Review of Income and Wealth
Series 37, No. 2, June 1991

CONSTRUCTING A TAX-BENEFIT MODEL: WHAT ADVICE CAN ONE GIVE?

BY HOLLY SUTHERLAND

London School of Economics

This paper takes the form of advice given to a hypothetical government on the construction of a tax-benefit model. The importance of the quality of the data and the care taken in its adjustment is emphasised. The choice of the unit of analysis, the requirements of the income data, methods of updating and the coverage and representativeness of the dataset as a whole is discussed. Caution in making use of estimates of behavioural change is advised. Throughout, examples drawn from the construction of TAXMOD, a model for the U.K., are used to illustrate particular points.

1. INTRODUCTION

The proposition that lies behind this paper is that the author has been asked to advise a hypothetical government on setting up a tax-benefit model for their country so they may design and analyse the effect of a series of tax and social security reform proposals. This paper, the documentation of this exercise, has two audiences in mind. The first consists of those who are unfamiliar with tax-benefit models and the problems associated with their construction. This audience can put themselves alongside the hypothetical government and what follows can serve as a guide to these models. The second audience are those already familiar with tax-benefit models. For them, this exercise is intended as a description of the dialogue that needs to take place between the eventual user of the model (the "government" in this case) and those giving advice on its construction.

In what follows, the hypothetical government is written as the *government* to distinguish it from any actual government. Much of the argument also applies to non-government institutions who wish to set up a model.

The arguments in the paper are based on experience over the past five years of constructing a model of the U.K. tax-benefit system, known as TAXMOD[1] and on the experience of trying to give advice to those interested in constructing new models. It has also drawn on the author's limited knowledge of some other tax-benefit models both in the U.K. [for example by the Institute for Fiscal Studies (see Dilnot *et al.*, 1988)] and for other countries such as SYSIFF for France (see Bourguignon *et al.*, 1988) and the SPSD/M produced by Statistics

Note: I would like to thank the participants of the 21st General Conference of the International Association for Research in Income and Wealth at Lahnstein, W. Germany, in particular A. Harding, B. Kennedy and A. Nakamura for their helpful comments. I would also like to thank A. B. Atkinson and J. Gomulka for their continuing guidance. All faults that remain are my own.

[1]More details of this model as well as descriptions of some other tax-benefit models and examples of their use are available in Atkinson and Sutherland (1988). The results from the model reported in this paper are for 1989-90 from TAXMOD version 7.3, which uses data from the 1982 *Family Expenditure Survey*. Copies of the model are available on floppy disk from the author at marginal cost.

Canada. Models are being built for an increasing number of countries using a variety of methods. For example, see Gegesy *et al.* (1989) for Hungary; Galler (1989) and Hüther *et al.* (1989) for W. Germany. Overviews of the current state of the art of micro-simulation, of which tax-benefit models form a part, have been published at various times, for example in Mertz (1988) and, focussing on government income tax models only, OECD (1988).

I have tried to keep the arguments as general as possible, but have demonstrated particular points with reference to the U.K. tax-benefit system and the particular solutions to problems chosen in the case of TAXMOD.

The hypothetical scene is set as follows; this *government* was keen to reap the political rewards of introducing a radical and effective package of reforms. Their officials, under pressure to produce speedy solutions to long-standing problems, were optimistic that a model based on micro-data, capable of simulating the effects of the various proposals, would quickly allow them to chose between competing alternatives. The author was happy to promote micro-modelling of the proposed reforms as a tool in policy analysis, but was concerned that the complexity of this approach and its limitations should be understood.

In particular, it was important for the author to make clear at the beginning that any model would only be as good as the data available to it. If, for example, there were no sources of information on holdings or transfers of capital assets then it would not be possible to model the taxation of capital. If coverage of the data source were limited in some way, for example to households in urban areas or to individuals in employment, then the task of assessing the *national* effect of change would be impossible. The bulk of her advice, and of this paper, concerned understanding what the data were capable of modelling and what would have to be left to other methods to analyse. Clearly, the problems to be encountered were all a matter of degree as it is unlikely that *perfect* data for any task will ever be available. It is always a question of balance as to whether a particular exercise is worthwhile. If it is, then it is important to assess the extent to which assumptions made affect the final results. In the absence of a method of systematically analysing the sensitivity of the model, some examples of how alternative assumptions would affect the results in the case of TAXMOD are included as illustrations in what follows.

Section 2 explains why the *government* needed a tax-benefit model and lists the type of reforms it was considering. It concludes that a static general model of the whole current tax and benefit system and any likely changes to it was what was required. Section 3 describes the data that should be included, with particular attention to the income data. Section 4 discusses problems that arise when the data do not accurately describe the required population and describes one approach to these problems. In section 5 the sorts of adjustments that need to be made to model the current system and a policy change are discussed and ways in which actual departures from the official rules can be modelled are considered. This is followed in section 6 by a brief description of the difficulties involved in modelling the indirect or dynamic effects of change in the tax-benefit system. In the final section the recommendations made to the *government* by the author are summarised.

2. The Initial Requirement

On closer examination it was discovered that the *government's* policy objectives were not always clear, consistent or precisely formulated. Various departments and ministries within the *government* had different objectives which included:

* revenue neutrality,
* redistribution of cash towards families with children,
* improvement of work incentives,
* lowering of income tax rates,
* producing a tax system that is accepted as "fair,"
* targetting expenditure particularly on the disabled and those caring for them,
* simplifying the tax and benefit systems and reducing administrative costs,
* reducing dependency on means-tested benefits.

It was not to be the author's job to prioritise these proposals, but to provide a tool for the analysis of the possible reforms, either separately or in combinations. The requirement was for a model which would estimate the total revenue cost of any reform, the impact of each component of a package of reforms, the overall distributional effect, the effect on particular groups in the population, and the incentive effects, both overall and on particular groups. This was in line with the approach in TAXMOD which uses data from a representative sample of households and calculates the benefit entitlements and tax liabilities for each unit under the present system, according to current rules. The rules governing the tax or benefit system are changed according to the policy proposal and the taxes and benefits are calculated again for each unit. This allows net incomes before and after the policy change to be calculated. The difference between these, appropriately weighted and added up for each family in the sample, would provide the *government* with their estimate of the revenue cost of the change for the whole economy. It would also be straightforward to calculate marginal tax rates for members of each family unit, by repeating the calculations after adding a marginal amount to incomes, and comparing net incomes with those previously calculated. It would also be possible to use the other information in the survey to focus on the effect of a change in policy on particular groups in the population—by region, family composition, age, employment status, income range or whatever else was of interest.

