Sample Surveys for Household Income and Expenditure Information*

By J. E. G. Utting

University of Cambridge, Department of Applied Economics

I. Introduction

This paper discusses a number of problems in the planning and execution of household income and expenditure studies in relation to two specific inquiries which have recently been undertaken in the United Kingdom, and are at present at the stage of the analysis of results. The bodies responsible for these two investigations are the Ministry of Labour and the Cambridge University Department of Applied Economics. The Ministry of Labour inquiry is an official one and was undertaken on a national scale. The Department of Applied Economics inquiry is unofficial and local; it was restricted to the administrative county of Cambridgeshire.

The design of any survey is necessarily influenced by the purpose which it is intended to serve. Even within the general objective of collecting household income and expenditure information there may be different purposes for which that information is obtained. Such differences in purpose have in fact led to some of the differences between these two studies.

The prime object of the Ministry of Labour inquiry is to obtain new weights for a cost-of-living index; there have been earlier official household expenditure studies in Britain which have had the same object, but none has been carried out since 1937. For this purpose it is necessary to have a rather detailed description of the commodities bought and information about the quantity purchased as well as the price paid. On the other hand, such an investigation does not normally need to cover all classes of society or any persons living in institutions such as schools, students' hostels, nurses' homes and the like. However, there is also a considerable demand for information on the expenditure patterns

---

* A lecture delivered 1953 to the Volkswirtschaftliche Abteilung des Instituts für Auslandsforschung in Zürich and the Studiengruppe für Statistik, here printed with some alterations.

1 A fuller account of this study is given in «Sampling for Social Accounts - Some Aspects of the Cambridgeshire Survey», by J. E. G. Utting and Dorothy Cole, presented at the 28th Session of the International Statistical Institute, Rome, 1953.

2 The county lies mainly within a radius of 12–15 miles of the city of Cambridge. Its total population in 1951 was 166,863, of whom 81,463 were in the city. The remainder of the county is almost wholly rural.
of all sections of the community, for national income studies and for other purposes. This was recognized by the Ministry, and there was no attempt to limit the investigation to any particular income stratum or social group, although the inquiry excluded residents in hostels, schools and other institutions.

The Department of Applied Economics study is restricted to a purpose which is more like the second object of the official inquiry. It is part of an investigation in which the Department is trying to develop sampling methods for the collection of national income and social accounting information. It is not necessary for this purpose to have a knowledge of quantities as well as prices or to have a very detailed description of the goods bought. On the other hand, it is necessary that the coverage of the population should be as complete as possible (including residents of institutions) and that the sampling method employed should lead to good estimates of the aggregate transactions of the whole population as well as to averages for a particular group or groups.

II. Sample Design

The basic principles of sample design employed in the two studies are very similar though necessarily dependent upon the difference in scale. The nationwide Ministry of Labour inquiry started with a sample of some 20,000 addresses while the more limited Department of Applied Economics study used only 4,000 addresses. This is of course a very large proportion (7.4 per cent.) of the total number of addresses in the county of Cambridgeshire, but since the sampling error is related directly to the absolute size of the sample rather than to the sampling fraction it is inevitable that the latter should be rather high if reliable results are to be achieved in a small area.

In both inquiries the initial sample of addresses was obtained by a random selection from local authority rating lists. The Ministry of Labour sample was mainly selected in two stages: the first stage being a selection of local administrative areas which was made with probability proportional to the population of the area; and the second stage being a selection of equal numbers of addresses within the chosen administrative areas. An exception to the general sampling scheme was made in the case of a number of the largest cities which were set on one side and within which the sample of addresses was chosen with a uniform sampling fraction which would give each household in these cities approximately the same chance of being selected as was given to households in smaller areas by the two-stage sampling used for those areas.

