

ON A PILOT SURVEY OF HOUSEHOLD INCOMES

JÁNOS AY – LÁSZLÓ VITA

With the transition to a market economy in Hungary, the conditions of conducting household income surveys have significantly changed. This is partly due to the proliferation of income sources, and partly to the introduction of taxation of income on a personal basis in 1988. Most of the newly emerged income items, e.g. entrepreneurial and property incomes, are “soft” with respect to measurement in contrast to typical income items under conditions of a planned economy and income policy. The significant extent of underground economy in Hungary has similar effects. The introduction of a comprehensive personal taxation has radically changed the attitude of population towards reporting true income since, at least in most respondents’ mind, doing so would be a “self-denunciation”. The fiasco of income survey attempted by the Hungarian Central Statistical Office (CSO) in 1993 is a convincing evidence for the aforesaid.

Therefore, by 1994, a new approach to gathering data on household income seemed to be necessary. A research project with this objective was launched by the Statistics Department of the Budapest University of Economic Sciences (BUES), sponsored by the Hungarian Scientific Research Fund (OTKA. Theme Number T 013505). The actual field work, i.e. the interviews in the selected households were conducted in April 1995 by students of the same university in Budapest and in two towns, while in four villages the interviewing was done by professional enumerators of the CSO.

In order to enhance the co-operation of the selected potential respondents, letters had been sent to about 2300 households, informing them about the objectives of the pilot survey, promising them absolutely confidential handling of the information obtained and asking for their co-operation. A pre-paid reply letter was attached, by which they could give three different answers to our request of co-operation: an affirmative or negative one, or they could state they might perhaps be persuaded at the time of the survey to co-operate. It was also stated in our letter that in case of no reply they would be visited by our enumerators.

88.6 per cent of the households approached chose this last option, while only 5 per cent of them denied the co-operation in writing. Such households were also requested to give the reason of their non-willingness to participate in the survey and to report some of the characteristics of their household. One third of such households did not give any

reason for their negative attitude, one sixth reported no faith in statisticians/politicians, and about one tenth reported lack of time.

Selection of the sample

Although the survey did not aim at producing nation-wide statistics on the living conditions and income of the population, because of cost considerations, we tried to spread the sample over a wide range in both social and regional sense. Namely, the Central Statistical Office recently experienced different attitude towards surveys in Budapest and the country-side respectively. Moreover, we thought that the attitude of replying to survey questionnaires might also differ in various strata of the population.

However, in order to obtain statistically significant data on income distribution and to assure its comparability with relevant data from other sources, at least for the capital city of Budapest, we concentrated two thirds of the sample to Budapest. The country-side was represented by a relatively small sample of households selected in two towns of county Fejér and four villages from county Pest. The selection of these two counties was deliberate, while the villages chosen were those where the professional enumerators conducting the regular household surveys of the CSO were also inclined to undertake our survey, however, in households not involved in the regular survey.

Originally we intended to select a random sample of households in Budapest and in the chosen towns and villages, respectively, but civil rights in Hungary did not allow the construction of a list of citizens' names with their addresses. At that time a list of dwellings with addresses could only be obtained. The post would not, however, deliver our afore-mentioned letters to addresses without names. At the same time, we were informed by the postal authority about the existence of documents, which describe mail-delivery districts. They order postmen to follow a given itinerary when delivering mail. The itinerary to be followed was given by listing a prescribed sequence of certain segments of the district, the segments consisting of blocks of houses, and indicating the number of floors and the number of private addresses within each block. Fortunately enough, the delivery of letters to all dwellings in any selected segment even without giving the names was possible. Thus, the primary sampling units became the mail-delivery districts, and one or more neighbouring segments within each primary unit were chosen as secondary sampling units. In each primary unit (district) the selected segments contained at least 23 households and a maximum of 45. Our afore-mentioned letter was then delivered to all private households within the selected segments. In Budapest we selected 36 districts, and 12 and 13, respectively in the two towns and in the four villages. The number of letters delivered in this manner was 1591 in Budapest, 398 in the two towns and 368 in the villages.

The enumerators had been instructed to visit first all the households (within the segment assigned to each of them), which gave a positive or "wavering" reply to our letters of approach, and then to supplement the sample with additional ones from among those, who had not replied at all. Each enumerator was given a target of completing the questionnaire in 12 households.

It should be obvious that, in spite of our effort to select a more or less random sample, the 578 households actually observed did not yield such a sample. On the one

hand, there were 118 households denying co-operation by answering so in writing and on the other hand, our enumerators also may have differed in their ability to convince the households actually visited. Still we hope that our findings with regard to testing our new approach to collect income statistics are valid.