It was made clear to the *government* that the specification of the rules governing both the current system and the various proposals would be required in detail. Some specific proposals had been publicly discussed, although not always with the degree of precision required to test them in a model. These included:

* the extension of a means-tested family benefit to become a universal benefit contingent on citizenship alone,
* a reduction in income tax rates accompanied by the abolition or limitation of selected tax reliefs,
* the introduction of an alternative minimum tax schedule,

* an increase in some or all indirect taxes,
* a complete overhaul and simplification of the taxation of capital transfer,
* a new non-contributory benefit for the disabled,
* the introduction of a property tax.

One of the first recommendations made to the *government* was that their proposals should be formulated in more detail—with the precision necessary to turn them into computer-readable code.

Although the dialogue is carried out with the *government*, it became clear that the various opposition parties had their own ideas for reform and that tax and benefit reform would not end with this particular exercise. What was needed was a general model of the system in this country and of all possible or likely changes to it. It was clear that the specification of rules describing policy should be as flexible as possible so that the model would be useful in analysing as yet unforeseen proposals in the future.

3. Assessing the Household Data

The *government* was to provide a dataset with which to construct a model. It was a sample survey of some 5,000 households dating from four years previously. It was conducted by personal interview and the quality of the data was believed to be high, including questions on income, expenditure, housing and family composition. In these respects it was similar to the dataset used in the U.K. model, the *Family Expenditure Survey* (FES) and was therefore thought to be a suitable source on which to base a model of the tax and benefit systems.

Unit of Analysis

The dataset was a household dataset, meaning that the sample was selected by choosing whole households. These were defined as groups living in the same dwelling with domestic arrangements in common. Households so defined could contain more than one family unit (single person or couple—legally married or *de facto*—and their dependent children). How often the household is not synonymous with the family unit, or in other words how many multi-family unit households there are (or, indeed, whether these are the appropriate units at all) depends on the specific social structure in the country. The FES shows that for the U.K. in 1982, 40 percent of households included more than one family unit, indicating that the two should not be treated as identical in this case. Although in other countries the relevant income units may be different, the discussion here will focus on three; household, family and individual.

Each piece of data was to be available at a household level (household income, housing costs, number of children aged under five and so on), but it was important for some purposes that data were available for other levels of analysis. In particular, the personal tax system in this country was based on the individual as tax unit, with no link between spouses. It was therefore vital that income information was available for every individual in the household, rather than as an aggregate. On the other hand, the means-tested family benefit was assessed on the income of the family unit and so it was necessary to be able to define a

family unit and calculate the necessary aggregate income variables. It had to be recognised that some definitions would not be possible. For example, in the U.K., TAXMOD is not able to distinguish legally married and *de facto* couples, a matter which would not be important in the country in question unless a policy that required them to be distinguished was proposed. It is of some significance in the U.K. where the tax treatment of couples depends on their legal status, but much of the means-testing carried out to calculate entitlement to social security benefits occurs on a family unit basis, regardless of the legal status of the couple. Some 4 percent of individuals aged 16 to 59 were co-habiting in 1986,[2] but account has not been taken of this in TAXMOD.

Another common problem in defining the family unit occurs with the definition of a "child" as this is often different in different contexts. For example, in the U.K., young adults aged over 18 but under 25 living with their parent(s) will be treated as independent adults by the tax and social security systems, although they will usually receive a lower level of social security benefit than those aged over 25. However, the FES does include them as part of the family unit if they are not married and have incomes too small to pay income tax or are full-time students. In TAXMOD, these "children" are split off from their parent's family unit to constitute their own. The definition of "child" that is used is a person for whom child benefit is received; someone aged under 16 or under 19 if in full-time education. Clearly the most appropriate definition of a child will depend on the particular institutional and social arrangements in the country in question.

A different, but related, question is the choice of the unit of analysis for the model as a whole. The choice of the unit usually corresponds to the main unit of assessment for taxes and benefits and it makes the most sense if it is the smallest unit for which one can assume that within it there is a sharing of resources. In the U.K. it is customarily assumed that resources are shared within the family and the family unit is usually chosen as the unit of analysis. However, there are no particularly good grounds for assuming that resources are always shared equitably and some evidence, such as that given in Brannen and Wilson (1987) for the U.K., that they are not. Women's groups, with a concern about the distribution of income between men and women and the impact of policy on this, will argue for the need to analyse at the level of the individual. A generally applicable model will need to take this requirement into account. However, problems will arise in having to make assumptions about the actual distribution of resources within the family when so little is known about this.

Similarly, if the family is selected as the unit of analysis, then assumptions must be made about transfers of goods or services within the household (for example housing services for secondary family units within the household). A common problem is that housing costs are only identified at the household level and surveys do not attempt to ask questions that would enable them to be apportioned among family units within the household. Some assumption has to

[2]Source: General Household Survey (OPCS, 1989). The effect of taking this into account in an FES-based model would depend on how many co-habitees were coded as single and how many as married in this survey, although evidence from the General Household Survey suggests that some would come from each category.

be made or housing costs cannot be modelled. There are clearly many plausible candidates; that all costs should fall on the head of household; that they are shared equally among adults in the household, or among the family units within the household; that they are shared but weighted according to characteristics (such as family unit income). In situations such as this, the choice made may depend as much on the modeller's own *a priori* assumptions as anything else, but the implications of the choice may be significant. For example, in the case of TAXMOD, the assumption is made that the head of household bears all the housing costs. An alternative assumption could be that secondary family units actually pay to the householder the amounts in rent and rates that are assumed in the calculation of householders' housing benefit.[3] If householders' own housing costs were also reduced by this amount, mean housing costs for those assumed to be liable for them would *fall* by 20 percent.

A similar, but more difficult problem exists with expenditure data. These are necessarily collected at the individual level, but are only meaningful at the household level, unless we are also provided with comprehensive information on the flows of income within the household or on individual consumption. This is clearly illustrated by the case of a household where all the food and clothing are bought by one person and all the fuel and interest payments are paid for by another. Income may only be received by one of them and clearly one cannot attribute indirect taxes on goods to those who actually carried out the purchases. If the entire analysis is to be carried out at the household level—that is the results presented in terms of gain and loss by household—then taxes on expenditure may be modelled without too many assumptions about incidence.