The Cambridgeshire sample was also drawn from rating lists but since it was restricted to a small area there was no need for a first stage selection of local authority districts. The sample was, however, a stratified sample with a

---

1 These are lists of land and buildings which are maintained for local taxation purposes. They are therefore kept up to date. A fuller account of the use of these lists as sampling frames is given in «Sampling for the Social Survey», by P. G. Gray and T. Corlett, *Journal of the Royal Statistical Society, Series A (General)*, Vol. CXIII, Part II, 1950.
varying sampling fraction. Since the level of expenditure on different items is largely dependent on income, and the variance of expenditures seems to be closely related to their size, it is clearly advantageous to use a sampling scheme which approximates to a stratification by income level with larger sampling fractions for the higher levels. There is, of course, no way of direct stratification by income group before making the inquiry, but it has been found useful in the United Kingdom to stratify by rateable value, which is related to the size and rental value of the dwellings. The Cambridgeshire sample used four different strata. The first division was between the purely rural part of the county and the city of Cambridge together with some adjacent dormitory areas. Within each of these groups there was a further subdivision into two strata by rateable value, the dividing line being drawn at somewhat different levels in the two areas. The sampling fractions used were based on information about variances obtained in pilot studies and were as follows:

<table>
<thead>
<tr>
<th>Stratum</th>
<th>Low R. V.</th>
<th>High R. V.</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>city</td>
<td>6.7</td>
</tr>
<tr>
<td>2</td>
<td>city</td>
<td>13.3</td>
</tr>
<tr>
<td>3</td>
<td>rural</td>
<td>6.7</td>
</tr>
<tr>
<td>4</td>
<td>rural</td>
<td>20.0</td>
</tr>
</tbody>
</table>

Both these samples consisted of a number of addresses whereas for analytical purposes what is required is the expenditure of households. A sample of households drawn with the same probability as the sample of addresses can readily be obtained by interviewing all the households resident at the selected addresses. Because some dwellings contain more than one household, the effect of this is to yield a sample of households which is rather larger in total numbers than the original sample of dwellings. If there is variation between one area and another in the proportion of multi-household dwellings, this will affect the distribution of the household sample between those areas. To overcome this difficulty, the Ministry of Labour reduced its household sample to the same size as the original sample of dwellings by deleting, at random, one unvisited address for every case where an additional household was found living at a selected dwelling. In the Cambridgeshire inquiry the same problem arises in relation to the four different strata, but in this case all households were visited and the results will be adjusted by re-weighting in inverse proportion to the achieved sampling fractions.

The same procedure can in principle be applied to residents of institutions, but since it is characteristic of these places that they normally have a large number of residents it may not be very convenient to attempt to obtain information from them all. The Cambridgeshire survey attempted to overcome this difficulty by making a first stage selection of institutions, with a high sampling fraction, and then interviewing only a random sample of the residents at each selected institution.

---

1 The correlation between household income and the rateable value of the dwelling occupied is not high - partly because large dwellings are often occupied by several households - but stratification nevertheless allows one to use a higher sampling fraction for the more variable part of the population.
III. The Information Required and Methods of Collection

As in the case of sample design there are a number of marked similarities between the methods of collection which were used in the two studies. There are, however, certain differences which are of fundamental importance and which do not depend primarily on the differences of scale or of scope. The chief of these differences is that the Cambridgeshire inquiry relied entirely on memory for the whole of its information, whereas the basic expenditure information in the official inquiry was collected by asking respondents to keep a record of their expenditure over a period of three weeks. It is not possible in a short paper to discuss all the arguments for and against these two methods of collecting information on household income and expenditure, but only to indicate some of the reasons why the Department of Applied Economics has chosen to rely on memory.

First, past experience suggests that one may expect a considerably higher rate of response to an inquiry conducted entirely by interview methods than to one in which people are asked to keep records. Since the problem of non-response is one of the principal difficulties in judging the validity of the results obtained in these inquiries it is obviously important to try any procedure which can be expected to lead to an appreciable increase in the response rate. The Ministry of Labour made a payment to respondents, provided all members of the household took part, as an incentive to co-operation 1.

Secondly, there seems little doubt that the mere fact of keeping a record of expenditure may tend to distort the normal expenditure pattern of the person who is keeping the record. This cannot be the case if the information sought relates to some past period although it then becomes important to avoid distortion which is due to faulty memory. Such distortion is, however, also present in records. These depend on the memory of the person who is completing the record and it seems likely that, at best, the record will only be completed at the end of each day. In many cases, the period that passes before the expenditure is recorded will be appreciably longer, and in some cases the form will only be filled in at the end of the total period to which it relates. In these circumstances the memory of the respondent is likely to be at least as faulty as that of a person who is being prompted and guided by a good interviewer.