The questionnaire

The questionnaire consists of two blocks: the items of Block *A* are related to the household as a whole, while items of Block *B* pertain to persons of the household as individuals having earned any kind of income in 1994. It is rather long and complex. This complexity is due to our original intention of conducting two or three consecutive surveys representing different approaches to and philosophies of income inquiry. However, financial considerations forced us to conduct a single survey, and we did not want to put aside our original plan either. Thus, we had to combine the planned several questionnaires into one. In fact, this solution also had an advantage, viz. the direct comparability of the results based on the various approaches. With some of the questions we finally tried to inspire the households' confidence.

The first item of the questionnaire was the number of persons in the household as on July 1 of 1994 and on 15 April 1995, respectively. The second item enumerated all individuals of the household born in 1981 and before (in descending order of age), indicating their year of birth, sex, highest school attainment, all sources of income they had in 1994 and the length of their presence in the household in 1994 (measured in months). Items 3 to 11 described the characteristics of the household's dwelling (size, tenancy, type of heating, amenities, etc.) and the expenditure connected with it.

Item 12 enquired about the average monthly net money income in 1994, item 13 about the estimated savings resulting from the consumption of goods and services produced by the household, item 14 requested the respondent to rate the level of living resulting from the income derived, while under item 15 and 16 they were requested to give an estimate on the amount of money income they would consider adequate for a decent life in 1994 and in 1995, respectively.

Items 17 to 20 were connected with the stock and use of car(s) in the household and the costs involved, items 21 and 22 requested the same information about the cottage and pets in the possession of the household. Items 23 to 28 put very detailed questions on all expenditures of the household, for 1994 as monthly averages and as a monthly figure for March 1995. Item 29 enumerated all durable household goods and appliances with the year, way and cost of acquisition.

Item 30 asked for information about the relation of their usual monthly income and expenditures, and under item 31 what they did if expenditure exceeded income. As a supplementary question to these, under items 32, 33 and 35, we asked whether they derived any income from leasing, selling, swapping or inheriting property, while item 34 asked if they had any debt. Item 36 inquired the amount derived from agricultural commodities produced, item 37 and 39 the amount received as gift and/or aid, item 38 the amount of aid rendered to (members) of other households. Item 40 inquired the amount spent on acquiring foreign currency. (At that time, but even now it is fairly frequent to keep savings in the form of some "hard" currency.)

It is also important to mention that the respondents were requested to assess the quality (accuracy) of every reply given in the form of a concrete amount and to indicate it by inserting a mark for it in the cell provided for this purpose after each amount. Amounts judged as perfect and accurate received mark 5, and with descending quality the marks went down to 1, this latter indicating a very unreliable reply. If someone was not able to give the amount a code 6 was written in the cell, a code 7 when the answer was denied, and the code 9 when the amount in question was 0 in the case of the interviewed household.

Block *B* was supposed to give a full account of the income received and taxes paid by all household members enumerated under item 2 of Block *A*. The first question inquired about the amount the household member contributed to the common budget of the household, the second question was how much he/she retained for "private" expenditures. Then 11 different tax-exempted income sources (e.g. fellowship, pension, casual work, interests, etc.) were enumerated. After these items 9 taxable income sources followed in the sequence of the tax-declaration form, and finally the amount of income tax paid was asked for. In this block it was also asked if the respondents relied on their income tax returns when reporting the amount of tax paid and the amounts of various taxable income items.

It is obvious that all income sources reported in item 2 of Block *A* must also appear as concrete amounts in the corresponding cells of Block *B*. To implement this requirement was given as an instruction to our enumerators. As a consequence only a refusal to answer our direct questions on income in Block *B* could have resulted the situation reflected by the figures of Table 1.

Table 1

Per cent of persons reporting not only the source but also the amount of their income

Source	Budapest	The two towns	The four villages
Employment	67.5	91.3	97.6
Entrepreneurship	37.3	.	.
Unemployment allowance	51.6	.	.
Property	.	.	.
Pension	93.9	100.0	92.2
Social benefits	64.7	89.6	.
Other	37.7	.	.

The income and expenditure indicators

The raw data collected by the questionnaire were transformed into several indicators for the purpose of further analyses. Two types of analyses were planned. The first type of analysis aimed at finding out if it were possible to acquire household income data of acceptable quality in Hungary in our days. To this purpose annual income data calculated for the households as units were used exclusively. The other type of analysis had the objective of studying the relationship between the households' income situation and their living conditions. To this purpose mostly monthly incomes calculated on a per capita

basis were used. In this article only some important findings of the first type of analysis will be reported.