It did seem to the *government* that there were strong arguments in favour of choosing the household as the unit of analysis. Indirect taxes could be modelled and the assignment of housing costs would be straightforward. However, it was pointed out that this would imply that resources were evenly distributed within households and that any adverse effect of one of the policy proposals on, say, the elderly living with relatives, would not be clearly observable if their income was pooled with that of the other household members. Ideally, all possible levels of analysis should be attempted, bearing in mind the major assumptions that must be built in to make this possible.

Income Data

The data on income in the survey were detailed, providing individual income from different sources and in most cases documenting the time period over which the income was received. It was possible to define income variables which relate to a short time period (the previous week or the previous month) or a longer one (the previous year). Both of these were necessary as (as in the U.K.) income tax was assessed on cumulative annual income, but means-tested benefits were calculated on a more short-term basis. In the case of TAXMOD, it is assumed that income is received at the same rate, from the same sources, throughout the year. If the tax and benefit systems are proportional this assumption should not

[3]Since April 1989 in Scotland and April 1990 in England and Wales, domestic rates have been replaced by the poll tax which, being an individual tax, does not present these problems.

distort the revenue or income distribution estimates of the model, since observations are collected throughout the year and one person's temporary fall in income is matched by another's temporary rise. However, in any other case, some distortion will occur. Furthermore, where a tax or benefit is assessed on income received in a period *different* from that to which the data refers, then the model cannot capture the precise detail of the system.

In spite of the detail in which it is provided, the quality of the income data should not be taken for granted. Whether the grossed-up aggregate amounts for each source of income are equal to aggregates from other sources, in particular to those in the national accounts, needs to be investigated. One such study by Atkinson and Micklewright (1983) for the FES concluded that (for 1970 to 1977) many of the aggregates compared favourably with the U.K. national accounts, once differences in definition and some aspects of differential non-response had been taken into account. The two areas of particular concern were self-employment income and investment income. There was believed to be significant under-reporting of both these sources of income and lower response rates by the self-employed, in the first case, and by those with very high income (and a large share of the investment income), in the second. In TAXMOD, to compensate for under-reporting, *ad hoc* increases in these variables have been introduced to bring them up to levels comparable with those given by the national accounts. The treatment of differential non-response is discussed in the next section.

However, as Atkinson and Micklewright point out, there are uncertainties about using the national accounts as a point of reference for all purposes. In some cases the national accounts estimates are partially based on the survey data itself. Some national accounting definitions of income are not the most appropriate for our purpose. For example, lump sum pension payments are included in the U.K. accounts item "pensions and other benefits from life assurance and superannuation schemes," whereas TAXMOD deliberately excludes lump sum incomes.

In addition to assessing the validity of the survey income *aggregates* and making corresponding adjustments, the *distribution* of income from each source ought also to be explored. This is clearly a major task which has been attempted for earnings in the U.K. by Thatcher (1968) and Atkinson, Micklewright and Stern (1988).

4. Coverage and Representativeness of the Population

The basic assumption behind the kind of general purpose model being proposed is that the data used can be considered to be representative of the required (in this case, national) population. There are a number of respects in which this will not be the case and for which adjustments will need to be made. The size of these adjustments and an assessment of their effect on the final result needs always to be borne in mind.

The Relevant Population

The first main assumption was that the survey should include the population relevant to the measures required—all the tax paid and the benefits received in the period required (usually the financial year).

For many reasons it may be tempting to exclude certain sorts of families that appear in the survey data; because they are very large and do not fit into a convenient data structure; because they are odd or unusual, for example including foster children; because they seem unlikely or impossible, for example families apparently with zero or negative incomes, or because they are in transition between employment statuses or marital statuses and are difficult to reconcile with tax and benefit rules. However, it is important to include all these families. It is often the "odd" cases that end up losing out from "simplifying" reforms. Changing jobs or living arrangements are legitimate activities that need to be modelled if the overall results are to be realistic.

This survey will not cover the entire resident population as it is a *household* survey and will exclude those not living in households, such as the armed forces living in barracks, people in hospital, nursing homes, prison etc. In the U.K., using information from the 1981 census, we can estimate that the numbers of pensioners living in institutions amount to 1.3 million tax units, about four percent of the total.[4] Furthermore, for some purposes it is not the resident population that is the most relevant. For example, state pensions may be payable to nationals who have made the requisite contributions but who are living abroad. These will not be counted in the adding up of the resident household population. In 1987, over four percent of state retirement pensions were paid to ex-residents of Great Britain living overseas. Not to take account of either of these missing components of the relevant population means that TAXMOD's assessment of the cost of increased spending on the retired would be under-estimated by up to nine percent.

The survey should include observations interviewed at all stages in the year so that where there is substantial seasonal variation, for example in unemployment or cash benefit payments for heating, account is taken of this. It is assumed that by sampling families taken at points throughout the year, the full range of changes that can occur are taken into account, such as wage rises, holidays, periods of sickness or unemployment.

In the second place, the data should be representative of the population in the actual year being modelled. However quickly the data is made available to analysts once it has been collected, it cannot be provided for the *current* year or for *future* years which are usually the relevant ones for the analysis of policy proposals. There is of course *some* virtue in modelling the introduction of a proposed policy for a year in the past but, to the extent that relevant aspects of a population change in the intervening period, this is less useful than modelling the current (or future) population. Adjusting for growth in money incomes or for inflation is usually a straightforward indexing by published data—for example from the national accounts for self-employment income or from the appropriate element in the Retail Prices Index (RPI) for rent. However, this may be seriously misleading. The earnings distribution has widened in recent years in the U.K. While the RPI rose by 41 percent between 1982 and 1989, the *New Earnings Survey 1989, Table A15* shows that the earnings of full-time adult men in the bottom decile grew by 54 percent between 1982 and 1989 while the earnings of

[4]Where the non-household population aged over retirement age represents five percent of the total over retirement age and assuming that all of them in fact receive a state retirement pension.