Thirdly, and less important, there is much less experience of memory methods already available than there is of record keeping. Since the Ministry of Labour's modified type of record-keeping inquiry was to be conducted at the same time as the Department's, this was clearly an incentive to experiment with quite a different method of collecting the information. Some at least of the results of the two inquiries are likely to be sufficiently comparable to permit some assessment of the relative effectiveness of the two methods.

---

1 The Cambridge inquiry has obtained the co-operation of 75.1 per cent. of the households in the original sample. In the Ministry of Labour inquiry, 67 per cent. of all households selected completed records of their expenditure.
It is a common feature of the two investigations that neither attempted to collect general expenditure information over a long period from any one household. The Ministry of Labour asked for records of all expenditure during each of three consecutive weeks while the Department of Applied Economics only asked people to remember all expenditure during the seven days immediately preceding the interview. In this respect there is a very marked contrast with some past inquiries in which people have been asked to keep records, or even to make memory estimates, for as long a period as a year.

Both inquiries supplemented this short period information on all expenditure by asking about expenditure during a longer period on certain kinds of goods and services. This is obviously desirable in many cases since it will help to reduce the variance of expenditure on goods which are not bought very frequently; but this advantage may be offset by the inaccuracy of memory or unwillingness to keep records for long periods. The Ministry of Labour used an interview to obtain regular housing expenditure (rent, rates, mortgage payments, etc.), the last gas, electricity and telephone bills, and details of coal purchases and house repairs, decorations and improvements during the last twelve months. It also asked for a separate record of other expenditure of a regular kind, such as licences, motor tax, insurance of all kinds, education fees and season tickets. The Department of Applied Economics covered all these items, except coal, in a similar way, but only asked for coal expenditure in the last three months. It asked people to remember their purchases of major durable goods (motor-cars, radios, refrigerators, vacuum cleaners, furniture, and the like) during the preceding twelve months; and for certain other important, but less memorable kinds of goods (clothing, jewellery, sports equipment, electrical and gas appliances, household textiles, etc.), it asked for expenditure in the last four weeks, as well as in the last week.

In addition to the expenditure information, the Ministry of Labour obtained details of income from all sources. The Department of Applied Economics also sought this information, but only from half of the households interviewed. The other half was asked to give savings information by answering questions about changes in their asset holdings, such as are asked in the Survey of Consumer Finances in the United States and in the savings surveys of the Oxford University Institute of Statistics in Great Britain. In this way it is hoped to obtain two independent estimates of personal saving, one as a difference between income and expenditure, and the other as changes in assets, and to make some comparison of the two methods.

Both inquiries are concerned with household income and expenditure, but both obtain this by interrogating individuals and aggregating their separate answers. The only exceptions to this principle are those items of expenditure.

which are covered by interview method by the Ministry of Labour (rent, rates, fuel, etc.). Information about these, and a very similar group of items in the Cambridge inquiry was obtained on only one form for the whole household, although it was sometimes necessary to interrogate more than one person to obtain all the details. This is because these expenditures are usually only incurred by one person in the household, although different individuals may be responsible for different items. Other expenditures, however, such as food and clothing, are not only incurred by different members of the household, but often by more than one member even in so short a period as a week. Both inquiries therefore asked every individual member of the household, except children with negligible personal expenditure, about all other items of expenditure, income or saving. So far as possible, provision was also made for any individual to give his information privately without revealing it to others in the household, although this is not always easy when there is limited space in which an interview can take place.

An important difference between the two investigations in respect of the information they were seeking is that the official inquiry asked for a detailed description and quantity of each purchase while the Cambridgeshire study asked only for the value, and listed items in fairly broad general groups. This is a direct reflection of the different purposes which the two studies are to serve. It is not clear how much the memory method would be affected if quantities and descriptions were required. Certainly the interview would take rather longer because of the need to record each item separately; whether it would impose much additional strain on the memory of the respondent is less certain, since much of the information given is in any case built up item by item.

