy 84, 11 - 44.

HECKMAN, J.J., 1981, Heterogeneity and State Dependence, in: S. ROSEN, ed., Studies in Labor Markets (Chicago), 91 - 139.

HECKMAN, J.J. and I.R. WALKER, 1988, Using Goodness of Fit and Other Criteria to Choose Among Competing Duration Models: A Case Study of Hutterite Data, in: C.C. CLOGG, ed., 1987, Sociological Methodology (Washington, D.C.).

HOTZ, V.J., F.E. KYDLAND, and G.L. SEDLACEK, 1988, Intertemporal Substitution and Labor Supply. Econometrica 56, 335 - 360.

JOHNSON, T.R. and J. PENCAVEL, 1982, Forecasting the Effects of a Negative Income Tax Program. Industrial and Labor Relations Review 35, 221 - 234.

JOHNSTON, J., 1984, Econometric Methods, 3rd ed. (New York).

JUDGE, G.G., W.E. GRIFFITHS, R.C. HILL, H. LUTKEPOHL, and T.C. LEE, 1985, The Theory and Practice of Econometrics. 2nd ed. (New York).

KEELEY, M.C., 1981, Labor Supply and Public Policy: A Critical Review (New York).

KMENTA, J., 1986, Elements of Econometrics. 2nd ed. (New York).

LIETMEYER, V., 1986, Microanalytic Tax Simulation Models in Europe: Development and Experience in the German Federal Ministry of Finance, in: ORCUTT, MERZ, and QUINKE, eds., 139 - 152.

MACURDY, T., 1981, An Empirical Model of Labor Supply in a Life-Cycle Setting. Journal of Political Economy 89, 1059-1085.

MACURDY, T., 1983, A Simple Scheme for Estimating an Intertemporal Model of Labor Supply and Consumption in the Presence of Taxes and Uncertainty. International Economic Review 24, 265 - 289.

MCCLUNG, N., 1986, The Art of Transfer Policy Analysis, in: ORCUTT, MERZ and QUINKE, eds., 101 - 112.

MERZ, J., 1986, Structural Adjustment in Static and Dynamic Microsimulation Models, in: ORCUTT, MERZ, and QUINKE, eds., 423 - 446.

NAKAMURA, A. and M. NAKAMURA, 1978, On Microanalytic Simulation and Its Applications in Population Projection. Journal of the Operational Research Society 29, 349 - 360.

NAKAMURA, A. and M. NAKAMURA, 1981, A Comparison of the Labor Force Behavior of Married Women in the United States and Canada, with Special Attention to the Impact of Income Taxes. Econometrica 49, 451 - 490.

NAKAMURA, A. and M. NAKAMURA, 1983, Part-time and Full-time Work Behavior of Married Women: A Model with a Doubly Truncated Dependent Variable. Canadian Journal of Economics 16, 229 - 257.

NAKAMURA, A. and M. NAKAMURA, 1985 a, The Second Paycheck: A Socioeconomic Analysis of Earnings (New York).

NAKAMURA, A. and M. NAKAMURA, 1985 b, Dynamic Models of the Labor Force Behavior of Married Women Which Can Be Estimated Using Limited Amounts of Past Information. Journal of Econometrics 27, 273 - 298.

NAKAMURA, A. and M. NAKAMURA, 1985 c, Rational Expectations and the Firm's Dividend Behavior. Review of Economics and Statistics 67 (4), 606 - 615.

NAKAMURA, A. and M. NAKAMURA, 1985 d, Incorporation of Micro and Macro Information in Microanalytic Models. Unpublished working paper.

NAKAMURA, A., M. NAKAMURA, and H. DULEEP, forthcoming, Alternative Approaches to Model Choice. Journal of Economic Behavior and Organization.

NAKAMURA, M., A. NAKAMURA, and D. CULLEN, 1979, Job Opportunities, the Offered Wage, and the Labor Supply of Married Women. American Economic Review 69, 787 - 805.