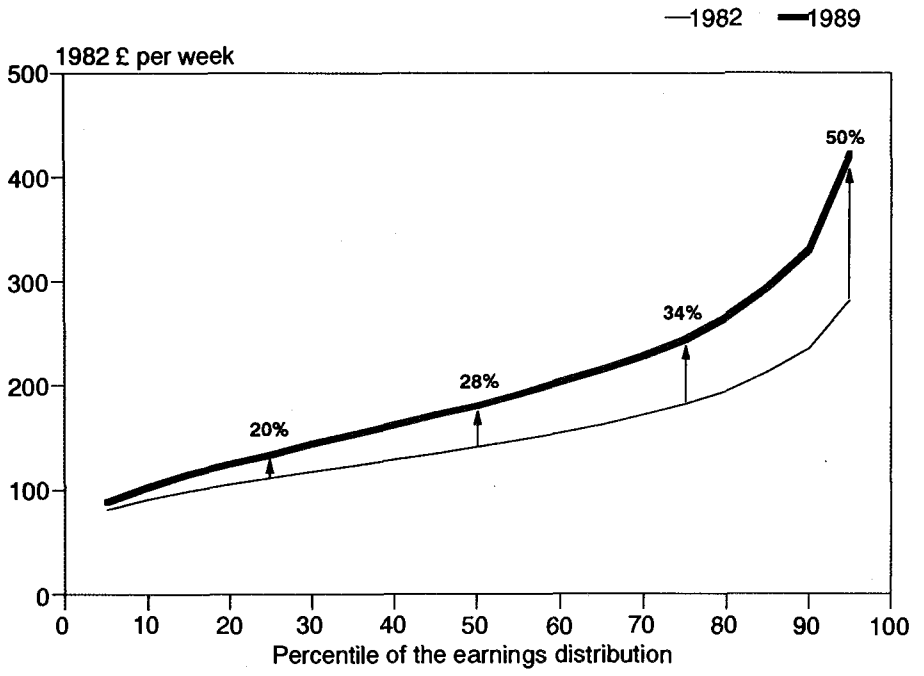

Figure 1. Real Shifts in the Earnings Distribution. Weekly, Male, Full-Time Earnings 1982–89

those in the top decile grew by an average of 81 percent in the same period. (The corresponding figures for women are 68 and 90 percent.) See Figure 1 for an illustration of how differential real earnings growth has affected the distribution of male, full-time earnings since 1982.

In an attempt to allow for this dispersion, earnings may be updated by a set of indices which vary with the point on the distribution of earnings, as shown in the figure. Inflating earnings according to their place in the earnings distribution has a noticeable effect on TAXMOD's results. For example, the estimate of expenditure on Family Credit rises by 14 percent if earnings are differentially adjusted.[5]

Apart from adjusting to take account of inflation and the impact of differential income growth, some relevant characteristics of the population can change dramatically in a short space of time and it is important to adjust the data to reflect these shifts. For example, in the U.K., owner occupation grew from 58.7 percent at the end of 1982 to 65.1 percent in 1988.[6] The number of people registered as unemployed fell from 2,917 million on average in 1982 to 1,799 in 1989.[7] Structural shifts cannot adequately be treated by simple indexing of individual data items. For example, the recent growth in owner-occupation in the U.K. cannot simply be allowed for by multiplying up the mortgage interest

[5]Family Credit is a means-tested benefit for families with children and at least one adult in full-time employment or self-employment.

[6]*Housing and Construction Statistics 1988*, Table 2.22.

[7]*Department of Employment Gazette*, June 1983 and May 1990, Table 2.1.

paid by owner-occupiers in the year the data were collected to give the correct current total. The distribution of interest payments would not be correct, giving under-estimates for the cost of the tax relief on mortgage interest. Also, this type of adjustment would not allow for the fact that people who have recently become owner-occupiers are different from the old owner-occupiers, both in their other characteristics (such as income) and in the amount they borrow. A more ambitious approach would be to make some of the tenants "buy" their housing, making assumptions about who they are, how much they pay, when they purchased and so forth. This approach changes the characteristics of the sample from the survey to make them appear as they would today, following an "ageing" technique such as that described by King (1987). This may be a realistic way to solve some of the problems of structural change, but would alter the character of the model as a whole. Ageing one characteristic of the family would require that all the other characteristics also be aged. This is a major task which has been attempted by the builders of dynamic micro-simulation models that are founded on the construction of pseudo longitudinal data, examples of which are described in Galler (1989) and Harding (1990). This approach was not considered feasible in the time available to the *government* for the construction of their model, particularly since the viability of the method would crucially depend on being able to make accurate underlying assumptions. It was recommended that available resources should be concentrated on processing the data as speedily as possible after its collection, to minimise the problem of structural change for the static model.

Representativeness

The second main requirement was that the survey must be large enough and the sampling techniques such that all types of family were represented. The full range of variation in the many relevant characteristics of families should be present in the dataset. Furthermore, these families must be represented in the same proportions as in the actual population if the model results are to predict the effect on this population. Even where the sample size and design appear to be well conceived, there are reasons why a particular survey will not accurately represent the population.

Unless responding to the survey is compulsory (or the complete information can be extracted from administrative sources), there will be some degree of non-response, either to the whole questionnaire by the household or by some individuals or to just some of the questions by some or all of the household. Some people will not wish to give large quantities of information about themselves and others will be particularly sensitive about income or other specific questions.[8] In the U.K., according to Kemsley *et al.* (1980), the response rate to the FES averages about 70 percent each year. Given that non-response is defined as occurring in households where substantive income or expenditure questions are refused by even only one individual (in which case, the whole household is

[8]Particular care should be taken when using data from panel studies. Attrition rates can lead to later waves of data becoming severely defective in terms of their ability to represent the whole population.

rejected from the final sample), this is considered to be a high rate of response.[9] It is likely that non-response will not be evenly distributed across households of all types. For the FES, where very little information is available on non-respondents, comparisons have been made between the characteristics of respondents and information from other sources, for example the 1971 and 1981 censuses. These indicate that differential response is associated with age of head of household, presence of children and employment status.[10] In addition, comparisons with the *Survey of Personal Incomes* suggests that those with high incomes are less likely to respond than others.[11]

On being presented with a new dataset, designed at least to be representative of a particular population, some simple comparisons with independent information should be made. If the survey data are multiplied up to give the population total of units, the characteristics of these can be compared with those of external data from censuses, other surveys, administrative statistics and national accounts (where these are from independent sources). In the first place the type of characteristics that are likely to lead to differential response, such as numbers of unemployed or self-employed, and numbers of children or pensioners can be examined. This can be followed by cross-classifying these categories where possible. The definitions of the characteristics can be varied, for example to see if it is the people who say they are unemployed who respond at a different rate than average, or if it is those actually receiving unemployment benefits who are responding at a different rate. Account will often need to be taken of differences in definition between the survey and the external source—does a "pensioner" mean a person old enough to receive a National Insurance retirement pension or someone who actually does receive one, for example. Furthermore, the most recent external population totals may post-date the survey and can be used to adjust for change since the data were collected.