IV. Allocation of Interviews over Time

Since both studies are concerned with quite short periods of time, except for relatively few items of expenditure, it was necessary to have some method of allowing for seasonal variations in expenditure or income. Both did this by spreading the inquiry over the whole of a year. The total sample was divided into a number of roughly equal parts, and each of these provided information for one of a series of consecutive periods. The Ministry of Labour inquiry had a fresh sub-sample of households beginning to keep records in each week of the year, while the Department of Applied Economics tried to collect information from different sub-samples during each of 26 fortnightly periods.

The Ministry’s sub-samples were so arranged that there was an even distribution over the whole year within each of the twelve administrative regions of the United Kingdom, so that it will be possible to analyse the results seasonally by regions, but not within individual towns or rural areas. The Department used an entirely random allocation over time within Cambridge; but, in order to permit some concentration of interviewing effort in the rural area, this was divided into thirteen districts which were visited in rotation, each being visited twice in the course of the whole year.
V. Estimation Procedures

The estimation procedures which are being used in both studies are largely determined by the desirability of obtaining information about seasonal variations of expenditure. The Ministry of Labour has information from expenditure records which begin on any day of the week other than Saturday or Sunday. In the Cambridgeshire inquiry the week's expenditure information refers to the seven days immediately preceding the day of interview which again could be any day of the week, except that interviews very rarely took place on Sunday. To overcome this difficulty the Department of Applied Economics is grouping together all the returns relating to weeks which end within a given period of 28 days. For analysis purposes these weeks will be regarded as equivalent and will be used to provide estimates of «an average week’s expenditure» for each 28-day period. There will thus be thirteen such averages in the course of the year and seasonal variations will appear as differences between the averages. Any seasonal change which occurs within a 28-day period will, of course, be obscured in the averaging process. Aggregate expenditure in the course of the year can then be estimated by regarding each 28-day period as the sum of four average weeks, and aggregating over the thirteen periods. Since the thirteen sub-samples are all independent the sampling error of the estimate can readily be obtained from the sampling errors for the average weeks. The Ministry of Labour is using a similar estimation procedure for the information obtained from its weekly records; it is combining them together in 8 periods of 4 weeks and 4 periods of 5 weeks which are so arranged that three consecutive periods cover a quarter (4+4+5 = 13 weeks).

In this way both inquiries could obtain estimates of total expenditure which were based entirely on short-period information relating to either one week or three weeks, respectively. It may be expected, though, that better estimates of some of the items will be obtained from the information which is available regarding expenditure in a longer period. In the Cambridge inquiry it is proposed to deal with those items which are covered for four weeks in a way exactly analogous to that already described. All four-week periods ending within a given 28 days will be regarded as equivalent and combined together for estimation purposes.

Some of the other long-period information relates to regular items of expenditure. These can all be converted into annual figures and the whole sample can be used to make estimates of annual expenditure. The expenditure recorded will be the rate ruling at the time of interview, so that if there is any trend in this expenditure over the course of a year it will be allowed for automatically.

In some cases the expenditure recorded related to a year, but this is the twelve months immediately preceding the interview. This may create a difficulty if there is any trend in expenditure on these items, since these periods of twelve months will in nearly every case be different from the year to which the estimates are intended to relate. An attempt will be made to find evidence of any such
trend by dividing the total sample into sub-samples relating to different time periods. If any significant trend is found the estimates will be adjusted appropriately.

In the Cambridgeshire inquiry these procedures will be used to obtain two different estimates of annual expenditure on all those items which are covered for more than a week. In general, it may be expected that the estimate derived from the longer period information will be more reliable than that given by the weekly figures, but this may not always be the case. Provided the two estimates do not prove significantly different it should be possible to use a statistical adjustment procedure which will yield a third estimate which is better than either. Where there is a significant difference between the two estimates it will be necessary to make a subjective judgment about which is to be preferred. It may be noted that, although the short-period information will lead to an estimate of total expenditure which is automatically the same as the total of estimated expenditure on the separate items, this will not necessarily be the case when longer period information is used. Here again there will be a need for statistical adjustment procedures which may need to be supplemented by an element of judgment. In the Ministry of Labour inquiry items covered for longer periods are being eliminated from the weekly records and no short-period estimates will be calculated for those items.

The method already discussed for estimating expenditure will also apply to the estimates of income or of savings.