ORCUTT, G.H., 1968, Research Strategy in Modeling Economic Systems, in: D.G. WATTS, ed., The Future of Statistics (New York), 71 - 95.

ORCUTT, G.H., S. CALDWELL, and R. WERTHEIMER, 1976, Policy Exploration Through Microanalytic Simulation, The Urban Institute (Washington, D.C.).

ORCUTT, G.H., M. GREENBERGER, J. KORBEL, and A. RIVLIN, 1961, Microanalysis of Socioeconomic Systems: A Simulation Study (New York).

ORCUTT, G.H., J. MERZ, and H. QUINKE, eds., 1986, Microanalytic Simulation Models to Support Social and Financial Policy (Amsterdam and New York).

ORCUTT, G.H., H.W. WATTS, and J.B. EDWARDS, 1968, Data Aggregation and Information Loss, American Economic Review 58, 773 - 787.

PECHMAN, J.A., 1965 a, A New Tax Model for Revenue Estimating, in: Proceedings of a Symposium on Federal Taxation, American Bankers' Association (Washington, D.C.).

PECHMAN, J.A., 1965 b, A New Tax Model for Revenue Estimating, in: A.T. PEACOCK and G. HAUSER, eds., Government Finance and Economic Development, OECD (Paris).

PENCAVEL, J., 1986, Labor Supply of Men: A Survey, in: O.C. ASHENFELTER and R. LAYARD, Handbook of Labor Economics, Vol. 1, 3 - 102.

ROBINSON, C. and N. TOMES, 1985, More on the Labour Supply of Women. Canadian Journal of Economics 18, 156 - 163.

SMITH, J.B. and M. STELCNER, 1988, Labour Supply of Married Women in Canada, 1980. Canadian Journal of Economics 21, 857 - 870.
STELCNER, M. and J. BRESLAW, 1985, Income Taxes and the Labour Supply of Married Women in Quebec. Southern Economic Journal 51, 1053 - 1072.
WERTHEIMER, R., S.R. ZEDLEWSKI, J. ANDERSON, and K. MOORE, 1986, Dynasim in Comparison with Other Microsimulation Models, in: ORCUTT, MERZ, and QUINKE, eds., 187 - 206.

Female Labour Supply in Micro-Simulation Models

ALDI J.M. HAGENAARS

I. Introduction

Micro-simulation models have become very popular in the last few years. They are presently used to analyze the effect of various potential changes in government policy, to predict the changes in household composition and size, to analyze the effect of a change in the age composition of the population on the demand for housing, health care, and care for the elderly, to name but a few of the applications. All these models have in common that they try to use information at the micro-level of households or individuals, rather than aggregate statistics or some "modal" or "representative" family or person.

In this book we are concerned with micro simulation models of the effect of changes in tax and transfer policy. By taking a representative sample of the population one is interested in, analyzing the proposed or expected changes at the micro level, and then aggregating these micro changes for the whole sample, the effect of proposed changes in tax structure and transfers can be calculated with considerably more precision than by the use of aggregate information.

In spite of this progress in policy analysis by the use of micro-simulation models, one large problem, however, still remains. Micro-simulation models usually consider the effect of a certain policy change on the income distribution, the government budget, and various other aspects one may be interested in, **assuming that people do not change their behaviour** as a result of this change in policy. Hence, a considerable number of micro-simulation models is basically static in nature, and ignore behavioral changes as a result of government policy.

This may sometimes lead to dramatic under- or over-estimates of the budgetary cost involved in a certain new policy. As an example a recent Dutch change in student grants may serve: the new system provided a basic income for all full-time students, instead of a grant that depended on the income of the parents of the student. On top of this (relatively low) basic income, loans free of rent would be made available to all students.

In the year following this change in system, the number of students increased by a large percentage. This change had not been anticipated, and caused major financial problems. Within two years after the introduction of the new system, the basic income