The Grossing-up Procedure

To make the survey totals agree with external information, while introducing as small a change as possible, a re-weighting procedure needs to be carried out. This "grossing-up" procedure is described in more detail in Atkinson, Gomulka and Sutherland (1988). The characteristics chosen to be "control totals" relate to income range, family composition, housing tenure and employment status, plus some additional totals relating to the receipt of specific benefits. The weight for each individual family depends on its particular combination of these characteristics. For each characteristic, the weighted sum of the families with that characteristic is equal to the external total (i.e. the population total). The need

[9]In surveys where missing information on key questions is not grounds for exclusion of the household altogether, the response rate may appear to be higher, but an additional problem arises in dealing with the missing information. The most straightforward solution is to reject households (or families, where household information is not required) with missing values on necessary variables and to deal with the problem in the same way as is used in the U.K. for TAXMOD. This may have an unacceptable impact on the representativeness of the remaining households and the alternative of imputing missing values will need to be considered.

[10]See Kemsley (1975) and Redpath (1986).

[11]See Central Statistical Office (1979).

for a grossing-up procedure such as this is borne out by comparing TAXMOD results with and without differential weights. For example, gross expenditure on child benefit (a tax-free, universal benefit paid on behalf of all children) is over-estimated by some 25 percent if uniform weights are used. This is because, grossed-up to give the population number of tax units using uniform weights, the model produces 16.7 million children, instead of the 12.5 million known to exist from administrative statistics.

Some Remaining Problems

However, when assessing the effect of grossing-up on the data, it is important to look not just at the dimensions of the problem that initially concerned us and where we have made explicit correction, but to also look at other aspects of the data, to assess how much these have been distorted by the procedure. For example, the number of single women receiving widows' benefits has fallen by 17 percent in the U.K. since 1982.[12] In the TAXMOD grossing-up procedure, an external total corrects for this and the weights of widows are lower than they would be otherwise. However, for the total number of single women to remain the same, which it must since this is controlled for, the weights for other single women must be higher. This in turn may mean that the weights of other groups with which single women have relevant characteristics in common (for example, some single men) may have to be lower than otherwise. The repercussions of introducing a new dimension of control are not easily foreseen, especially when the number of these dimensions becomes large. Furthermore, the larger the number of control totals becomes, the smaller the number of observations in each "cell" (i.e. with each combination of characteristics being controlled for). This can result in some individual weights being very large or very small (varying say from one tenth of the mean weight to ten times the mean weight). This *need* not be a problem, but could present difficulties if observations with large weights also had characteristics that were not controlled for but which became relevant in a policy change.

So, the advice to the *government* was that, on the one hand it is important to be economical with the number of control totals, to keep cell size up, and on the other hand we may wish to introduce additional totals to control for the full range of characteristics that may be relevant to policy analysis. In practice, this tension has been largely resolved by the fact that few relevant totals are available for the U.K. The main limitations on the choice of these external totals has been the availability of recent data collected independently from the sample survey. For example, the age of head of household was one of the variables found to explain differential response and this would be an obvious choice as one of the control totals. However, there is no U.K. source of the distribution of age of head of household since the 1981 census. Our experience has been that it is not worth including a total that only approximates to the actual dimension of control required. This tends to increase the range of the size of the weights without the desired improvement in the data.

However, complete availability of external information describing the current system would not solve all the problems. Indeed, it could be argued that if one

[12]Source: *Government Expenditure Plans 1990/91-1992/93*, CM1014, Table 14.11.

were to introduce explicit totals to describe the whole tax-benefit system, using external information on all possible dimensions of each tax and benefit, one could accurately model the current system but that, with a limited sample size, the model would probably be very poor at estimating the effect of any *change* in the system. Effects of change would be dominated by observations with very high weights and the specific characteristics of those observations would determine the outcome. The very adjustments introduced to accurately model the current system may make the modelling of change *less* accurate. For example, if claimants of a particular income-tested benefit are under-represented in the original data, an explicit total can be introduced to correct for this. As with widows, this will *lower* the weights of similar families, including those with incomes slightly too high to be currently entitled to the benefit. Modelling a change in the benefit that extends the range of income over which families are entitled will draw in these families with lowered weights, underestimating the cost of making the benefit more generous (and the cash gain to the population of potential recipients). It may be that the dimensions controlled for in grossing-up should be kept general and broadly defined and that specific problems are best dealt with in an *ad hoc* way that does not have repercussions elsewhere.

It should be remembered that grossing-up and updating procedures are not good substitutes for high quality, recent data. In particular, these procedures cannot accurately model all the relevant aspects of structural change. The treatment of the increase in owner-occupied housing advocated here does not fully solve the problems in modelling the change in the composition of owner-occupiers that has taken place. Nor is it clear that by modelling the lowering of unemployment and employment through grossing-up at the same time as adjusting the earnings distribution through the updating procedure, we end up with an accurate description of the current labour force. The most strongly worded piece of advice to the *government* was that they should begin collecting new data for a future model as soon as possible.

5. Modelling the Current System and a New One

The Current System

Apart from income data, which are clearly the most crucial input into a tax-benefit model, other variables necessary to model the current system may not be available in the precise form required. For example, in the U.K. there is tax relief on mortgage interest payments but, in the case of some 20 percent of mortgages in the 1982 FES, the interest payment is not distinguished from the capital repayment. In these cases a method described in Atkinson, Gomulka and Smith (1988) to impute the missing mortgage interest figures uses information on the characteristics of the home-owners and the dwellings they live in for the cases where interest payments are known. Using this method, the model predicts the revenue cost of mortgage interest tax relief for 1989-90 to be £5.5 billion. Using an alternative and simpler method of substituting the mean interest figure from all the non-missing cases for the missing values gives a revenue estimate of £4.9 bn, a reduction of 11 percent.

In order for the model to consistently treat taxes and benefits under the current system and those introduced as the policy change, the model should calculate liabilities and entitlements in both cases and not simply adopt the amounts from the data for the current system. However, the information on benefit receipt is useful, particularly for benefits that are contingent on contribution conditions or specific characteristics (such as disability) that may not be otherwise observable in the data.