Income and expenditure data of households are widely known to be dependent on household size and, by our presumptions, also on the character of the dwelling area. Thus, the first thing we had to do was to examine if the sample of 366 Budapest households had the same composition by these two variables as the parent population. It was found that this was not the case and, therefore, a re-weighting of our sample by these two variables was necessary. This re-weighting was made in such a way that the number of the households was raised to the actual number of Budapest households in the reference year 1994. The composition of these 772 thousand households obtained by the re-weighting mentioned is already in perfect coincidence with the structure of the parent population. This could, however, be done only with the Budapest households, since the conditions of a similar re-weighting for the towns and villages were not met.

From now on, mostly data of the re-weighted Budapest sample will be used.

With respect to the accuracy of the raw data, the following observations should be made. None of our questions provoked particular protesting effect. Only the question about the amount spent on buying foreign currency in 1994 seems to stand out in this respect, since 7 per cent of the Budapest respondents refused to answer this question. As to the proportion of “I do not know” answers, the questions related to the expenditures on cars, the question about the savings achieved by goods and services produced for self-consumption as well as the question about expenditures on repairing services are worth mentioning. In these cases the proportion of “I do not know” answers ranged from 13 to 20 per cent.

The modal accuracy code for most amounts was 4, but the average marks show characteristic differences. The accuracy of the raw data, as assessed by the respondents, seems to depend on two things: the “freshness” of the amount inquired after and the regularity of the item in question. For example, the mean accuracy code for the 1994 average monthly expenditure on medicine was 3.92, while for the same spending in March 1995 it was 4.22. At the same time, the mean mark for the 1994 monthly rent was 4.21 and 4.40 for the rent paid in March 1995. Some further information on the accuracy of a few important items is summarised in Table 2.

Table 2

The mean accuracy mark of selected items

Item	Budapest	The two towns	The four villages
Average net money income per month in 1994	4.14	3.97	4.04
Savings from self-consumption	3.82	3.37	3.67
The monthly income necessary for a decent life in 1994	4.02	3.64	4.19
The monthly income necessary for a decent life in 1995	3.98	3.60	4.27

Before conducting the survey we hoped quite a few respondents would rely on their income tax returns when answering certain questions. Table 3 furnishes some information in this respect.

Table 3

Percentage of persons relying on their income tax returns when answering certain questions

Subject of the question	Budapest	The two towns	The four villages
The amount of income tax paid	48.7	48.1	15.7
The amounts of taxable items	26.5	32.2	14.1

The figures in Table 3 obviously disappointed us, since we had hoped that many more respondents would depend on their income tax returns when answering questions of this type.

After this short characterisation of our data, we attempt to summarise the kind of indicators we formed from the raw data in order to measure the households' living standards. All the six indicators we formed were calculated on a household basis and for the whole reference year 1994. The six indicators were the following.

Declared Income (DINC) – the annual net money income of the household, calculated from the amounts declared in Items 12 and 13 of Block A;

Declared Household Budget (HHB) – the household members' contribution to the common household budget plus the amounts retained for their own purpose (based on the first two questions of Block B);

Household Expenditure (EXP) – the expenditure of the household excluding those spent on purchasing real estate, car and on the construction of new dwelling (based on the detailed questioning of household expenditures in Block A);

Total Household Expenditure (TEXP) – the total expenditure of the household including expenditures on purchasing real estate, car and on the construction of new dwelling;

Gross Household Income (GINC) – the sum of the household members' gross incomes reported in Block B of the questionnaire;

Net Household Income (NINC) – the sum of the household members' net incomes based on the gross income items and income tax amounts reported in Block B.

Table 4

The number of households allowing to construct the indicators

Indicator	Budapest		The two towns	The four villages
	before	after		
	the re-weighing			
DINC	358	757 472	132	78
HHB	317	674 585	130	56
TEXP	366	772 409	132	80
EXP	366	772 409	132	80
GINC	291	628 810	127	77
NINC	193	447 113	102	71
Observed households	366	772 409	132	80

From now on only the corresponding acronyms will be used throughout instead of the full names of the six indicators.

Four of the above six indicators are of income-type by their nature, while the other two are of expenditure-type with different content. The underlying assumption of

distinguishing between EXP and TEXP is that EXP is presumably financed from the current receipts of the household members, while financing TEXP may require former savings or loans, too. It can also be seen that five out of the six indicators are net of income tax and other possible deductions.

Since respondents had the possibility to deny answering any particular question, not all six indicators could be calculated for every household, but only to their certain sets. This is shown by Table 4, which also gives the effect of re-weighting in the case of Budapest.