Once a model can describe the current tax and benefit system, it is possible to start validating its output. By now, much of the available relevant external information will have been used in updating and grossing-up the data. Therefore, it would not be surprising to find that expenditure on a benefit contingent only on characteristics controlled for in these processes (for example, child benefit or widows' benefits in the case of TAXMOD) is now estimated to be more or less the same as administrative statistics would lead us to expect. However, there will be other model outputs for which there have been no explicit controls that will need to be cross-checked with other sources before the model can be considered to be a reliable tool for assessing the effect of policy *change*. Good examples would be income-dependent taxes and benefits. Although income range may be one of the control totals in the grossing-up procedure (as it is in the case of TAXMOD) and income from different sources may have been most carefully updated, these adjustments are essentially crude. It is the quality and the richness of the micro-data that the model depends on and which will need to be tested.

One of the differences between TAXMOD and the model to be developed by the *government*, is in the methods available for validating model results. TAXMOD results can be compared to official U.K. government figures and decisions taken about which assumptions to use, which set of grossing-up totals to select and so on according to how close the final result is to the official version. Of course, this is not a foolproof method as government figures may well be estimates obtained from the same data using similar techniques to those employed in TAXMOD. Favourable comparisons betwen models may result from errors in both of them. However, some comparison is better than nothing, and for a country where no such model exists at all there can be problems in choosing between different plausible options, apart from on *a priori* grounds.

Another sort of testing—of the accuracy of the tax-benefit calculations themselves—should not be overlooked. The ability of models to display the detailed results of their calculations for individual families (or other units) is invaluable in this respect. Not only do they facilitate the checking for errors, but they can most usefully demonstrate features of a tax-benefit system that have not necessarily been foreseen or understood. What appear at first sight to be programming errors may on closer inspection turn out to be features of the tax-benefit system that have not been fully appreciated. The educational nature of these models is worth remembering. A good example of this is the unexpected fact that a cut in the basic rate of income tax in the U.K. results in some *losers*. Examining individual cases from TAXMOD shows us that these arise for the following three reasons:

(1) Means-tested benefits are assessed on post-tax incomes. A fall in tax may result in loss of entitlement to benefit. Since receipt of these benefits

can act as a passport to some additional benefits in kind (such as milk for infants and free dental care), and also because payments of benefits are subject to a minimum entitlement, the *loss* in value of the benefit package can be greater than the gain from reduced tax.

(2) Mortgage interest relief is deducted at source at the basic rate of income tax regardless of the taxpayer status of the borrower. If the basic rate falls then the relief is *less* and the payments rise, even for non-taxpayers.

(3) Some investment income is taxed at source at a "composite" rate, slightly lower than the basic rate. Higher-rate taxpayers must pay additional tax on the income grossed-up at the basic rate. The larger the difference between the basic and higher rates, the greater the extra tax paid. High income people with large amounts of investment income may *lose more* from their increased higher rate tax than they gain from the fall in the basic rate.

Modelling a New System

Modelling the system that was current when the data were collected is relatively straightforward. While it will not often be the case that income questions in a survey will mirror those required by the income tax schedules or benefit claim forms existing at the time (which would be ideal), at least the concepts should be similar. However, this may not be the case when the *government* is considering radical policy change. If a new tax is to be introduced (such as on property) or a benefit designed for a new category of claimant not covered before, then *no* information in the area may have been collected in the survey. For example, the only information on disability in the U.K. FES is limited to information on current disability benefit receipt. If a benefit with different qualifying conditions were proposed, this would be difficult to model without some arbitrary assumptions about who in the sample survey should receive the benefit. Random assignment of the projected number of claims or the matching and merging of data from other sources would be two alternative options. However, the introduction of another source of uncertainty in a general model would need to be assessed on its own merits. The *government* was warned that while the merging data from different sources could be useful, the approach had its limits.

Modelling Departures from The Rules

If the rules governing direct taxes and benefits are specified in detail and applied to data, corrected for as many defects as possible, the results may still not correspond to what is observed *ex post.* This could be because the response to a policy change has itself had an effect on incomes, or because of exogenous changes. These are discussed in the next section. It could also be because the rules describing the tax-benefit system do not adequately describe what happens in fact. There may be a number of reasons for this. Individuals may *break* the rules and, for example, may evade taxes by withholding information from the authorities. It can fairly safely be assumed that these people will either refuse to respond to the survey and will not be represented or will similarly withhold or understate income in their responses. In that sense the data, and the application

of the rules, will mirror what happens in the real world. However, even if measures were introduced that were effective in reducing evasion these would be difficult to model. Also people may not claim benefits to which they are entitled, either because of the social and other costs attached to claiming or because they are not aware that they may be entitled. Where this occurs to a substantial degree, as in the U.K. where take-up of Family Credit is estimated as 50 percent on a caseload basis (and 65 percent in expenditure terms), it is important that this effect is included in the modelling.[13] Otherwise not only would the cost of this benefit be overestimated, but also the effect on the incomes of families who do not claim would remain hidden.

In addition there may be *mistakes* made in the official calculation of entitlements. How often mistakes are a problem for the individuals concerned and for the accuracy of the model may be indicated by the U.K. government's own estimate of 9.1 percent of Income Support payments that were incorrect in Great Britain in 1988-89.[14] Unless we have evidence to the contrary we can assume that mistakes are symmetrically distributed about the value predicted by the rules and are independent of the characteristics of the family. In this case they would, on average, make no difference to the final revenue estimate. However, this would affect the distribution of the amount of benefit received. In general, they are akin to sampling error in the survey in the sense that they increase the uncertainty with which the estimate is made. If there is evidence that mistakes are biased in one direction or that certain sorts of people are more likely to receive incorrect payments, then any substantial effect should be modelled.

6. Static vs. Dynamic

One of the key debates within the *government* was over the effect on incentives of introducing a reduced rate of income tax financed by a lowering of tax thresholds. Whether or not the effect on labour supply would result in an additional revenue cost was a matter of great speculation and argument and there was an unwillingness to propose a precise new revenue-neutral rate and threshold without knowing the size and direction of the behavioural effect.

It was therefore considered to be of key importance to include in any model a simulation of the behavioural response to policy change and to produce a dynamic estimate of the revenue cost which took account of any individual adjustment to pre-tax earnings that resulted from the change. This particular example highlighted the excitement felt by *government* officials about the prospects of a fully dynamic model which would tell them the final or "true" cost of each package of reforms (or design for them a truly revenue-neutral package). Furthermore, they did not just want to produce fully dynamic cost estimates, they were also concerned about the impact of changed incentives to work (or, in the case of some of the other policy proposals, incentives to do other things such as split families or re-allocate capital assets among family members) on particular groups in the population for social policy or political considerations.

[13] *Hansard* written answers 17th March 1989, col. 391-392.

[14] Income Support is the main means-tested benefit. Source: *Government Expenditure Plans 1990/91-1992/93*, CM 1014, Table 14.31.

These concerns were appreciated by the author, but she felt there were a number of points that the *government* should understand about the problems of incorporating the estimation of indirect effects that might cause them to be less optimistic.

The understanding of behavioural response to changes in income is limited to particular cases and is even then a subject of great uncertainty. Most of the research on "behavioural response" has concentrated on the labour supply response of men or women to changes in net income. Even within this area, estimates differ widely and it would never be clear which estimate should be used in a practical application or on what basis a specific result might be generally applied. For example, Hausman's widely quoted U.S. estimates (1981) excluded the self-employed, people aged under 25 or over 55, farmers, single women without children and the disabled. If one were to apply these results to the labour force in the TAXMOD dataset, *more than half* of it would consist of people excluded by Hausman.[15] Furthermore, one would be assuming that U.S. estimates were applicable to the U.K. some 14 years later. As Burtless (1986) says in the case of the U.S. income maintenance experiments:

> "In comparison to the large number of studies of experimental labor supply response, there have been only few studies attempting to generalize the findings from the experiments to the U.S. population. Predictions of the nationwide response to a negative income tax are rare because they are costly to obtain."

The effect on each individual of the two tax changes described above—a lower initial rate of tax and lower thresholds—would depend on their position in the income tax schedule. The group who were brought into tax would face both higher marginal rates and higher average rates. The group whose marginal rate fell because of the new reduced rate band might face either lower or higher average rates, depending on the precise new rate and threshold chosen. The group whose marginal rate remained unaffected because their taxable income was above the reduced rate limit could also face higher or lower average rates. This would depend on whether the amount they lost in paying tax on a larger proportion of their income was offset by the amount they gained by paying less tax on the lowest tranche of taxable income. The changes in marginal and average rates induced by such a policy change in the U.K. are shown in Table 1. This shows that the same proportion of families face increases and decreases in their average rate and that the marginal rates of the individuals involved may rise, fall or remain unchanged.

The direction of any predicted labour supply response would not be the same for all the groups and separate estimates would be needed for each. Indeed, the effects may well be further differentiated by sex, by whether the person was in full-time or part-time work and would vary according to local labour market conditions. Furthermore, there may be cross-substitution effects in families with more than one earner. If it were feasible to provide estimates of labour supply response with some certainty in all these separate cases, then an accurate estimate of the dynamic effect of these tax changes would be possible.

[15]A number of different plausible definitions were tried, all giving similar results.

TABLE 1

CHANGES IN AVERAGE AND MARGINAL TAX RATES AFTER REDUCING INCOME TAX THRESHOLDS AND THE STARTING INCOME TAX RATE

	Percent Who Experience		
	A Fall	A Rise	No Change
Average rates (families)	41	41	18
Marginal rates for heads	20	4	76
Marginal rates for wives	24	22	54

Source: TAXMOD for the U.K. in 1989–90.

Notes: The introduction of a reduced rate band of 15 percent on taxable incomes up to £5,000 per year above tax thresholds is financed by a fall in thresholds of 45 percent.

"No change" includes families or individuals who pay no tax in either system.

Marginal rates are only calculated for individuals with earned income.

However, variation in the estimates of response may also arise because the models of behavioural response are different; or because the size of the marginal change under examination varies; or because the estimation procedures vary. Killingsworth and Heckman (1986) reproduce results for one set of variable definitions using different estimation procedures showing a variation in elasticities of female labour supply of between +0.09 and −0·05. It is most unlikely that a model incorporating a single elasticity chosen somehow from such a possible range would inspire much confidence in the results.

The scope for manipulation of model results becomes greater with the increased choice of inputs in the form of estimates of behavioural effects. Indeed, the *government* officials appreciated the point made by McLure and Zodrow (1987) with reference to the modelling work carried out in the analysis of the recent U.S. tax reform:

> "Static estimates are almost certainly wrong, but they are less subject to manipulation than are 'dynamic' estimates. While the economists in the Office of Tax Analysis have been traditionally quite free from effective pressure to 'shade' revenue estimates up or down for political purposes, they might be less successful in resisting such pressure in a world of widespread dynamic revenue estimation, which would inevitably entail making controversial assumptions concerning the dynamic responses of the economy to tax reform."

Research has concentrated on the effect of changes in income arising from changes in taxes or transfers on *labour supply*. The difference between the dynamic and static effects of changes affecting behaviour other than labour supply (such as savings decisions) may be of as much significance as any of these well studied, but still not fully understood labour supply effects. A *fully* dynamic model would require that all changes in behaviour induced by changes in policy that had some impact on net incomes should be modelled. To chose some and not others would be misleading.

Furthermore, an estimate of the behavioural effect will not be the end of the story. Either the initial change in policy or the behavioural responses to it may

have an effect on the macroeconomy which will need to be taken into consideration in a fully dynamic model. For example, reducing tax burdens on low paid workers may have the effect of lowering their bargaining power in the next wage negotiation. Lower wages themselves will have a macroeconomic effect. On the other hand, tax cuts that are inflationary may reduce real living standards in the end by more than the initial increase in net income. The government's own behaviour needs to be predicted. If its response to inflationary pressures is to increase interest rates, this will increase the nominal incomes of savers and decrease those of borrowers, apart from any additional macro effect.

In addition, there is the question of what time-scale a policy change should be assessed over. It is not clear that a time horizon could be established in which the full dynamic effect of a change could be considered to be fully worked through. Once any but the shortest time-period is of interest, static modelling techniques are clearly no longer appropriate and the construction of a dynamic *longitudinal* model, which allows a population to age in all respects, needs to be considered.

Although the main recommendation to the *government* was that estimates of indirect effects should only be attempted with extreme caution, it was made clear that in cases where a policy change was expected to offer significant incentives to change behaviour, claims that any *static* estimate is realistic need in turn to be treated with caution. The implications of the pattern and size of changes in marginal tax rates should then be considered carefully.

7. Recommendations and Conclusions

The recommendations made to the *government* are summarised as follows:

(1) to formulate policy proposals with sufficient precision that they may be translated into computer-readable form,
(2) to decide which unit of analysis is most appropriate for the policy questions at issue, without having to make unacceptable assumptions about the resources available to that unit; where possible to allow for variation in the unit of analysis,
(3) to find alternative methods of assessing policy options that cannot be realistically modelled using these methods,
(4) to be cautious about attempting dynamic estimation in a generally applicable tax-benefit model,
(5) to collect new data, designed for the purpose and to process it as quickly as possible.