Some results

The underlying idea of forming the six indicators was that the investigation of their consistency with each other may give useful clues if household incomes of acceptable quality can be procured at all in the present days, and if the answer to this question is affirmative, what the appropriate way of this might be. Before coming to these points, however, we briefly introduce the distribution of the Budapest households by the six indicators formed.

Table 5

Summary statistics of the Budapest households' distribution by the six indicators
(thousands of HUF)

Indicator	Mean	Standard deviation	Lower quartile	Median	Upper quartile
DINC	486	433.92	204	384	620
HHB	475	374.75	216	372	600
TEXP	678	767.56	277	464	746
EXP	585	531.65	276	452	702
GINC	644	731.35	192	400	780
NINC	503	594.98	180	333	539

Just to orientate the foreign reader we note the average exchange rate for 1994 was 105.13 Hungarian Forints (HUF) to 1 USD.

First of all we wanted to get an idea about the magnitude of sampling errors of the means of the six indicators. To facilitate the estimation of standard errors we relied on the cross-classification of households by size and type of dwelling area and regarded the cells of this cross-classification as strata. In order to find the standard errors we assumed simple random sampling within the strata and the weights of the strata were taken from the parent population. Just to demonstrate the results we are mentioning that the standard error of the mean of EXP was 27.64 thousand HUF.

The summary statistics in Table 5 are not strictly comparable to each other, since they refer to different sets of households. Therefore, they are not appropriate for the purposes of consistency analyses either. In principle, both for comparison and for consistency analysis only those households can be used, for which all six indicators are available. Such a requirement would, however, restrict the number of comparable households very much. Therefore, we finally limited this "availability requirement" only to the three

indicators DINC, HHB and EXP. This yielded 671 thousand “comparable” households, which is only 13 per cent less than the re-weighted number of all households.

Leaving out GINC is straightforward enough, since its gross character is very different from the net character of the other five indicators. TEXP could, however, easily be omitted, too, since, in contrast to the other four indicators, it was not only related to the current receipts of the households. NINC was finally left out only because its retention would considerably narrow down the set of households that could be examined. But its omission can also be justified by the “message” of Table 1. The selection of these three indicators can, however, be also justified by the consideration that DINC and HHB are the simplest possible measures of living standards, while EXP is a much more detailed measure of the same thing. Table 6 gives some summary statistics on the distribution of the comparable Budapest households by the three selected indicators.

Table 6

The distribution of the comparable Budapest households by the three selected indicators
(Thousands of HUF)

Indicator	Mean	Standard deviation	Lower quartile	Median	Upper quartile
DINC	470	314.44	228	384	610
HHB	477	375.26	216	372	600
EXP	557	418.74	277	452	694

From Tables 5 and 6 it is apparent that the standard deviation of DINC and EXP is considerably less in Table 6 than in Table 5. According to our investigations, this phenomenon stems from the greater variability of the households left out from Table 5 with regard to both DINC and EXP. On the other hand, both tables include practically all households, for which HHB is available and thus the distribution of this indicator is almost identical in the two cases.

The inter-consistency of the three selected indicators, what was in fact our focal interest, has been examined from several angles. First of all let the intercorrelations of the three indicators be shown in Table 7.

Table 7

Intercorrelations of the indicators DINC, HHB and EXP

Indicator	DINC	HHB	EXP
DINC	1.0000	0.7447	0.6672
HHB	0.7447	1.0000	0.5070
EXP	0.6672	0.5070	1.0000

The intercorrelations are not very strong, but they are not negligible either, all of them being highly significant. (The corresponding two-tailed p-values are 0 without exception.)

The inspection of Table 6 seems to suggest that the distribution of the households by the size of DINC and HHB is very similar, while their distribution by the size of EXP is rather different from the two distributions mentioned before. This is even more explicitly shown by Table 8, which gives the means and standard deviations of the differences between DINC and HHB, DINC and EXP, and HHB and EXP, respectively. The means and standard deviations are given for three different classes of the households. In the first class the first of the two paired indicators is smaller than the second one, in the second class the two indicators in question are nearly equal and in the third class the first of the two paired indicators is greater than the second one. Two indicators paired were considered nearly equal if the ratio of the first one to the second one varied between 0.95 and 1.05, and their order of magnitude was also qualified accordingly. Table 8 also gives the percentage distribution of all comparable households by the three classes defined above.