The last recommendation was the most important. A number of general guidelines for the design of the survey and its subsequent treatment were proposed as follows:

(a) The data should have the capacity to be analysed at several levels. Not only should the different appropriate levels be identifiable, but where possible the data should be defined at each level.
(b) The income data should be as accurate as possible and in sufficient detail to model not only the current system, but also any envisaged system with precision.
(c) The population coverage should be as extensive as is feasible.

(d) The method of updating incomes should be as disaggregated as possible.
(e) The grossing-up procedure should adjust for differential non-response. It should also take account of the changes in characteristics of the population between the time the data were collected and the date of the tax-benefit system being modelled. If the survey is processed quickly and if it is repeated regularly so that successive versions of the model have the most recent data possible, then this second adjustment [and those in (d) above] should be small.
(f) The effect of the assumptions made should be monitored. In particular, the imputing of missing information should be carried out with care and, where any external data are available for validation, the assumptions should be tested against plausible alternatives.

This paper has been addressed to two audiences. Those unfamiliar with tax-benefit models should be aware of the precise requirements they have in terms of data and other inputs, such as the detailed specification of policies and the choice of unit of analysis. Those familiar with the methods and pitfalls of model construction need to develop methods of conveying the limitations of these models. The dialogue between those who build the models and those who might make use of them needs to be improved.

References

Atkinson, A. B., Gomulka, J., and Smith, J. M., The Treatment of Mortgage Interest in the FES, in Atkinson, A. B. and Sutherland, H. (eds.), *Tax-Benefit Models*, STICERD Occasional Paper 10, London School of Economics, 1988.

Atkinson, A. B., Gomulka, J., and Sutherland, H., Grossing-up FES Data for Tax-Benefit Models, ESRC Programme on Taxation, Incentives and the Distribution of Income Discussion Paper No. TIDI/105, London School of Economics, 1988.

Atkinson, A. B. and Micklewright, J., On the Reliability of Income Data in the Family Expenditure Survey 1970-77, *Journal of the Royal Statistical Society*, Series A, 146, 33-61, 1983.

Atkinson, A. B., Micklewright, J., and Stern, N. H., Comparison of the FES and the New Earnings Survey 1971-1977, in Atkinson, A. B. and Sutherland, H. (eds.), *Tax-Benefit Models*, STICERD Occasional Paper 10, London School of Economics, 1988.

Atkinson, A. B. and Sutherland, H., (eds.), *Tax-Benefit Models*, STICERD Occasional Paper 10, London School of Economics, 1988.

Bourguignon, F., Chiappori, P-A., and Sastre-Descals, J., SYSIFF: A Simulation Program of the French Tax-Benefit System, in Atkinson, A. B. and Sutherland, H. (eds.), *Tax Benefit Models*, STICERD Occasional Paper 10, London School of Economics, 1988.

Brannen, J. and Wilson, G., (eds.), *Give and Take in Families: Studies in Resource Distribution*, Allen and Unwin, London, 1987.

Burtless, G., The Work Response to a Guaranteed Income: A Survey of Experimental Evidence, in Munnell, A. H. (ed.), *Lessons from the Income Maintenance Experiments*, Federal Reserve Bank of Boston and The Brookings Institution, Washington, D.C., 1986.

Central Statistical Office, Bias in the FES—Some Results, unpublished, 1979.

Dilnot, A., Stark, G., and Webb, S., The IFS Tax and Benefit Model, in Atkinson, A. B. and Sutherland, H. (eds.), *Tax-Benefit Models*, STICERD Occasional Paper 10, London School of Economics, 1988.

Galler, H. P., Policy Evaluation by Microsimulation—The Frankfurt Model, Paper prepared for the 21st General Conference of the International Association for Research in Income and Wealth, Lahnstein, W. Germany, 1989.

Gegesy, F., Juhász, M., and Szivós, P., The Hungarian Microsimulation System, Paper prepared for the 21st General Conference of the International Association for Research in Income and Wealth, Lahnstein, W. Germany, 1989.

Harding, A., Dynamic Microsimulation Models: Problems and Prospects, Welfare State Programme Discussion Paper WSP/48, London School of Economics, 1990.

Hausman, J. A., Labor Supply, in Aaron, H. J. and Pechman, J. A. (eds.), *How Taxes Affect Economic Behavior*, Brookings, Washington DC, 1981.

Hüther, M., Müller, M., Petersen, H-G., and Schäfer, B., Microsimulation of Alternative Tax and Transfer Systems for the Federal Republic of Germany, Paper prepared for the 21st General Conference of the International Association for Research in Income and Wealth, Lahnstein, W. Germany, 1989.

Kemsley, W. F. F., Family Expenditure Survey—A Study of Differential Response Based on a Comparison of the 1971 Sample with the Census, *Statistical News*, No. 31, 3-21, 1975.

Kemsley, W. F. F., Redpath, R. U., and Holmes, M., *Family Expenditure Survey Handbook*, HMSO, London, 1980.

Killingsworth, M. R. and Heckman, J. J., Female Labour Supply: A Survey, in Ashenfelter, O. C. and Layard, R. (eds.), *Handbook of Labour Economics*, Vol. 1, North-Holland, Amsterdam, 1986.

King, A., The Distribution of Current Disposable Incomes: An Estimation using Techniques of Microanalytic Simulation, NIEIR Working Paper No. 1, National Institute of Economic and Industry Research, Australia, 1987.

McLure, C. E. and Zodrow, G. R., Treasury I and the Tax Reform Act of 1986: The Economics and Politics of Tax Reform, in *Journal of Economic Perspectives*, 1 (1), 37-58, 1987.

Mertz, J., Microsimulation—a Survey on Principles, Developments and Applications with Focus on the Static Case and the Static Sfb3-Microsimulation Model as an Example, Sfb3 Working Paper No. 268, Frankfurt/Mannheim, 1988.

OECD, *A Comparative Study of Personal Income Tax Models*, OECD, Paris, 1988.

OPCS, *General Household Survey 1986*, HMSO, 1989.

Redpath, R. U., Family Expenditure Survey: A Second Study of Differential Response, Comparing Census Characteristics of FES Respondents and Non-respondents, *Statistical News*, 72, 13-16, 1986.

Thatcher, A. R., The Distribution of Earnings of Employees in Great Britain, *Journal of the Royal Statistical Society*, Series A, 31 (2), 133-170, 1968.