Table 8

The order of magnitude of the three indicators

Relationship of the indicators	Percentage of households	Mean	Standard deviation
		of the differences in 1000 HUF	
DINC is			
Smaller	} than/to HHB	-210	492.1
Nearly equal		0	7.3
Greater		212	244.0
Together	100.0	-7	252.9
DINC is			
Smaller	} than/to EXP	-245	303.4
Nearly equal		2	16.3
Greater		151	165.4
Together	100.0	-87	313.9
HHB is			
Smaller	} than/to EXP	-251	325.1
Nearly equal		0	14.4
Greater		200	378.1
Together	100.0	-80	396.0

Table 8 clearly shows that DINC and HHB are nearly equal for about two thirds of all comparable households, and the average differences in both directions are about equal.

However, the differences between DINC and EXP, as well as those between HHB and EXP show a quite different feature. They are namely not symmetric and in only 6 per cent of the cases there is no considerable difference at all. It would be worth to devote deeper analyses in order to find out what the factors are behind this phenomenon. At present we only have certain conjectures about the causes of these discrepancies. While underreporting may be supposed in case of DINC and HHB, for EXP overreporting can be assumed. Furthermore, in our experimental survey inquiring about expenditures was much more detailed than about incomes, which may have yielded a fuller account of expenditures than incomes.

Of course, we also examined how the cross-classification of comparable households looks like by certain identical intervals of any pair of the three indicators. Table 9 displays the results of one of these cross-classifications performed.

Table 9

The cross-classification of the households by DINC and EXP

DINC	EXP							Total
	-150	150-250	250-350	350-500	500-650	650-800	800-	
-150	13 859	17 369	17 987	-	-	-	-	49 215
150-250	31 009	30 366	33 785	32 121	9 321	-	-	136 602
250-350	8 869	15 588	19 606	44 232	979	2 679	3 264	95 217
350-500	979	4 637	36 498	33 522	38 591	14 290	15 288	143 805
500-650	-	-	10 509	20 113	28 642	13 265	19 627	92 156
650-800	-	3 040	-	-	14 521	22 936	28 533	69 030
800-	-	1 040	-	4 926	8 559	15 347	55 436	85 308
Total	54 716	72 040	118 385	134 914	100 613	68 517	122 148	671 333

Percentages of households appearing in the diagonal of the above and the two other tables constructed in a similar way are given below.

The per cent of households that fall into the same interval of various pairs of indicators:

Pair of indicators	Per cent
DINC and HHB	80.7
DINC and EXP	30.4
HHB and EXP	31.3

Table 9 and the column of figures above reveal inconsistencies, which are striking enough. This is, however, in fairly good agreement with the differences found before. (See Tables 6 and 8.) It is also worth mentioning that almost the same percentages were obtained for the households that claimed regularly to have a balanced household budget.

Conclusions

Though our experiences with the letters sent to potential respondents are negative with regard to reply, we are still convinced about the usefulness of contact letters before the enumerators call on households the first time. Such letters namely may

- avoid frustration of enumerators caused by resolute refusal,
- contribute to the elimination of eventual mistrust to receive unknown persons,
- make the detailed introduction of the subject of the survey by the enumerator unnecessary or, at least, reduce the time required by this,
- incite the respondent to think the subject of the survey over and thus to give more informed answers to the questionnaire.

As for anonymity, we may report rather mixed experiences. On the one hand, anonymity reduces the self-denunciation effect mentioned in the introduction. On the other hand it makes checking and amending questionnaires posteriorly impossible, which may create unwanted problems.

With respect to the use of income tax returns in the course of interviews we also found negative attitude in contrast to our hopes and expectations. If, however, income tax returns are not used as background documents in the course of collecting income statistics, it is not advisable to adjust the timing of the survey to the deadline for declaring incomes to the tax authority. Instead, the survey is to be conducted as soon as possible after the end of the reference year. This proposition stems from the experience gained and already reported with the accuracy codes we developed and used in our pilot survey.

Our most important conclusion from the pilot survey is that expenditures covered by current receipts, rather than incomes, could and should be collected, processed and published in order to reveal the “true” living standards of population in Hungary, and in our days. We can support this claim by the following facts. For Budapest in 1994 the average income in our sample (both DINC and HHB) practically coincided with the corresponding CSO figure based on the Household Expenditure Survey. In our survey, however, the mean of EXP and TEXP exceeded average income by about one fifth and one third, respectively. At the same time, CSO published almost equal means for expenditure and income. In addition, the CSO expenditure concept is supposed to be equivalent with our TEXP. Therefore, we think our EXP is in closer correspondence with prevailing Hungarian reality than any figure based on a direct inquiry on income.

Based on the aforesaid we think the most promising way to get reliable statistics on living standards of population is a survey taken immediately after the reference year, and with a questionnaire focusing on various expenditure items. A survey like this would be, with regard to its content, similar to traditional household expenditure surveys, however, it would neither be so deeply detailed nor lasting so long in time.